

```
In [94]: import wikipedia
import nltk
from nltk.corpus import stopwords
from nltk.tokenize import word_tokenize

t = wikipedia.search("google", results=30)
print t
plist = []
for p in t:
    print p
    page = wikipedia.page(p, auto_suggest=False)
    plist.append(page)
tokenized_docs_list = []
from nltk.tokenize import word_tokenize
for i in plist:
    tokenized_docs = word_tokenize(i.content.encode('ascii','ignore'))
    tokenized_docs_list.append(tokenized_docs)
```

```
[u'Google', u'Google Search', u'Google+', u'Google Play', u'Google Docs, Sheets, and Slides', u'Google Books', u'Google Translate', u'.google', u'Google hacking', u'Google Account', u'Google Chrome', u'Google Maps', u'Gmail', u'Google Glass', u'Google Hangouts', u'Thuppakki', u'G Suite', u'AdWords', u'Google Classroom', u'Google Doodle', u'Google Earth', u'Google Analytics', u'List of Google products', u'Google Drive', u'Googleplex', u'Vikas Gupta', u'Google Brain', u'Google Traffic', u'Google Dashboard', u'Motorola']
```

```
Google
Google Search
Google+
Google Play
Google Docs, Sheets, and Slides
Google Books
Google Translate
.google
Google hacking
Google Account
Google Chrome
Google Maps
Gmail
Google Glass
Google Hangouts
Thuppakki
G Suite
AdWords
Google Classroom
Google Doodle
Google Earth
Google Analytics
List of Google products
Google Drive
Googleplex
Vikas Gupta
Google Brain
Google Traffic
Google Dashboard
Motorola
```

```
In [106]: print len(tokenized_docs_list)
```

29

Remove the stopwords

```
In [107]: from nltk.corpus import stopwords
nltk.download('stopwords')
print "and" in stopwords.words()
t_nsw_list = []
for p in tokenized_docs_list:
    tokenized_nsw = []
    for i in p:
        if not i in stopwords.words():
            tokenized_nsw.append(i)
    t_nsw_list.append(tokenized_nsw)
print len(t_nsw_list)
```

```
[nltk_data] Downloading package stopwords to /Users/han/nltk_data...
[nltk_data]   Package stopwords is already up-to-date!
True
29
```

Stem the words

```
In [101]: from nltk.stem.porter import PorterStemmer
          from nltk.stem.snowball import SnowballStemmer
          from nltk.stem.wordnet import WordNetLemmatizer

          porter = PorterStemmer()
          #snowball = SnowballStemmer("english")
          wordnet = WordNetLemmatizer()
          after_stem_list = []
          for p in t_nsw_list:
              words = []
              for i in p:
                  words.append(porter.stem(i))
              after_stem_list.append(words)
```

[u'unilev', '(', ')', 'dutch-british', u'transnat', u'consum', u'good', u'compani', u'co-headquart', 'rotterdam', ' ', u'netherland', 'london', ' ', u'unit', 'kingdom', '.', u'it', u'product', u'includ', 'food', ' ', u'beverag', ' ', u'clean', u'agent', u'person', u'product', '.', 'It', 'world', "s", 'largest', u'consum', u'good', u'compani', u'measur', '2012', u'revenu', ' ', 'world', "s", 'largest', u'produc', 'food', u'spread', ' ', u'margarin', '.', 'It', u'europ', '7th-most', u'valuabl', u'compani', '.', u'unilev', 'one', 'oldest', u'multin', u'compani', ' ', u'product', u'avail', 'around', '190', u'countri', '.', u'unilev', u'own', '400', u'brand', ' ', u'focus', 'thirteen', u'brand', u'sale', 'one', 'billion', u'euro', ':', 'axe/lynx', ' ', 'dove', ' ', 'om', ' ', 'becel/flora', ' ', 'heartbrand', 'ice', u'cream', ' ', 'hellmann', "s", ' ', 'knorr', ' ', 'lipton', ' ', 'lux', ' ', 'magnum', ' ', 'rama', ' ', 'rexona', ' ', 'sunsilk', 'surf', '.', 'It', u'dual-list', u'compani', u'consist', u'unilev', 'n.v.', ' ', u'base', 'rotterdam', ' ', u'unilev', 'plc', ' ', u'base', 'london', '.', 'the', 'two', u'compani', u'oper', u'singl', u'busi', ' ', 'common', 'board', u'director', '.', u'unilev', u'organis', 'four', 'main', u'divis', u'food', ' ', u'refresh', '(', u'beverag', 'ice', 'cream', ')', ' ', 'home', 'care', ' ', u'person', 'care', '.', 'It', 'research', u'develop', u'facil', u'unit', 'kingdom', '(', 'two', ')', ' ', u'netherland', ' ', 'china', ' ', 'india', u'unit', u'state', '.', u'unilev', u'found', '1930', 'merger', 'dutch', u'margarin', u'produc', u'margarin', u'uni', 'british', u'soapmak', 'lever', u'brother', '.', u'dure', 'second', 'half', '20th', u'centuri', u'compani', u'increasingli', u'diversifi', 'maker', u'product', 'made', u'oil', u'fat', ' ', u'expand', u'oper', u'worldwid', '.', 'It', 'made', u'numer', u'corpor', u'acquisit', ' ', u'includ', 'lipton', '(', '1971', ')', ' ', u'brook', 'bond', '(', '1984', ')', ' ', u'chesebrough-pond', '(', '1987', ')', ' ', 'best', u'food', '(', '2000', ')', ' ', 'ben', '&', u'jerri', "s", '(', '2000', ')', ' ', u'alberto-culv', '(', '2010', ')', ' ', 'dollar', 'shave', 'club', '(', '2016', ')', '.', u'unilev', u'divest', u'special', u'chemic', u'busi', 'ici', '1997', '.', 'In', '2015', ' ', 'leadership', 'paul', 'polman', ' ', u'compani', u'gradual', u'shift', u'focu', u'toward', 'health', u'beauti', u'brand', 'away', 'food', u'brand', u'show', 'slow', 'growth', '.', u'unilev', 'n.v.', u'primari', u'list', 'euronext', 'amsterdam', u'constitu', 'aex', 'index', '.', u'unilev', 'plc', u'primari', u'list', 'london', 'stock', u'exchang', u'constitu', 'ftse', '100', 'index', '.', 'the', u'compani', 'lux', 'euro', 'stoxx', '50', 'stock', 'market', 'index', '.', '==', u'histori', '==', '===', u'1870s1910', '===', 'In', '1872', ' ', 'antoon', u'jurgen', ' ', u'found', 'first', u'margarin', u'factori', 'world', u'oss', ' ', u'netherland', '.', 'then', ' ', '1888', ' ', 'samuel', 'bergh', ' ', u'oss', ' ', u'open', u'margarin', u'factori', 'kleve', '.', 'these', 'two', u'compani', u'merg', '1927', 'form', u'margarin', u'uni', '.', '===', u'1910s1920', '===', 'the', u'initi', u'harvest', 'palm', 'oil', 'british', 'west', 'africa', ' ', u'news', u'report', 'seen', 'back', 'england', u'show', u'worker', 'abroad', u'favour', u'condit', '.', 'In', '1911', ' ', u'compani', u'receiv', u'concess', '750,000', u'hectar', 'forest', 'belgian', 'congo', ' ', u'mostli', 'south', 'bandundu', ' ', 'system', u'forc', 'labour', u'oper', '.', '===', u'1920s1930', '===', 'In', '1922', 'lever', u'brother', u'acquir', 'mac', u'fisheri', ' ', 'owner', 'T.', 'wall', '&', u'son', '.', 'In', u'septemb', '1929', ' ', u'unilev', u'form', 'merger', u'oper', 'dutch', u'margarin', u'uni', 'british', u'soapmak', 'lever', u'brother', ' ', 'name', u'result', u'compani', 'portmanteau', 'name', u'compani', '.', '===', u'1930s1940', '===', 'In', u'1930', u'busi', 'grew', 'new', u'ventur', u'launch', 'africa',

