# Experiment for Extracting Multiple Dimensions from Texts

Bing Ma and Hai Zhuge

## Purpose

The purpose of this experiment is to extract multiple classification dimensions on a set of texts.

## Input

The input is a set of texts from DUC2002 dataset (available at the folder named “Input Texts”), which consists of the following two classes:

1. The earthquake class (the IDs of the texts are a0, a1, a2, a3, a4, a5, a6, a7, a8, a9, a10, a11). It contains the Iran Earthquake class (a0, a1, a2, a3), the Northern California Earthquake class (a4, a5, a6, a7) and the Himalayan Earthquake class (a8, a9, a10, a11).
2. The hurricane class (the IDs of the texts are b0, b1, b2, b3, b4, b5, b6, b7, b8). It contains the Hurricane Andrew class (b0, b1, b2), the Hurricane Gilbert class (b3, b4, b5) and the Hurricane Hugo class (b6, b7, b8).

## Method

We use the bottom-up text clustering approach proposed in the paper “Automatic Construction of Classification Dimensions on Texts” to construct the classification dimensions on the input based on common words between texts. The common words on the currently extracted classification dimension are deleted from the input texts to extract the next dimension.

## Output

The output are classification dimensions consisting of class trees. Each class in the class trees labelled by their common words contains a set of texts.

The results show that (1) the first dimension is the major dimension, and its structure fits the classification structure of the input texts, i.e., containing two class trees: earthquake class and hurricane class; (2) the second dimension contains seven class trees, which do not contain the two classes of the first dimension because the words representing them are removed; and, (3) the number of class trees of the next dimension tends to increase and the height of the class trees tends to decrease because with the next dimension has fewer common words between texts.

The following is the extracted dimensions.

### Dimension 1 (Dim1)

Number of class trees: 2

#### Tree1:

Root-common words: ['hurricane', 'storm', 'damage', 'coast', 'official']

Root-texts: ['b1', 'b2', 'b0', 'b5', 'b4', 'b3', 'b8', 'b7', 'b6']

subclass1-common words: ['hurricane', 'storm', 'people', 'damage', 'coast', 'official', 'area', 'field']

subclass1-texts: ['b1', 'b2', 'b0', 'b5', 'b4', 'b3', 'b8', 'b7']

subclass11-common words: ['hurricane', 'storm', 'people', 'damage', 'florida', 'coast', 'official', 'area', 'field']

subclass11-texts: ['b1', 'b2', 'b0', 'b5', 'b4', 'b3', 'b8']

subclass111-common words: ['hurricane', 'people', 'storm', 'damage', 'florida', 'coast', 'mexico', 'official', 'area', 'field']

subclass111-texts: ['b1', 'b2', 'b0', 'b5', 'b4', 'b3']

subclass1111-common words: ['hurricane', 'people', 'damage', 'florida', 'storm', 'coast', 'new orleans', 'miami', 'mexico', 'official', 'gulf', 'area', 'field', 'orleans']

subclass1111-texts: ['b1', 'b2', 'b0', 'b5', 'b4']

subclass11111-common words: ['hurricane', 'florida', 'damage', 'people', 'storm', 'coast', 'landfall', 'new orleans', 'miami', 'gulf', 'mexico', 'area', 'field', 'official', 'orleans']

subclass11111-texts: ['b1', 'b2', 'b0', 'b5']

subclass111111-common words: ['damage', 'hurricane', 'loss', 'industry', 'louisiana', 'florida', 'yesterday', 'storm', 'landfall', 'new orleans', 'insurer', 'underwriter', 'night', 'market', 'miami', 'hurricane andrew', 'death', 'coast', 'bahamas', 'people', 'southern florida', 'gulf', 'bn', 'mexico', 'andrew', 'official', 'area', 'field', 'hugo', 'south carolina', 'orleans', 'carolina']

subclass111111-texts: ['b1', 'b2', 'b0']

subclass1111111-common words: ['insurance', 'hurricane', 'loss', 'claim', 'damage', 'florida', 'louisiana', 'industry', 'reinsurance', 'storm', 'texas', 'yesterday', 'death', 'company', 'hurricane andrew', 'market', 'new orleans', 'insurer', 'underwriter', 'group', 'gulf', 'landfall', 'bahamas', 'southern florida', 'last night', 'people', 'mexico', 'cent', 'miami', 'balcombe group', 'bn', 'oil refining industry', 'andrew', 'power', 'brunt', 'property damage', 'industry analyst', 'level', 'area', 'field', 'gulf coast', 'night', 'wind', 'hugo', 'south carolina', 'year', 'insurance analyst', 'coast', 'official', 'household', 'home', 'sunday', 'century', 'hundred', 'one', 'analyst', 'carolina', 'orleans']

subclass1111111-texts: ['b0', 'b2']

subclass1111112-common words: ['damage', 'louisiana', 'mile', 'industry', 'florida', 'hurricane', 'storm', 'yesterday', 'landfall', 'loss', 'new orleans', 'centre', 'city', 'exposure', 'miami', 'hurricane andrew', 'andrew', 'north-west', 'market', 'homestead', '2m people', 'dollars 42 bn', 'official', 'havoc', 'coast', 'night', 'insurer', 'underwriter', 'plant', 'bahamas', 'southern florida', 'bill', 'area', 'field', 'monday', 'death', 'gulf', 'mexico', 'hour', 'hugo', 'south carolina', 'people', 'much', 'orleans', 'carolina', 'bn']