'latin', 'america', '.', 'the', 'nazi', u'occup', u'europ', 'second', 'world', 'war', 'meant', u'unilev', u'unabl', 'reinvest', u'capit', u'europ', ',', 'instead', u'acquir', 'new', u'busi', 'UK', 'US', '.', 'In', '1943', u'acquir', 'T.', 'J.', 'lipton', ',', u'major', 'stake', u'frost', u'food', '(', 'owner', u'bird', 'eye', 'brand', ')', u'batche lor', u'pea', ',', 'one', 'largest', u'veget', u'canner', 'UK', '.', 'In', '1944', ',', u'pepsod', u'acquir', '.', 'after', '1945', u'unilev', "s", u'success', 'US', u'busi', '(', 'lever', u'brother', 't.j.', 'lip ton', ')', 'began', u'declin', '.', 'As', 'result', ',', u'unilev', 'be gan', u'oper', '^', u'hand', '"', u'polici', u'toward', u'subsidiar i', ',', 'left', 'american', u'manag', u'devic', '.', '===', u'1950s196 0', '===', 'sunsilk', 'first', u'launch', 'UK', '1954', '.', 'dove', 'f irst', u'launch', 'US', '1957', '.', u'unilev', 'took', 'full', 'owners hip', u'frost', u'food', '1957', ',', u'renam', u'bird', 'eye', '.', 't he', u'us-bas', 'good', 'humor', 'ice', 'cream', u'busi', u'acquir', '1 961', '.', 'By', u'mid-1960', u'laundri', 'soap', u'edibl', u'fat', 'st ill', u'contribut', 'around', 'half', u'unilev', "s", u'corpor', u'pro fit', '.', u'howev', 'stagnant', 'market', 'yellow', u'fat', u'increa s', u'competit', u'deterg', u'soap', 'procter', '&', u'gambl', u'forc', u'unilev', u'diversifi', '.', 'In', '1971', ',', u'unilev', u'acquir', u'british-bas', 'lipton', 'ltd', u'alli', u'supplier', '.', 'In', '197 8', ',', u'nation', 'starch', u'acquir', '\$', '487', 'million', ',', u'mark', 'largest', 'ever', u'foreign-acquisit', 'US', u'compani', 'po int', '.', '===', u'1970s1980', '===', 'By', u'1970', ',', u'acquisit', ',', u'unilev', u'gain', '30', 'cent', 'western', 'european', 'ice', 'c ream', 'market', '.', 'In', '1982', u'unilev', u'manag', u'decid', u're posit', u'unwioldi', u'conglomer', u'concentr', 'fmcg', u'compani', '.', 'In', '1984', u'unilev', u'acquir', u'brook', 'bond', '(', 'make r', 'PG', u'tip', 'tea', ')', '390', 'million', u'compani', "s", 'firs t', u'success', u'hostil', u'takeov', '.', 'In', '1986', u'unilev', u's trengthen', u'posit', 'world', 'skin', 'market', u'acquir', u'chesebrou gh-pond', '(', u'merg', 'chesebrough', u'manufactur', 'pond', "s", u'c ream', ')', ',', 'maker', 'rag', ',', 'pond', "s", ',', 'aqua-net', ',', 'cutex', ',', u'vaselin', u'anoth', u'hostil', u'takeov', '.', 'I n', '1989', ',', u'unilev', 'bought', 'calvin', 'klein', u'cosmet', ',', 'faberg', ',', 'elizabeth', 'arden', ',', 'latter', 'later', 'sol d', '(', '2000', ')', 'ffi', u'fragranc', '.', '===', u'1990', '===', 'In', '1993', u'unilev', u'acquir', u'breyer', 'kraft', ',', 'made', u'compani', 'largest', 'ice', 'cream', u'manufactur', u'unit', u'stat e', '.', 'In', '1996', u'unilev', u'merg', 'elida', u'gibb', 'lever', u'brother', 'UK', u'oper', '.', 'It', u'purchas', u'helen', u'curti', ',', u'significantli', u'expand', u'presenc', u'unit', u'state', 'sham poo', u'deodor', 'market', '.', 'the', u'purchas', 'brought', u'unile v', u'suav', u'finess', u'hair-car', 'product', u'brand', u'degre', u'd eodor', 'brand', '.', 'In', '1997', u'unilev', 'sold', u'special', u'ch emic', u'divis', ',', u'includ', u'nation', 'starch', '&', u'chemic', ',', 'quest', ',', 'unichema', 'crossfield', u'imperi', u'chemic', u'in dustri', '4.9', 'billion', '.', u'unilev', u'establish', u'sustain', u'agricultur', u'programm', '1998', '.', '===', u'2000', '===', 'In', 'april', '2000', u'unilev', 'bought', 'ben', '&', u'jerri', "s", 'sli m', 'fast', '1.63', 'billion', '.', 'later', 'year', ',', u'compani', u'acquir', 'best', u'food', '13.4', 'billion', '.', 'the', u'bestfoo d', u'acquisit', u'increas', u'unilev', "s", 'scale', u'food', 'americ a', ',', u'ad', u'brand', 'knorr', 'hellmann', "s", 'portfolio', '.', 'the', u'transact', 'second', 'largest', 'cash', u'acquisit', 'world', u'busi', u'histori', '.', 'In', u'exchang', 'european', u'regulatori', u'approv', 'deal', ',', u'unilev', u'divest', 'well-known', u'brand',