subclass1111112-texts: ['b1', 'b2']

subclass2-common words: ['hurricane', 'storm', 'wind', 'coast', 'damage', 'official']

subclass2-texts: ['b6', 'b7', 'b8', 'b3', 'b5', 'b4', 'b2', 'b0']

subclass21-common words: ['storm', 'hurricane', 'wind', 'coast', 'damage', 'service', 'official']

subclass21-texts: ['b6', 'b7', 'b8', 'b3', 'b5', 'b4', 'b2']

subclass211-common words: ['storm', 'hurricane', 'wind', 'coast', 'official', 'service', 'damage']

subclass211-texts: ['b6', 'b7', 'b8', 'b3', 'b5', 'b4']

subclass2111-common words: ['storm', 'hurricane', 'wind', 'coast', 'official', 'service', 'damage']

subclass2111-texts: ['b6', 'b7', 'b8', 'b3', 'b5']

subclass21111-common words: ['storm', 'hurricane', 'wind', 'coast', 'official', 'service', 'damage']

subclass21111-texts: ['b6', 'b7', 'b8', 'b3']

subclass211111-common words: ['storm', 'riley', 'hurricane', 'charleston', 'wind', 'service', 'fury', 'official', 'coast', 'mayor joseph p riley jr', 'number', 'damage', 'ten', 'decade', 'time', 'jr']

subclass211111-texts: ['b6', 'b7', 'b8']

subclass2111111-common words: ['riley', 'hugo', 'hurricane', 'city', 'city hall', 'storm', 'wind', 'tin roof', 'charleston', '135 mph', 'mph', 'eye', 'center', 'coast', 'chamber', 'city command post', 'electric service', 'loss', 'reporter', 'fury', 'mayor joseph p riley jr', 'damage', 'speed', 'official', 'phone', 'street', 'water', 'ten', 'decade', 'number', 'strength', 'day', 'service', 'thursday', 'night', 'command post', 'hall', 'one', 'post', 'spot', 'place', 'roof', 'time', 'jr']

subclass2111111-texts: ['b6', 'b7']

subclass2111112-common words: ['storm', 'charleston', 'hurricane', 'warning', 'wind', 'riley', 'shelter', 'thousand', 'south carolina', 'fury', 'official', 'number', 'life', 'occupant', 'resident', 'people', 'mayor joseph p riley jr', 'coast', 'columbia', 'service', 'forecaster', 'ten', 'decade', 'area', 'field', 'damage', 'time', 'carolina', 'jr']

subclass2111112-texts: ['b7', 'b8']

subclass2112-common words: ['people', 'hurricane', 'cancun', 'storm', 'jamaica', 'coast', 'mile', 'wind', 'mph', 'gilbert', 'city', 'hotel', 'cozumel', 'mexico', 'today', 'area', 'field', 'island', 'resort', 'inch', 'official', 'sheets', 'damage', 'florida', 'dominican republic', 'zone', 'north', 'national weather service', 'telephone', 'maximum sustained wind', 'report', 'edward seaga', '1935 storm', '256 people', 'two category 5 hurricane', 'spokesman', 'communication', '408 people', 'mississippi coast', 'catastrophic damage', 'civil defense official', 'hurricane camille', 'hurricane gilbert', 'tree', 'intense storm', 'hour', '1935 hurricane', '2635 inch', 'barometric pressure', 'florida keys', '15 mph', 'national hurricane center', 'type', 'thousand', 'strong', 'hurricane center', 'deadly', 'pressure', 'morning', 'eye', 'center', 'term', 'one', 'sustained wind', 'united states', 'service', 'book', 'record', 'seaga', 'weather service', 'republic', 'keys', 'states', 'camille']

subclass2112-texts: ['b3', 'b5', 'b4']

subclass21121-common words: ['people', 'cancun', 'hurricane', 'storm', 'mile', 'merida', 'mph', 'today', 'gilbert', 'jamaica', 'tourist', 'mexico', 'resort', 'mexico city', 'island', 'cozumel', 'wind', 'coast', 'peninsula', 'hotel', 'texas', '160 mph wind', 'yucatan coast', 'sheets', 'rain', 'puerto progreso', 'north', 'florida', 'west', 'damage', 'city', 'official', 'zone', 'coastal area', 'capital', 'shelter', 'supply', 'national weather service', 'dominican republic', 'telephone', 'area', 'field', '2631 inch', 'edward seaga', '1935 storm', '256 people', 'two category 5 hurricane', '10 inch', '5 storm', '408 people', 'mississippi coast', 'catastrophic damage', 'civil defense official', 'hurricane camille', 'hurricane center', 'national hurricane center', 'gulf coast', 'intense storm', '1935 hurricane', '2635 inch', 'barometric pressure', 'florida keys', 'thousand', 'spokesman', 'communication', '15 mph', 'maximum sustained wind', 'hour', 'sheet', 'tree', 'landfall', 'type', 'one', 'sustained wind', 'hurricane gilbert', 'electricity', 'strong', 'deadly', 'report', 'town', 'township', 'morning', 'term', 'storm wind', 'pressure', 'eye', 'center', 'household', 'home', 'state', 'country', 'united states', 'book', 'record', 'weather service', 'month', 'inch', 'progreso', 'service', 'republic', 'keys', 'seaga', 'states', 'camille']

subclass21121-texts: ['b3', 'b5']

subclass21122-common words: ['cancun', 'hurricane', 'people', 'jamaica', 'mile', 'wind', 'coast', 'storm', 'cozumel', 'gilbert', 'hotel', 'edt', 'today', 'mph', 'tuesday', 'area', 'field', 'mexico', 'inch', 'rig', 'dominican republic', 'florida', '175 mph', 'resort', 'sheets', 'official', 'hour', 'maximum sustained wind', 'communication', 'city', 'six people', 'hotel zone', 'island', 'us national weather service', 'north', 'report', 'prime minister edward seaga', 'hurricane gilbert', 'five people', 'sunday', '256 people', 'two category 5 hurricane', '1935 storm', 'bad disaster', '408 people', 'civil defense official', 'catastrophic damage', 'mississippi coast', 'low pressure', 'oil company', 'hurricane camille', 'western hemisphere', 'gulf', 'telephone', 'intense storm', 'miami', '1935 hurricane', 'barometric pressure', '2635 inch', 'florida keys', 'mark zimmer', 'new orleans', 'figure', 'damage', 'tree', '15 mph', 'strong', 'director', 'modern history', 'type', 'thousand', 'spokesman', 'eye', 'center', 'deadly', 'national hurricane center', 'forecaster', 'term', 'morning', 'worker', 'service', 'history', 'monday', 'seaga', 'hurricane center', 'united states', 'national weather service', 'weather service', 'book', 'record', 'occupant', 'resident', 'company', 'one', 'republic', 'pressure', 'disaster', 'states', 'camille', 'zone', 'keys', 'hemisphere', 'zimmer', 'orleans']

subclass21122-texts: ['b4', 'b5']

#### Tree2:

Root-common words: ['quake', 'earthquake', 'temblor']

Root-texts: ['a0', 'a2', 'a1', 'a3', 'a11', 'a8', 'a10', 'a9', 'a7', 'a5', 'a6', 'a4']

subclass1-common words: ['quake', 'earthquake', 'temblor']

subclass1-texts: ['a4', 'a5', 'a6', 'a7', 'a11', 'a9', 'a8', 'a10', 'a1', 'a2', 'a0']

subclass11-common words: ['quake', 'earthquake', 'temblor']

subclass11-texts: ['a4', 'a5', 'a6', 'a7', 'a11', 'a9', 'a8', 'a10', 'a1', 'a2']

subclass111-common words: ['quake', 'earthquake', 'temblor', 'area', 'field']

subclass111-texts: ['a4', 'a5', 'a6', 'a7', 'a11', 'a9', 'a8', 'a10', 'a1']

subclass1111-common words: ['quake', 'earthquake', 'temblor', 'area', 'field']

subclass1111-texts: ['a4', 'a5', 'a6', 'a7', 'a11', 'a9', 'a8', 'a10']

subclass11111-common words: ['quake', 'earthquake', 'temblor', 'damage', 'area', 'field']

subclass11111-texts: ['a4', 'a5', 'a6', 'a7', 'a11', 'a9', 'a8']

subclass111111-common words: ['quake', 'earthquake', 'temblor', 'damage', 'mile', 'report', 'area', 'field', 'building']

subclass111111-texts: ['a4', 'a5', 'a6', 'a7', 'a11', 'a9']

subclass1111111-common words: ['quake', 'earthquake', 'temblor', 'damage', 'mile', 'magnitude', 'report', 'area', 'field', 'building']

subclass1111111-texts: ['a4', 'a5', 'a6', 'a7', 'a11']

subclass11111111-common words: ['damage', 'quake', 'earthquake', 'temblor', 'san francisco', 'mile', 'magnitude', 'major quake', 'francisco', 'building', 'report', 'area', 'field', 'bay bridge', 'time', 'bridge', 'major earthquake', 'major temblor']

subclass11111111-texts: ['a4', 'a5', 'a6', 'a7']

subclass111111111-common words: ['san francisco', 'mile', 'richter scale', 'damage', 'reading', 'major earthquake', 'quake', 'earthquake', 'temblor', 'magnitude', 'building', 'report', 'francisco', 'area', 'field', 'bay bridge', 'widespread heavy damage', 'severe damage', 'seismograph', 'bridge', 'ground motion', 'measure', 'time', 'widespread damage', 'heavy damage', 'scale', 'motion', 'major quake']

subclass111111111-texts: ['a4', 'a5', 'a6']

subclass1111111111-common words: ['san francisco', 'damage', 'richter scale', 'candlestick', 'mile', 'francisco', 'building', 'fan', 'upper deck', 'reading', 'quake', 'earthquake', 'temblor', 'world series', 'stadium', 'major earthquake', 'bridge', 'collins', 'level', 'candlestick park', 'report', '225 mile', '7 reading', 'san francisco bay bridge', 'televised footage', '30-foot section', 'san francisco giants', 'oakland athletics', 'tenfold increase', 'severe damage', 'widespread heavy damage', 'abc-tv', 'reno', 'system', 'one number', 'every increase', 'magnitude', 'nev', 'san francisco earthquake', 'phone', 'bureau', 'area', 'field', 'sacramento', 'seismograph', 'northeast', 'ground motion', '10 time', 'electricity', 'injury', 'measure', 'power', 'one', 'heavy damage', 'section', 'scale', 'series', 'increase', 'step-up', 'park', 'footage', 'deck', 'giants', 'athletics', 'motion', 'number', 'time']