'oxo', ',', 'royco', u'batchelor', '.', 'the', 'year', u'compani', 'bo
 ught', u'worldwid', 'mustard', u'product', 'firm', u'maill', '.', u'mai
 ll', 'three', u'boutiqu', u'europ', ',', 'sell', 'mustard', 'pump', u't
 radit', u'maill', 'fashion', '.', u'pari', ',', 'dijon', ',', u'franc',
 'london', ',', 'UK', '.', 'the', u'merg', 'best', u'food', u'unilev',
 u'approv', u'isra', 'anti', 'trust', u'agenc', '.', 'In', '2001', u'un
 ilev', 'split', 'two', u'divis', ':', 'one', u'food', 'one', 'home',
 u'person', 'care', '.', 'In', 'UK', u'merg', 'lever', u'brother', 'eli
 da', u'faberg', u'busi', 'lever', u'faberg', u'januari', '2001', '.',
 'In', u'septemb', '2002', ',', u'compani', 'sold', u'specialti', u'oi
 l', u'fat', u'divis', ',', u'loder', 'croklaan', ',', 'rm814', 'millio
 n', '(', '218.5', 'million', ')', 'ioi', u'corpor', ',', 'kuala', 'lump
 ur', ',', u'malaysia-bas', 'oil', 'palm', u'compani', '.', 'As', 'par
 t', 'deal', ',', u'loder', 'croklaan', 'brand', u'maintain', '.', 'als
 o', '2002', u'unilev', 'sold', 'mazola', ',', 'argo', '&', u'kingsfor
 d', ',', 'karo', ',', 'golden', u'griddl', ',', 'henri', '"s', u'bran
 d', ',', 'along', u'sever', 'canadian', u'brand', ',', 'ach', 'food',
 u'compani', ',', 'american', u'subsidiari', u'associ', 'british', u'fo
 od', '.', 'In', '2004', u'unilev', 'sold', 'share', 'rushdi', 'food',
 u'industri', 'bashir', u'famili', u'start', u'use', u'barack', 'bran
 d', 'name', '.', 'As', '2014', 'roshadi', 'food', u'industri', 'one',
 'three', 'largest', 'tahini', u'produc', 'israel', 'one', 'largest',
 u'produc', 'tahini', u'worldwid', '.', 'In', 'may', '2007', u'unilev',
 u'becam', 'first', u'large-scal', u'compani', 'commit', u'sourc', 'te
 a', u'sustain', 'manner', ',', u'employ', 'rainforest', u'allianc',
 ',', u'intern', u'environment', 'ngo', ',', u'certifi', 'tea', u'esta
 t', 'east', 'africa', ',', 'well', u'third-parti', u'supplier', 'afric
 a', u'part', 'world', '.', 'It', u'declar', 'aim', 'lipton', 'yellow',
 'label', 'PG', u'tip', 'tea', u'bag', 'sold', 'western', u'europ', u'c
 ertifi', '2010', ',', u'follow', 'lipton', 'tea', u'bag', u'global', '2
 015', '.', 'In', u'septemb', '2009', u'unilev', u'agre', u'acquir', u'p
 erson', u'busi', 'sara', 'lee', u'corpor', ',', u'includ', u'brand', 'r
 adox', ',', u'badedda', u'duschda', ',', u'strengthen', u'categori', 'le
 adership', 'skin', u'cleans', u'deodor', '.', 'the', 'sara', 'lee', u'a
 cquisit', u'complet', '6', u'decemb', '2010', '.', '==', '20102014',
 '==', 'On', '9', 'august', '2010', u'unilev', u'sign', 'asset', u'pur
 chas', 'agreement', 'norwegian', u'dairi', 'group', 'tine', ',', u'acqu
 ir', u'activ', u'diplom-i', 'denmark', '.', 'On', '24', u'septemb', '20
 10', u'unilev', u'announc', u'enter', u'definit', 'agreement', 'sell',
 u'consum', 'tomato', u'product', u'busi', 'brazil', u'cargil', '.', 'O
 n', '27', u'septemb', '2010', u'unilev', u'purchas', u'alberto-culv',
 ',', 'maker', u'person', 'household', u'product', u'includ', u'simpl',
 ',', 'vo5', ',', u'nexxu', ',', 'tresemm', ',', u'mr', '.', 'dash',
 ',', 'US', '\$', '3.7', 'billion', '.', 'On', '28', u'septemb', '2010',
 u'unilev', 'evga', u'announc', u'sign', 'agreement', u'unilev', 'woul
 d', u'acquir', 'evga', '"s', 'ice', 'cream', u'brand', '(', 'amongst',
 u'other', ',', 'scandal', ',', u'variet', 'karabola', ')', u'distribu
 t', 'network', u'greec', ',', u'undisclos', 'amount', '.', 'In', u'febr
 uari', '2011', u'unilev', u'announc', 'switch', '100', '%', u'cage-fr
 e', u'egg', u'product', u'produc', u'worldwid', '.', 'In', 'march', '20
 11', u'announc', u'unilev', u'enter', u'bind', 'agreement', 'sell', 'sa
 nex', 'brand', u'colgate-palmol', '672', 'million', ',', u'unilev', 'wo
 uld', u'acquir', u'colgate-palmol', '"s', u'laundri', u'deterg', u'bran
 d', 'colombia', '(', 'fab', ',', u'lavomat', 'vel', ')', 'US', '\$', '21
 5', 'million', '.', 'In', 'april', '2011', u'unilev', u'fine', '104',
 'million', 'european', u'commiss', u'establish', u'price-fix', 'carte
 l', u'europ', 'along', 'P', '&', 'G', ',', u'fine', '211.2', 'million',