subclass1111111111-texts: ['a4', 'a5']

subclass111111112-common words: ['quake', 'earthquake', 'temblor', 'magnitude', 'san francisco', 'major quake', 'report', 'california', 'damage', 'mile', 'area', 'field', 'santa cruz', 'building', 'francisco', 'cruz', 'bay bridge', 'northern california', 'bridge', 'time', 'eye', 'center', 'major earthquake', 'major temblor']

subclass111111112-texts: ['a6', 'a7', 'a5']

subclass1111111121-common words: ['quake', 'earthquake', 'temblor', 'magnitude', 'major quake', 'high probability', 'santa cruz', 'scientist', 'probability', 'bay area', 'tuesday quake', 'damage', 'fault', 'santa cruz mountain', '1988 report', 'california', 'san francisco', 'us geological survey', 'point', 'cruz', '1987 whittier quake', 'mile', 'strong quake', 'building', 'region', 'san andreas fault', 'tuesday', 'southern california', 'overpass', 'francisco', '65 magnitude', 'parkfield', 'area', 'field', 'survey', 'northern california', 'bay bridge', 'los angeles area', 'geological survey', 'report', 'time', 'eye', 'center', 'mountain', 'mountains', 'bridge', 'major earthquake', 'tuesday earthquake', 'major temblor', 'santa cruz mountains']

subclass1111111121-texts: ['a6', 'a7']

subclass2-common words: ['quake', 'earthquake', 'temblor']

subclass2-texts: ['a0', 'a2', 'a1', 'a3', 'a11', 'a8', 'a10', 'a9', 'a7', 'a5', 'a6']

subclass21-common words: ['people', 'quake', 'earthquake', 'temblor']

subclass21-texts: ['a0', 'a2', 'a1', 'a3', 'a11', 'a8', 'a10', 'a9', 'a7', 'a5']

subclass211-common words: ['people', 'quake', 'earthquake', 'temblor']

subclass211-texts: ['a0', 'a2', 'a1', 'a3', 'a11', 'a8', 'a10', 'a9', 'a7']

subclass2111-common words: ['people', 'quake', 'earthquake', 'temblor', 'state', 'country', 'agency', 'authority']

subclass2111-texts: ['a0', 'a2', 'a1', 'a3', 'a11', 'a8', 'a10', 'a9']

subclass21111-common words: ['people', 'quake', 'earthquake', 'temblor', 'government', 'state', 'country', 'agency', 'authority']

subclass21111-texts: ['a0', 'a2', 'a1', 'a3', 'a11', 'a8', 'a10']

subclass211111-common words: ['people', 'quake', 'earthquake', 'temblor', 'red cross', 'government', 'state', 'country', 'agency', 'authority', 'cross']

subclass211111-texts: ['a0', 'a2', 'a1', 'a3', 'a11', 'a8']

subclass2111111-common words: ['people', 'red cross', 'quake', 'earthquake', 'temblor', 'government', 'agency', 'authority', 'state', 'country', 'cross']

subclass2111111-texts: ['a0', 'a2', 'a1', 'a3', 'a11']

subclass21111111-common words: ['iran', 'people', 'red cross', 'united states', 'tehran', 'us government', 'quake', 'earthquake', 'temblor', 'agency', 'authority', 'state', 'country', 'government', 'cross', 'states']

subclass21111111-texts: ['a0', 'a2', 'a1', 'a3']

subclass211111111-common words: ['iran', 'organization', 'people', 'aid', 'assistance', 'red cross', 'tehran', 'victim', 'us government', 'quake', 'earthquake', 'temblor', 'agency', 'authority', 'official', 'state', 'country', 'united states', 'government', 'cross', 'today', 'states']

subclass211111111-texts: ['a1', 'a3', 'a0']

subclass2111111111-common words: ['iran', 'organization', 'plane', 'red cross', 'aid', 'assistance', 'medicine', 'supply', 'people', 'tehran', 'tent', 'iranians', 'medical supply', 'blanket', 'shipment', 'earthquake victim', 'mission', 'us government', 'us', 'development', 'quake', 'earthquake', 'temblor', 'official', 'us aid', 'new canaan', 'earthquake-stricken iran', 'food', 'pound', 'need', 'airport', 'agency', 'authority', 'victim', 'government', 'conn', 'state', 'country', 'morning', 'cross', 'support', 'year', 'united states', 'today', 'states', 'canaan']

subclass2111111111-texts: ['a1', 'a3']

subclass2111111112-common words: ['iran', 'people', 'aid', 'assistance', 'group', 'organization', 'rescuer', 'iranian red crescent', 'victim', 'official', 'american red cross', 'humanitarian organization', 'us government', 'red crescent', 'help', 'iranian agency', 'quake', 'earthquake', 'temblor', 'humanitarian group', 'relation', 'state department', 'washington', 'crescent', 'israel', 'red cross', 'regard', 'early thursday', 'tehran', 'agency', 'authority', 'united states', 'state', 'country', 'house', 'sign', 'death toll', 'today', 'thursday', 'cross', 'department', 'government', 'toll', 'states', 'iranian authority']