,', 'henkel', '(', u'fine', ')', '.', 'though', 'fine', 'set', 'higher', 'first', '(', u'discount', '10', '%', u'unilev', 'P', '&', 'G', u'admit', u'run', 'cartel', '.', 'As', u'provid', 'tip-off', u'lead', u'investig', '(', 'henkel', u'fine', '.', 'On', '24', 'august', '2011', u'announc', u'unilev', u'agre', 'sell', 'alberto', 'vo5', 'brand', u'unit', u'state', 'puerto', 'rico', '(', 'rave', 'brand', u'global', '(', 'brynwood', u'partner', 'VI', 'l.p.', 'On', '14', u'octob', '2011', u'announc', u'unilev', u'agre', u'acquir', '82', '%', u'russia-bas', u'beauti', u'compani', 'kalina', '.', 'On', '27', u'decemb', '2012', u'announc', u'unilev', 'would', 'phase', 'use', u'microplast', 'form', u'microbead', u'person', u'product', '2015', '.', 'In', u'januari', '2013', '(', u'unilev', u'agre', 'sell', u'skippi', 'peanut', 'butter', 'brand', '(', u'togeth', u'relat', u'manufactur', u'facil', u'littl', 'rock', '(', u'arkansa', '(', u'unit', u'state', 'weifang', '(', 'shandong', '(', 'china', '(', 'hormel', u'food', u'approxim', '\$', '700', 'million', '(', '433', 'million', '(', u'approxim', '540', 'million', ')', 'cash', '.', 'In', u'juli', '2013', u'unilev', u'increas', 'stake', 'indian', 'unit', '(', 'hindustan', u'unilev', '(', '67', '%', 'around', '2.45', 'billion', '.', 'On', '12', 'august', '2013', u'unilev', u'announc', u'sign', 'agreement', u'wish-bon', 'western', u'dress', u'brand', u'pinnacl', u'food', 'inc.', 'total', 'cash', u'consider', u'approxim', 'US', '\$', '580', 'million', '(', 'subject', u'regulatori', u'approv', '.', 'On', '6', u'septemb', '2013', u'unilev', u'enter', u'definit', 'agreement', u'acquir', 'premium', 'australian', 'tea', 'brand', 'T2', '.', 'On', '21', u'februari', '2014', u'unilev', u'sign', u'definit', 'agreement', 'meat', u'snack', u'busi', '(', u'includ', 'peperami', '(', 'uk/ireland', ')', 'bifi', '(', u'continent', u'europ', ')', 'jack', 'link', "s", '(', u'undisclos', 'amount', '.', 'In', 'march', '2014', u'unilev', u'agre', u'acquir', u'major', 'stake', u'china-bas', 'water', u'purif', u'compani', 'qinyuan', '(', u'make', 'water', u'purifi', '(', u'drink', 'water', u'equip', 'water', 'treatment', u'membran', '(', u'undisclos', 'price', '.', 'On', '22', 'may', '2014', u'compani', u'announc', 'sold', 'north', 'america', 'pasta', u'sauc', u'busi', u'includ', 'rag', 'bertolli', u'brand', u'japanes', u'compani', 'mizkan', 'deal', 'worth', '\$', '2.15', 'billion', '.', 'On', '10', u'juli', '2014', '.', u'unilev', u'announc', 'sold', 'slim-fast', 'brand', u'kaino', u'capit', '(', 'yet', u'retain', u'minor', 'stake', u'busi', '.', 'On', '2', u'decemb', '2014', '(', u'unilev', u'announc', u'acquir', 'talenti', 'gelato', '&', 'sorbetto', ':', u'minneapolis-bas', 'talenti', '(', u'found', '2003', '(', 'grown', u'best-sel', u'packag', 'gelato', u'unit', u'state', '.', 'On', '22', u'decemb', '2014', '(', u'unilev', u'announc', u'purchas', 'camay', 'brand', u'global', 'zest', 'brand', u'outsid', 'north', 'america', 'caribbean', 'procter', '&', u'gambl', '.', '====', 'hampton', 'creek', 'lawsuit', '====', 'In', u'novemb', '2014', '(', u'unilev', 'subject', 'media', 'backlash', 'due', 'lawsuit', 'rival', 'hampton', 'creek', '.', 'In', 'suit', '(', u'unilev', u'reveal', 'hampton', 'creek', '(', u'seiz', 'market', 'share', '"', u'loss', u'caus', u'unilev', '(', u'irrepar', 'harm', '.', '"', u'unilev', u'use', 'standard', u'ident', u'regul', u'claim', 'hampton', 'creek', "s", '(', 'just', 'mayo', '"', u'product', u'fals', u'advertis', "n't", 'contain', u'egg', '.', 'the', 'washington', 'post', u'headlin', 'suit', 'read', '(', 'big', 'food', "s", 'weird', 'war', 'over', 'the', u'mean', u'mayonnais', '.', '"', 'the', u'lo', u'angel', u'time', 'began', u'stori', '(', 'big', 'tobacco', '(', 'big', 'oil', '(', 'big', 'mayo', '?', '"', 'A', 'wall', 'street', 'journal', 'writer', u'describ', '(', 'giant', u'corpor', u'gener', 'huge', u'quantit', 'free', u'advertis', 'brand', u'equiti', u'tini', 'rival', u'su',

'.', '""', 'eat', 'drink', u'polit', u'headlin', u'controversi', '``',
 u'unilev', "'s", u'bulli', u'backfir', ' ', u'boost', 'hampton', 'cree
 k', '""', '.', u'neg', 'media', u'coverag', 'big', 'mayo', 'lawsuit',
 u'goe', 'viral', 'case', u'studi', 'PR', 'blunder', '""', '.', '==',
 '2015present', '==', 'In', 'march', '2015', ' ', u'unilev', u'confir
 m', u'reach', 'agreement', u'acquir', 'ren', u'skincar', ' ', 'britis
 h', u'nich', u'skincar', 'brand', '.', u'thi', u'follow', 'may', '201
 5', u'acquisit', u'prestig', u'skincar', 'brand', 'kate', u'somervil',
 u'skincar', 'llc', '.', 'In', u'juli', '2015', ' ', u'compani', u'sepa
 r', u'spread', u'busi', ' ', u'includ', 'flora', 'I', 'Ca', "n't", u'be
 liev', 'It', "'s", 'not', 'butter', '!', u'brand', ' ', u'standalon',
 u'entiti', u'name', u'unilev', u'bake', ' ', u'cook', u'spread', '.',
 'the', u'separ', 'first', u'announc', u'decemb', '2014', 'made', u'res
 pons', u'declin', u'worldwid', u'sale', 'product', u'categori', '.', 'I
 n', u'octob', '2015', ' ', u'unilev', u'agre', u'acquir', 'italian', 'p
 remium', 'ice', 'cream', 'maker', 'grom', u'undisclos', 'fee', '.', 'I
 n', u'juli', '2016', ' ', u'unilev', 'bought', 'US', 'start-up', 'dolla
 r', 'shave', 'club', u'report', '\$', '1bn', '(', '764m', ')', 'cash',
 u'compet', 'male', u'groom', 'market', '.', 'In', u'septemb', u'unile
 v', u'acquir', 'seventh', u'gener', 'inc.', '\$', '700', 'million', '.',
 'On', u'februari', '17', ' ', '2017', ' ', u'significantli', 'smaller',
 'kraft', 'heinz', 'made', '\$', '143', 'billion', 'bid', 'food', u'consu
 m', u'product', 'giant', u'unilev', 'the', 'deal', u'declin', u'unile
 v', u'abandon', u'februari', '19', 'UK', 'prime', u'minist', 'theresa',
 'may', u'order', u'scrutini', 'deal', '.', '==', u'oper', '==', u'unile
 v', u'organis', 'four', 'main', u'divis', ':', u'person', 'care', '(',
 u'product', 'skin', 'hair', u'product', ' ', u'deodor', 'oral', u'prod
 uct', ')', ';', u'food', '(', u'product', u'soup', ' ', u'bouillon',
 ' ', u'sauc', ' ', u'snack', ' ', u'mayonnais', ' ', 'salad', u'dres
 s', ' ', u'margarin', u'spread', ')', ';', u'refresh', '(', u'product',
 'ice', 'cream', ' ', u'tea-bas', u'beverag', ' ', u'weight-manag', u'pr
 oduct', u'nutrit', u'enhanc', u'stapl', 'sold', u'develop', u'market',
 ')', ';', 'home', 'care', '(', u'product', 'home', u'product', u'inclu
 d', u'powder', ' ', u'liquid', u'capsul', ' ', 'soap', u'bar', u'clea
 n', u'product', ')', '.', 'In', u'financi', 'year', u'end', '31', u'dec
 emb', '2013', ' ', u'unilev', 'total', u'turnov', '49.797', 'billion',
 '36', '%', u'person', 'care', ' ', '27', '%', u'food', ' ', '19', '%',
 u'refresh', '18', '%', 'home', 'care', '.', u'unilev', u'invest', 'tota
 l', '1.04', 'billion', 'research', u'develop', '2013', '.', u'unilev',
 'one', 'largest', 'media', u'buyer', 'world', ' ', u'invest', 'aroun
 d', '6', 'billion', '(', 'US', '\$', '8', 'billion', ')', u'advertis',
 u'promot', '2010', '.', u'unilev', "'s", 'largest', u'intern', u'compe
 titor', 'nestl', 'procter', '&', u'gambl', '.', 'It', u'face', u'compet
 it', 'local', u'market', u'specif', 'product', u'rang', u'numer', u'com
 pani', ' ', u'includ', 'beiersdorf', ' ', 'conagra', ' ', u'danon',
 ' ', 'henkel', ' ', u'mar', ' ', 'pepsico', ' ', 'reckitt', u'bencki
 s', 'S.', 'C.', 'johnson', '&', 'son', '.', u'unilev', u'fine', 'autori
 t', u'concurr', u'franc', '2016', u'price-fix', u'person', u'hygien',
 u'product', '.', '==', u'product', '==', u'unilev', "'s", u'product',
 u'includ', u'food', ' ', u'beverag', ' ', u'clean', u'agent', u'perso
 n', u'product', '.', 'the', u'compani', u'own', '400', u'brand', ' ',
 u'organis', 'four', 'main', u'categori', '-', u'food', ' ', u'refres
 h', ' ', 'home', 'care', ' ', u'person', 'care', '.', u'unilev', "'s",
 'current', u'largest-sel', u'brand', u'includ', ':', 'axe/lynx', ';',
 'ben', '&', u'jerri', "'s", ';', 'dove', ';', 'flora/becel', ';', 'hea
 rtbrand', ';', "hellmann's/best", u'food', ';', 'knorr', ';', 'lipton',
 ';', 'lux/radox', ';', 'omo/surf', ';', u'rexona/sur', ';', 'sunsilk',