subclass2111111112-texts: ['a3', 'a0']

subclass211111112-common words: ['iran', 'people', 'united states', 'tehran', 'quake', 'earthquake', 'temblor', 'red cross', 'us government', 'hostage', 'agency', 'authority', 'embassy', 'state', 'country', 'government', 'cross', 'states']

subclass211111112-texts: ['a0', 'a2', 'a1']

subclass2111111121-common words: ['iran', 'people', 'united states', 'quake', 'earthquake', 'temblor', 'us government', 'offer', 'humanitarian group', 'islamic republic news agency', 'hostage', 'embassy', 'rubble', 'relief', 'group', 'tehran', 'american red cross', 'thousand', 'agency', 'authority', 'state', 'country', 'red cross', 'government', 'states', 'cross']

subclass2111111121-texts: ['a0', 'a2']

subclass21112-common words: ['people', 'nepal', 'india', 'quake', 'earthquake', 'temblor', 'area', 'field', 'bihar state', 'katmandu', 'worker', 'official', 'house', 'sign', 'indian state', 'state', 'country', 'agency', 'authority', 'household', 'home']

subclass21112-texts: ['a9', 'a11', 'a10', 'a8']

subclass211121-common words: ['people', 'nepal', 'bihar', 'mile', 'india', 'quake', 'earthquake', 'temblor', 'border', 'katmandu', 'area', 'field', 'richter scale', 'worker', 'bihar state', 'town', 'township', 'house', 'sign', 'child', 'southeast', 'indian state', 'official', 'agency', 'authority', 'state', 'country', 'household', 'home', 'scale']

subclass211121-texts: ['a9', 'a11', 'a10']

subclass2111211-common words: ['people', 'nepal', 'mile', 'bihar', 'india', 'quake', 'earthquake', 'temblor', 'rescue operation', 'border', 'richter scale', 'report', 'katmandu', 'area', 'field', 'new delhi', 'press trust', 'capital', 'west bengal', 'observatory', 'landslide', 'bihar state', 'am', 'edt', 'city', 'child', 'ground motion', 'town', 'township', 'seismograph', 'southeast', 'sikkim', 'school hostel', 'strong earthquake', 'social worker', 'several home', 'india news agency', 'damage', 'calcutta', '1000 people', 'indian state', 'house', 'sign', '100 people', 'building', 'casualty', 'severe damage', 'official', 'energy', 'state', 'country', 'saturday', 'portion', 'part', 'news agency', 'agency', 'authority', 'household', 'home', 'scale', 'delhi', 'operation', 'motion', 'trust', 'bengal', 'worker', 'hostel']

subclass2111211-texts: ['a9', 'a11']

subclass211122-common words: ['people', 'nepal', 'india', 'bihar state', 'area', 'field', 'quake', 'earthquake', 'temblor', 'official', 'region', 'worker', 'shrestha', 'katmandu', 'state', 'country', 'house', 'sign', 'aftershock', 'monsoon rain', 'assam', 'deadly quake', 'death toll', 'government', 'body', 'sunday', 'agency', 'authority', 'household', 'home', 'rain', 'toll', 'deadly earthquake']

subclass211122-texts: ['a8', 'a10', 'a11']

subclass2111221-common words: ['people', 'nepal', 'dharan', 'bihar state', 'river', 'india', 'worker', 'official', 'area', 'field', 'katmandu', 'quake', 'earthquake', 'temblor', 'house', 'sign', 'crack', 'monsoon rain', 'aftershock', 'medicine', 'state', 'country', 'region', 'helicopter', 'aid', 'assistance', 'bag', 'british army', 'death toll', 'rice', 'brick', 'police inspector g adhikari', 'one remaining', 'big calamity', 'man singh shrestha', 'body', 'deadly quake', 'road', 'official count', 'assam', 'tuesday', 'heavy rain', 'victim', 'shrestha', 'sunday', 'household', 'home', 'government', 'agency', 'authority', 'survivor', 'count', 'adhikari', 'street', 'nation', 'army', 'rain', 'toll', 'remaining', 'calamity']

subclass2111221-texts: ['a8', 'a10']

### Dimension 2 (Dim2)

Number of class trees:7

#### Tree1:

Root-common words: ['loss', 'claim', 'group', 'death', 'market', 'insurance', 'night', 'texas', 'company', 'year']

Root-texts: ['b0', 'b2']

#### Tree2:

Root-common words: ['mile']