';', 'tresemm', ';', 'magnum', ';', u'vaselin', 'vo5', '.', u'unilev',
 "s", 'standard', u'industri', u'classif', u'code', '10890', ':', u'man
 ufactur', 'food', u'product', 'n.e.c.', ',', '10410', ':', u'manufactu
 r', u'oil', u'fat', ',', '10420', ':', u'manufactur', u'margarin', 'sim
 ilar', u'edibl', u'fat', '.', '==', u'corpor', u'affair', '==', '===',
 'legal', u'structur', '===', u'unilev', 'two', u'hold', u'compani',
 ':', u'unilev', 'n.v.', ',', u'regist', 'head', u'offic', 'rotterdam',
 ',', u'netherland', u'unilev', 'plc', ',', u'regist', u'offic', 'port',
 'sunlight', u'merseysid', ',', u'unit', 'kingdom', 'head', u'offic',
 u'unilev', u'hous', 'london', ',', u'unit', 'kingdom', '.', u'unilev',
 'plc', u'unilev', 'n.v.', u'subsidiari', u'compani', u'oper', u'nearl
 i', u'practic', u'singl', u'econom', u'entiti', ',', 'whilst', u'remai
 n', u'separ', 'legal', u'entiti', u'differ', u'sharehold', u'separ', 's
 tock', u'exchang', u'list', '.', 'there', u'seri', 'legal', u'agreemen
 t', 'parent', u'compani', ',', u'togeth', 'special', u'provis', u'respe
 ct', u'articl', u'associ', ',', 'known', u'foundat', u'agreement', '.',
 'A', 'key', u'requir', u'agreement', u'peopl', u'board', 'two', 'paren
 t', u'compani', '.', 'An', u'equalis', 'agreement', u'regul', 'mutual',
 u'right', u'sharehold', u'unilev', 'plc', u'unilev', 'n.v.', u'object',
 u'ensur', ',', u'principl', ',', 'make', u'financi', u'differ', 'hold',
 u'share', u'unilev', 'plc', 'rather', u'unilev', 'n.v.', '(', 'vice',
 'versa', ')', '.', '===', 'senior', u'manag', '===', u'unilev', "s",
 'highest', u'execut', u'bodi', u'unilev', 'leadership', u'execut',
 ',', 'led', 'chief', u'execut', '(', u'current', 'paul', 'polman',
 ')', '.', '==', 'logo', '==', 'In', '1930', ',', 'logo', u'unilev',
 u'use', 'helvetica', 'font', u'cap', '.', 'In', '1969', ',', u'typefa
 c', 'logo', u'chang', ',', 'basic', 'design', u'remain', '.', 'the', 'c
 urrent', u'unilev', u'corpor', 'logo', u'introduc', '2004', u'design',
 'wolff', u'olin', ',', 'brand', u'consult', u'agenc', '.', 'the',
 "U", "", 'shape', 'made', '25', 'distinct', u'symbol', ',', 'icon',
 u'repres', 'one', u'compani', "s", u'sub-brand', u'corpor', u'valu',
 '.', 'the', 'brand', u'ident', u'develop', 'around', 'idea', '``', u'a
 d', u'vital', 'life', '.', '``', '==', u'advertis', '==', 'dove', ':',
 'dove', u'describ', u'dedic', '``', 'help', '...', 'women', 'develop',
 u'posit', 'relationship', 'way', 'look', u'help', u'rais', 'self-estee
 m', u'realiz', 'full', u'potenti', '``', '.', '(', 'dove', ',', '``',
 'our', 'vision', '``', ')', 'dove', u'employ', 'use', u'advertis', u'p
 roduct', 'display', u'messag', u'posit', 'self-esteem', '.', 'In', u'se
 ptemb', '2004', 'dove', u'creat', 'real', u'beauti', 'campaign', ',',
 u'focus', u'predomin', 'women', u'shape', 'colour', '.', 'later', '200
 7', 'campaign', u'further', u'includ', 'women', u'age', '.', u'thi', 'c
 ampaign', u'consist', u'mostli', u'advertis', ',', 'shown', u'televis',
 u'popularis', 'internet', '.', 'dove', 'fell', u'scrutini', u'gener',
 'public', 'felt', 'dove', u'advertis', u'describ', 'opinion', u'cellul
 it', 'still', u'unsightli', ',', 'women', "s", u'age', u'process', u's
 ometh', u'asham', '.', u'lynx/ax', ':', 'axe', ',', 'known', 'lynx',
 u'unit', 'kingdom', ',', u'republ', 'ireland', ',', 'australia', 'ne
 w', 'zealand', ',', u'toiletri', 'brand', u'market', u'toward', 'youn
 g', 'age', '16', '24', '.', u'it', u'market', '``', 'tongue-in-cheek',
 'take', u"mate", 'game', '``', '``', ',', u'suggest', 'women', u'insta
 ntli', 'drawn', u'use', u'product', '.', u'unlik', 'dove', "s", 'lon
 g', u'run', u'beauti', 'campaign', 'lynx', u'advertis', 'often', u'crea
 t', 'mini', u'seri', '``', u'advertis', u'base', 'around', 'singular',
 'product', 'rather', u'commun', u'overarch', 'idea', '.', u'thi', u'ad
 vertis', 'campaign', u'thrive', u'controversi', '.', u'use', u'imag',
 u'compani', u'know', u'receiv', u'complaint', u'garner', 'brand', 'fre
 e', u'public', u'notoriet', '.', 'A', 'wide', u'variet', u'advert',

u'ban', u'countri', 'around', 'world', '.', 'In', '2012', 'lynx',
 "'s", "'clean", u'ball', "", 'advert', u'ban', '.', u'thi', 'advert',
 u'design', u'televis', ',', u'show', u'attract', 'young', 'woman', u'cl
 ean', u'variou', 'sport', u'ball', '.', 'In', '2011', 'UK', 'lynx',
 "'s", 'shower', 'gel', 'campaign', u'ban', '.', 'the', 'poster', 'lyn
 x', 'shower', 'gel', u'show', 'woman', u'undon', 'bikini', 'shower', 'b
 each', ',', u'headlin', ':', '``', 'the', 'cleaner', 'dirtier', 'get',
 '.', '``', 'both', u'advertis', u'campaign', 'make', 'stark', u'compar
 ison', 'women', u'sexual', u'portray', u'advertis', u'sale', u'effici',
 '.', 'lynx', u'commonli', u'portray', 'women', 'visual', u'advertis',
 'hyper', 'sexual', ',', u'flawless', u'stereotyp', u'attract', u'arou
 s', ',', u'age', u'satur', ',', 'use', 'lynx', 'product', '.', 'thei
 r', 'target', u'audienc', u'age', '16-24', u'singl', '.', '==', u'envir
 onment', 'record', '==', u'unilev', u'declar', 'goal', u'decoupl', u'en
 vironment', 'impact', 'growth', ',', ':', u'halv', u'environment', 'foo
 tprint', u'product', 'next', '10', u'year', ';', u'help', '1', 'billio
 n', u'peopl', u'improv', 'health', u'well-b', ';', u'sourc', u'agricult
 ur', 'raw', u'materi', u'sustain', '.', '===', 'palm', 'oil', '===',
 u'unilev', u'criticis', u'greenpeac', u'caus', u'deforest', ',', u'uni
 lev', u'target', '2008', u'greenpeac', 'UK', ',', u'criticis', u'compan
 i', u'buy', 'palm', 'oil', u'supplier', u'damag', 'indonesia', "'s",
 u'rainforest', '.', 'By', '2008', ',', 'indonesia', u'lose', '2', '%',
 u'remain', 'rainforest', 'year', ',', 'fastest', u'deforest', 'rate',
 u'countri', '.', 'the', u'unit', u'nation', u'environment', u'program
 m', u'state', 'palm', 'oil', u'plantat', u'lead', u'caus', u'deforest',
 'indonesia', '.', u'furthermor', ',', 'indonesia', 'fourteenth', 'large
 st', u'emitt', u'greenhous', u'gase', u'larg', 'due', u'destruct', u'ra
 inforest', 'palm', 'oil', u'industri', ',', u'contribut', '4', '%', 'gl
 obal', 'green', u'hous', u'ga', u'emiss', '.', u'accord', u'greenpeac',
 ',', 'palm', 'oil', u'expans', u'take', 'place', u'litl', 'oversight',
 'central', 'local', u'govern', u'procedur', u'environment', 'impact',
 u'assess', ',', u'land-us', u'plan', u'ensur', 'proper', u'process',
 u'develop', u'concess', u'neglect', '.', u'plantat', u'off-limit',
 ',', 'law', ',', 'palm', 'oil', u'plantat', u'establish', 'well', u'il
 leg', 'use', 'fire', 'clear', 'forest', u'area', u'commonplac', '.',
 u'unilev', ',', u'found', 'member', u'roundtabl', u'sustain', 'palm',
 'oil', '(', 'rspo', ')', ',', u'respond', u'publicis', 'plan', 'obtai
 n', 'palm', 'oil', u'sourc', u'certifi', u'sustain', '2015', '.', 'It',
 u'claim', 'goal', '2012', u'encourag', 'rest', u'industri', u'becom',
 '100', '%', u'sustain', '2020', '.', 'In', 'cte', u"d'ivoir", '(', u'i
 vori', 'coast', ')', ',', 'one', u'unilev', "'s", 'palm', 'oil', u'supp
 lier', u'accus', u'clear', 'forest', u'plantat', ',', u'activ', u'threa
 ten', u'primat', u'speci', ',', u'miss', 'waldron', "'s", 'red', u'colo
 bu', '.', u'unilev', u'interven', 'halt', u'clearanc', u'pend', u'resul
 t', u'environment', u'assess', '.', u'accord', u'amnesti', u'intern',
 '2016', u'unilev', 'palm', 'oil', u'provid', 'wilmar', u'intern', u'pr
 ofit', '8', '14', 'year', 'old', 'child', 'labor', u'forc', 'labor',
 '.', 'some', u'worker', u'extort', ',', u'threaten', 'paid', 'work',
 '.', 'some', u'worker', u'suffer', u'sever', u'injuri', 'toxic', u'ba
 n', u'chemic', '.', 'In', '2016', u'singapore-bas', 'wilmar', u'inter
 n', u'world', 'biggest', 'palm', 'oil', 'grower', '.', '===', 'paper',
 'use', '===', 'for', u'year', ',', u'unilev', u'purchas', 'paper', u'p
 ackag', 'asia', 'pulp', '&', 'paper', ',', 'third', 'largest', 'paper',
 u'produc', 'world', ',', u'label', '``', 'forest', u'crimin', "", u'd
 estroy', '``', u'preciou', 'habitat', "", 'indonesia', "'s", 'rainfor
 est', '.', 'In', '2011', ',', u'unilev', u'cancel', 'contract', 'asia',
 'pulp', '&', 'paper', ',', u'greenpeac', u'execut', 'director', 'phil',

'radford', u'commend', u'compani', u'effort', 'made', u'toward', 'fores
t', u'protect', ' ', ' ', ' ', u'take', 'rainforest', u'conserv', u'serious',
' ', ' ', ' ', ' ', ' ', 'rainforest', u'allianc', ' ', ' ', u'unilev', u'cer
tifi', 'tea', u'product', 'rainforest', u'allianc', 'scheme', ' ', 'th
e', u'compani', u'state', 'least', '50', '%', 'tea', u'product', u'orig
in', u'certifi', u'farm', ' ', ' ', u'compar', u'allianc', 's', '30', '%',
'minimum', u'entri', 'point', ' ', ' ', u'unilev', u'decid', 'scheme', u'fa
irtrad', ' ', ' ', u'accord', u'compani', 's', u'analysi', ' ', ' ', u'fairtra
d', 'might', ' ', ' ', 'lack', 'scale', u'organiz', u'flexibl', u'certifi',
u'industri', 'tea', u'estat', ' ', ' ', ' ', ' ', ' ', u'critic', ' ', ' ', 'th
e', 'rainforest', u'allianc', u'certif', 'scheme', u'criticis', u'offe
r', u'produc', 'minimum', u'guarante', 'price', ' ', ' ', u'therefor', u'lea
v', u'vulner', 'market', 'price', u'variati', ' ', 'the', u'altern', u'c
ertif', ' ', ' ', u'fairtrad', ' ', ' ', u'howev', u'receiv', 'similar', u'criti
c', 'well', ' ', 'the', 'rainforest', u'allianc', u'certif', u'furthem
or', u'criticis', u'allow', 'use', 'seal', u'product', 'contain', 'mini
mum', '30', '%', u'certifi', 'content', ' ', ' ', u'accord', u'endang', u'in
tegr', u'certif', ' ', ' ', ' ', u'juli', '-', u'septemb', '2016', 'salmon
ella', 'affair', ' ', ' ', ' ', ' ', 'salmonella', 'affair', u'cereal', 'isr
ael', ' ', ' ', 'In', u'juli', '2016', ' ', ' ', u'rumour', 'salmonella', u'co
ntamin', u'cereal', 'spread', 'among', u'isra', u'consum', ' ', u'init
i', ' ', ' ', u'unilev', u'provid', 'public', u'inform', 'subject', u'quer
i', 'matter', u'initi', u'rebuf', u'compani', u'non-stori', u'nonsens',
' ', 'On', 'night', '26', u'juli', '2016', ' ', ' ', u'unilev', u'stop', u't
ransfer', u'cornflak', u'retail', u'chain', ' ', 'On', '28', u'juli',
' ', 'yediote', 'ahronot', u'report', u'ten', u'thousand', u'box', 'bre
akfast', 'cereal', u'destroy', ' ', 'By', '28', u'juli', ' ', ' ', u'despi
t', u'compani', 's', u'assur', u'noth', u'contamin', u'releas', u'cons
umpt', ' ', ' ', u'mani', u'custom', u'stop', u'buy', u'unilev', u'product',
u'start', 'throw', 'away', u'cornflak', 'made', u'unilev', ' ', 'the',
u'compani', 'withheld', u'inform', u'affect', u'product', u'date',
' ', 'On', '2', 'august', '2016', ' ', ' ', u'globe', u'report', u'compan
i', u'publish', u'inform', 'telma', u'cereal', u'handl', u'packag', 'li
ne', u'contamin', u'discov', 'telma', u'announc', 'made', ':', ' ', 'W
e', u'stress', 'telma', u'product', u'store', u'home', 'safe', 'eat',
' ', u'accord', u'compani', 's', 'strict', u'procedur', ' ', ' ', u'ever
i', u'product', 'batch', u'check', 'put', 'hold', ' ', 'these', u'produ
ct', u'market', 'test', u'result', u'product', u'seri', u'return', ' ',
u'confirm', 'well', ' ', 'If', 'flaw', u'discov', ' ', ' ', 'batch', u'marke
t', u'store', ' ', ' ', 'case', ' ', ' ', 'In', u'follow', u'day', 'healt
h', u'minist', ' ', ' ', 'yakov', 'litzman', ' ', ' ', u'threaten', 'pull', u'uni
lev', 's', u'licenc', 'israel', ' ', 'He', u'accus', u'unilev', u'li
e', u'ministri', u'regard', u'salmonella-infect', 'breakfast', u'cerea
l', ' ', 'On', '7', 'august', ' ', ' ', u'globe', u'report', u'contamin', 'm
ay', u'sourc', 'pigeon', u'faec', ' ', ' ', 'health', u'ministri', 'said',
'might', u'sourc', u'contamin', 'pigeon', u'faec', u'possibl', u'sour
c', ' ', ' ', u'globe', 'said', u'product', 'line', u'automat', '(', ' ',
'without', 'human', u'hand', ' ', ' ', ' ', u'possibl', u'sourc', 'human',
'slim', u'chanc', ' ', 'On', '8', 'august', '2016', ' ', ' ', u'isra', 'heal
th', u'minist', u'suspend', u'manufactur', u'licenc', u'unilev', u'carr
i', 'number', u'correct', ';', 'action', 'came', u'inspect', 'arad', 'p
lant', ' ', ' ', u'state', ' ', ' ', u'thi', u'seri', u'neglig', u'mistak', ' ',
u'incid', u'malici', 'intent', 'firm', 's', u'manag', u'qualiti', 'con
trol', u'procedur', ' ', ' ', ' ', 'An', u'investig', 'led', 'prof.', 'itam
r', 'grutto', 'eli', 'gordon', u'conclud', 'event', u'caus', u'neglig',
' ', 'On', '23', u'septemb', u'report', u'cereal', u'produc', '18th',
'20th', 'arad', 'plant', u'trace', 'salmonella', ' ', ' ', ' ', u'clas

```
s', u'action', '====', 'A', u'file', u'class', 'action', 'must', 'firs
t', u'approv', u'isra', 'court', '', u'approv', 'case', 'held', '.',
'for', 'sum', '1.2', 'million', u'ni', '(', '~', '$', '329k', 'usd',
')', u'unilev', u'hide', u'contamin', u'mislead', 'public', 'for', 'su
m', '76', 'million', u'ni', '(', '~', '$', '23m', 'usd', ')', u'unile
v', '15-year-old', 'teen', u'hospitalis', u'salmonellosi', u'allegedl
i', u'contract', u'unilev', u'product', '====', 'salmonella', 'affair',
'tehnia', '====', 'On', '31', 'august', '', u'unilev', u'state', 'tehn
ia', u'product', u'produc', 'rjm', u'contamin', 'salmonella', '.', '==
=', u'kodaikan', '===', 'In', '2015', '', 'indian', 'rapper', 'sofia',
'ashraf', u'releas', 'music', 'video', '`', u'kodaikan', 'Wo', "n't",
'', "'", 'set', 'beat', 'nicki', 'minaj', "s", '`', 'anaconda',
'', "'", u'criticis', u'unilev', u'dump', u'mercuri', u'wast', 'grou
nd', 'indian', 'town', u'kodaikan', '.', u'unilev', u'acknowledg', 'too
k', u'thermet', u'factori', u'acquisit', u'chesebrough-pond', '.',
u'accord', u'unilev', "s", 'statement', '', u'factori', 'sold', u'me
rcury-contamin', 'scrap', u'glass', 'local', 'dealer', '', u'prompt',
'hindustan', u'unilev', u'immedi', 'close', u'factori', '', 'plan',
'clean-up', u'affect', u'site', 'monitor', 'health', u'worker', '.',
u'unilev', "s", u'websit', u'state', u'wait', u'sinc', '2010', 'loca
l', u'govern', 'tamil', 'nadu', u'pollut', 'control', 'board', 'give',
u'permiss', 'clean-up', '.', '==', 'see', '==', 'list', 'food', u'comp
ani', '==', u'refer', '==', '==', u'extern', u'link', '==', u'offici',
u'websit']
```