Root-texts: ['a6', 'a11', 'a9', 'b1', 'a5', 'b4', 'a4', 'b2', 'b5', 'b7', 'b3', 'a7', 'a10']

subclass1-common words: ['mile']

subclass1-texts: ['a6', 'a11', 'a9', 'b1', 'a5', 'b4', 'a4', 'b2', 'b5', 'b7', 'b3', 'a7']

subclass11-common words: ['mile']

subclass11-texts: ['a6', 'a11', 'a9', 'b1', 'a5', 'b4', 'a4', 'b2', 'b5', 'b7', 'b3']

subclass111-common words: ['mile']

subclass111-texts: ['a6', 'a11', 'a9', 'b1', 'a5', 'b4', 'a4', 'b2', 'b5', 'b7']

subclass1111-common words: ['mile']

subclass1111-texts: ['a6', 'a11', 'a9', 'b1', 'a5', 'b4', 'a4', 'b2', 'b5']

subclass11111-common words: ['mile']

subclass11111-texts: ['a6', 'a11', 'a9', 'b1', 'a5', 'b4', 'a4', 'b2']

subclass111111-common words: ['mile']

subclass111111-texts: ['a6', 'a11', 'a9', 'b1', 'a5', 'b4', 'a4']

subclass1111111-common words: ['mile', 'report']

subclass1111111-texts: ['a6', 'a11', 'a9', 'b1', 'a5', 'b4']

subclass11111111-common words: ['mile', 'report']

subclass11111111-texts: ['a6', 'a11', 'a9', 'b1', 'a5']

subclass111111111-common words: ['mile', 'report']

subclass111111111-texts: ['a6', 'a11', 'a9', 'b1']

subclass1111111111-common words: ['mile', 'report', 'town', 'township']

subclass1111111111-texts: ['a6', 'a11', 'a9']

subclass11111111111-common words: ['region', 'magnitude', 'report', 'mile', 'seismologist', 'reading', 'town', 'township']

subclass11111111111-texts: ['a6', 'a11']

subclass112-common words: ['mile', 'city']

subclass112-texts: ['b7', 'b4', 'b3', 'b1', 'a9', 'a11']

subclass1121-common words: ['mile', 'city']

subclass1121-texts: ['b7', 'b3', 'b1', 'a9', 'a11']

subclass11211-common words: ['mile', 'city']

subclass11211-texts: ['b7', 'b3', 'b1', 'a9']

subclass112111-common words: ['mile', 'city', 'mph', 'line']

subclass112111-texts: ['b7', 'b3', 'b1']

subclass1121111-common words: ['beach', 'mile', 'mph', 'island', 'shelter', 'light', 'supply', 'city', 'line', 'bob sheets', 'wave', 'highway', 'puerto rico', 'effect', 'town', 'township', 'household', 'home', 'sheets', 'rico']

subclass1121111-texts: ['b7', 'b3']

subclass11212-common words: ['mile', 'border', 'report', 'am', 'edt', 'city', 'child', 'observatory', 'capital', 'town', 'township', 'house', 'sign']

subclass11212-texts: ['a9', 'a11']

subclass1122-common words: ['mile', 'city']

subclass1122-texts: ['b7', 'b4', 'b3', 'b1', 'a9']

subclass11221-common words: ['mile', 'city', 'line']

subclass11221-texts: ['b7', 'b4', 'b3', 'b1']

subclass112211-common words: ['mile', 'beach', 'island', 'city', 'line', 'bob sheets', 'sheets']

subclass112211-texts: ['b7', 'b4', 'b3']

subclass1122111-common words: ['beach', 'mile', 'edt', 'island', 'hour', 'fla', '200 mile', 'show', 'tide', 'isle', 'power line', 'evacuee', 'eye', 'center', 'coral gables', 'bob sheets', 'new york city', 'hospital', 'sheets', 'line', 'city', 'gables']

subclass1122111-texts: ['b7', 'b4']

subclass113-common words: ['mile']

subclass113-texts: ['b3', 'b5', 'b1', 'b4', 'a9']

subclass1131-common words: ['mile', 'hotel']

subclass1131-texts: ['b3', 'b5', 'b1', 'b4']

subclass11311-common words: ['mile', 'hotel', 'mph']

subclass11311-texts: ['b3', 'b5', 'b1']

subclass113111-common words: ['mile', 'today', 'mph', 'sheets', 'tourist', 'north', 'west', 'zone', 'hotel', 'town', 'township']

subclass113111-texts: ['b3', 'b5']

subclass12-common words: ['mile', 'office', 'magnitude', 'san jose', 'county', 'building', 'oakland', 'downtown', 'south', 'bridge', 'major fire', 'person', 'shopping mall', 'san francisco airport', '10 mile', 'marina district', 'emergency services', 'district', 'one person', 'eye', 'center', 'airport', 'another', 'jose', 'services', 'fire', 'mall']

subclass12-texts: ['a7', 'a5']

subclass2-common words: ['mile']

subclass2-texts: ['a6', 'a4', 'a11', 'a10', 'a9', 'b1', 'b2', 'b5', 'a5', 'b4', 'b7', 'b3']

subclass21-common words: ['mile']

subclass21-texts: ['a6', 'a4', 'a11', 'a10', 'a9', 'b1', 'b2', 'b5', 'a5', 'b4', 'b7']

subclass211-common words: ['mile']

subclass211-texts: ['a6', 'a4', 'a11', 'a10', 'a9', 'b1', 'b2', 'b5', 'a5', 'b4']

subclass2111-common words: ['mile']

subclass2111-texts: ['a6', 'a4', 'a11', 'a10', 'a9', 'b1', 'b2', 'b5', 'a5']

subclass21111-common words: ['mile']

subclass21111-texts: ['a6', 'a4', 'a11', 'a10', 'a9', 'b1', 'b2', 'b5']

subclass211111-common words: ['mile']

subclass211111-texts: ['a6', 'a4', 'a11', 'a10', 'a9', 'b1', 'b2']

subclass2111111-common words: ['mile']

subclass2111111-texts: ['a6', 'a4', 'a11', 'a10', 'a9', 'b1']

subclass21111111-common words: ['mile']

subclass21111111-texts: ['a6', 'a4', 'a11', 'a10', 'a9']

subclass211111111-common words: ['mile', 'seismologist']

subclass211111111-texts: ['a6', 'a4', 'a11', 'a10']

subclass2111111111-common words: ['mile', 'reading', 'seismologist']

subclass2111111111-texts: ['a6', 'a4', 'a11']

subclass21111111111-common words: ['mile', 'reading', 'seismologist', 'pasadena', 'technology']

subclass21111111111-texts: ['a6', 'a4']

subclass211111112-common words: ['mile', 'border', 'town', 'township', 'house', 'sign']

subclass211111112-texts: ['a10', 'a11', 'a9']

subclass2111111121-common words: ['darbhanga', 'region', 'border', 'mile', 'seismologist', 'shrestha', 'secretary', 'town', 'township', 'district', 'sleep', 'house', 'sign']

subclass2111111121-texts: ['a10', 'a11']

subclass2111112-common words: ['mile', 'industry', 'landfall', 'loss', 'exposure', 'centre', 'market', 'night']

subclass2111112-texts: ['b1', 'b2']

#### Tree3:

Root-common words: ['organization', 'aid', 'assistance', 'iran']

Root-texts: ['a1', 'a0', 'a3']

subclass1-common words: ['organization', 'aid', 'assistance', 'lebanon', 'hostage', 'dog', 'iran', 'soviet armenia', 'embassy', 'truck', 'hospital', 'party', 'iraq', 'armenia']

subclass1-texts: ['a1', 'a0']

subclass2-common words: ['organization', 'plane', 'aid', 'assistance', 'iran', 'mission', 'development']

subclass2-texts: ['a1', 'a3']

#### Tree4:

Root-common words: ['medical team', 'region', 'seismological observatory', 'mountainous india-nepal border region', 'tremor', 'subcontinent', 'late june', 'team', 'house', 'sign', 'witness', 'observatory', 'june']

Root-texts: ['a8', 'a11']

#### Tree5:

Root-common words: ['oil', 'group', 'year', 'america', 'portion', 'part']

Root-texts: ['b0', 'a2']

#### Tree6:

Root-common words: ['warning', 'death', 'county', 'million', 'billion']

Root-texts: ['b2', 'b8']

#### Tree7:

Root-common words: ['city']

Root-texts: ['b1', 'b6', 'b7', 'b3', 'a9', 'a11', 'b4']

subclass1-common words: ['mile', 'city']

subclass1-texts: ['b7', 'b4', 'b3', 'b1', 'a9', 'a11']

subclass11-common words: ['mile', 'city']

subclass11-texts: ['b7', 'b3', 'b1', 'a9', 'a11']

subclass111-common words: ['mile', 'city']

subclass111-texts: ['b7', 'b3', 'b1', 'a9']

subclass1111-common words: ['mile', 'city', 'mph', 'line']

subclass1111-texts: ['b7', 'b3', 'b1']

subclass11111-common words: ['beach', 'mile', 'mph', 'island', 'shelter', 'light', 'supply', 'city', 'line', 'bob sheets', 'wave', 'highway', 'puerto rico', 'effect', 'town', 'township', 'household', 'home', 'sheets', 'rico']

subclass11111-texts: ['b7', 'b3']

subclass112-common words: ['mile', 'border', 'report', 'am', 'edt', 'city', 'child', 'observatory', 'capital', 'town', 'township', 'house', 'sign']

subclass112-texts: ['a9', 'a11']

subclass12-common words: ['mile', 'city']

subclass12-texts: ['b7', 'b4', 'b3', 'b1', 'a9']

subclass121-common words: ['mile', 'city', 'line']

subclass121-texts: ['b7', 'b4', 'b3', 'b1']

subclass1211-common words: ['mile', 'beach', 'island', 'city', 'line', 'bob sheets', 'sheets']

subclass1211-texts: ['b7', 'b4', 'b3']

subclass12111-common words: ['beach', 'mile', 'edt', 'island', 'hour', 'fla', '200 mile', 'show', 'tide', 'isle', 'power line', 'evacuee', 'eye', 'center', 'coral gables', 'bob sheets', 'new york city', 'hospital', 'sheets', 'line', 'city', 'gables']

subclass12111-texts: ['b7', 'b4']

subclass2-common words: ['city']

subclass2-texts: ['b1', 'b6', 'b7', 'b3', 'a9', 'a11']

subclass21-common words: ['city']

subclass21-texts: ['b1', 'b6', 'b7', 'b3', 'a9']

subclass211-common words: ['city', 'mph']

subclass211-texts: ['b1', 'b6', 'b7', 'b3']

subclass2111-common words: ['city', 'mph', 'hugo']

subclass2111-texts: ['b1', 'b6', 'b7']

subclass21111-common words: ['city', 'loss', 'mph', 'close contact', 'hugo', 'contact']

subclass21111-texts: ['b1', 'b6']