Put the cleaned word into a string

```
In [108]: w_str_list=[]
          for p in after_stem_list:
              word_str = ''
              for i in p:
                  if (i!=p[0]):
                      word_str += ' '
                      word_str += i
              w_str_list.append(word_str)
```

```
In [109]: import csv
          f= open("wiki_result.csv","w")
          writer = csv.writer(f, delimiter='|')
          for i in w_str_list:
              p = ['0',i]
              writer.writerow(p)
          f.close()
```

Find a different sort of wiki pages

```
In [98]: t = wikipedia.search('asia',results=30)
plist = []
for p in t:
    print p
    if p == "XXX": continue
    page = wikipedia.page(p, auto_suggest=False)
    plist.append(page)

tokenized_docs_list = []

for i in plist:
    tokenized_docs = word_tokenize(i.content.encode('ascii','ignore'))
    tokenized_docs_list.append(tokenized_docs)
#-----#
t_nsw_list = []
for p in tokenized_docs_list:
    tokenized_nsw = []
    for i in p:
        if not i in stopwords.words():
            tokenized_nsw.append(i)
    t_nsw_list.append(tokenized_nsw)
#-----#
after_stem_list = []
for p in t_nsw_list:
    words = []
    for i in p:
        words.append(porter.stem(i))
    after_stem_list.append(words)
#-----#
w_str_list2 = []
for p in after_stem_list:
    word_str = ''
    for i in p:
        if (i!=p[0]):
            word_str += ' '
        word_str += i
    w_str_list2.append(word_str)
#-----#
```

Asia
 Unilever
 Eurasia
 Eric Hoffer
 Shamanism
 Asian giant hornet
 List of Asian pornographic actors
 Georgia (country)
 Calligraphy
 Hornet
 Asia Argento
 Time (magazine)
 Pacific War
 Boundaries between the continents of Earth
 Demographics of Russia
 List of most common surnames in Asia
 Humid subtropical climate
 Aedes albopictus
 Buddhism
 Lists of World Heritage Sites
 XXX
 Shanghai SIPG F.C.
 Alcohol flush reaction
 Asian Boyz
 KLM
 XXX (Asia album)
 Pre-Indo-European languages
 Tatars
 Diospyros kaki
 Indian subcontinent

```

In [110]: f= open("wiki_result.csv","a")
writer = csv.writer(f, delimiter='|')
for i in w_str_list2:
    p = ['1',i]
    writer.writerow(p)
f.close()
  
```

In [96]:

```

[u'Asia', u'Unilever', u'Eurasia', u'Eric Hoffer', u'Shamanism', u'Asia
n giant hornet', u'List of Asian pornographic actors', u'Georgia (count
ry)', u'Calligraphy', u'Hornet', u'Asia Argento', u'Time (magazine)',
u'Pacific War', u'Boundaries between the continents of Earth', u'Demog
raphics of Russia', u'List of most common surnames in Asia', u'Humid su
btropical climate', u'Aedes albopictus', u'Buddhism', u'Lists of World
Heritage Sites', u'XXX', u'Shanghai SIPG F.C.', u'Alcohol flush reacti
on', u'Asian Boyz', u'KLM', u'XXX (Asia album)', u'Pre-Indo-European la
nguages', u'Tatars', u'Diospyros kaki', u'Indian subcontinent']
  
```