### Dimension 3 (Dim3)

Number of class trees:12

#### Tree1:

Root-common words: ['cayman islands', 'hotel', 'island', 'ham radio operator', 'wednesday', 'resort', 'jamaica', 'spokesman', 'operator', 'boat', 'nothing', 'ham operator', 'islands', 'radio ham operator']

Root-texts: ['b3', 'b4']

#### Tree2:

Root-common words: ['industry', 'landfall', 'exposure', 'centre']

Root-texts: ['b1', 'b2']

#### Tree3:

Root-common words: ['today', 'mph', 'sheets', 'tourist', 'north', 'west', 'zone', 'hotel', 'town', 'township']

Root-texts: ['b3', 'b5']

#### Tree4:

Root-common words: ['mph', 'chamber', 'hugo', 'ten', 'decade']

Root-texts: ['b6', 'b7']

#### Tree5:

Root-common words: ['victim', 'call', 'number']

Root-texts: ['b8', 'a3']

#### Tree6:

Root-common words: ['god', 'secretary', 'donor', 'neighbor', 'crane', 'anonymity', 'condition', 'world']

Root-texts: ['a1', 'a10']

#### Tree7:

Root-common words: ['river', 'worker', 'bag']

Root-texts: ['a8', 'a10']

#### Tree8:

Root-common words: ['thing', 'ten thousand', 'thousand']

Root-texts: ['b8', 'a2']

#### Tree9:

Root-common words: ['border', 'town', 'township']

Root-texts: ['a9', 'a10', 'a11']

subclass1-common words: ['border', 'report', 'am', 'edt', 'child', 'capital', 'town', 'township']

subclass1-texts: ['a9', 'a11']

subclass2-common words: ['border', 'nepal', 'kingdom', 'others', 'town', 'township', 'police', 'men']

subclass2-texts: ['a9', 'a10']

#### Tree10:

Root-common words: ['building']

Root-texts: ['a6', 'a5', 'a7', 'b7', 'a9']

subclass1-common words: ['building', 'eye', 'center']

subclass1-texts: ['a6', 'a5', 'a7', 'b7']

subclass11-common words: ['magnitude', 'building', 'eye', 'center']

subclass11-texts: ['a6', 'a5', 'a7']

subclass111-common words: ['building', 'report', 'quake magnitude', 'frank baldwin', 'san francico', 'us geological survey earthquake center', 'magnitude', 'calif', 'eye', 'center', 'baldwin', 'francico']

subclass111-texts: ['a6', 'a5']

subclass112-common words: ['magnitude', 'probability', 'building', 'point', 'eye', 'center']

subclass112-texts: ['a6', 'a7']

subclass12-common words: ['chance', 'building', 'eyewitness', 'post', 'spot', 'place', 'next day', 'eye', 'center', 'ten', 'decade', 'others', 'day']

subclass12-texts: ['b7', 'a7']

#### Tree11:

Root-common words: ['collins', 'level']

Root-texts: ['a4', 'a5']

#### Tree12:

Root-common words: ['hill']

Root-texts: ['a0', 'a10', 'b0']

subclass1-common words: ['worker', 'hill', 'border', 'account']

subclass1-texts: ['b0', 'a10']

subclass2-common words: ['body', 'boulder', 'hill', 'earth', 'land']

subclass2-texts: ['a0', 'a10']

### Dimension 4 (Dim4)

Number of class trees:15

#### Tree1:

Root-common words: ['percent', 'point', 'information', 'report', 'length', 'size', 'lot', 'many']

Root-texts: ['b4', 'a6']

#### Tree2:

Root-common words: ['chance', 'eyewitness', 'post', 'spot', 'place', 'eye', 'center']

Root-texts: ['b7', 'a7']

#### Tree3:

Root-common words: ['plane', 'mission', 'development']

Root-texts: ['a1', 'a3']

#### Tree4:

Root-common words: ['us insurer', 'insurer', 'underwriter']

Root-texts: ['b0', 'b1']

#### Tree5:

Root-common words: ['bay', 'station', 'report', 'car']

Root-texts: ['b4', 'a5']

#### Tree6:

Root-common words: ['edt']

Root-texts: ['b4', 'b5', 'b7', 'a9', 'a11']

subclass1-common words: ['edt']

subclass1-texts: ['b4', 'b5', 'b7', 'a9']

subclass11-common words: ['edt', 'hour']

subclass11-texts: ['b4', 'b5', 'b7']

subclass111-common words: ['edt', 'hour', 'tuesday', 'inch']

subclass111-texts: ['b4', 'b5']

subclass2-common words: ['report', 'am', 'edt', 'child', 'capital']

subclass2-texts: ['a9', 'a11']

#### Tree7:

Root-common words: ['probability', 'one', 'system', 'increase', 'step-up']

Root-texts: ['b3', 'a7']

#### Tree8:

Root-common words: ['office']

Root-texts: ['b6', 'a5', 'b5', 'a7', 'a3']

subclass1-common words: ['office']

subclass1-texts: ['b6', 'a5', 'b5', 'a7']

subclass11-common words: ['office', 'window']

subclass11-texts: ['b6', 'a5', 'b5']

subclass111-common words: ['office', 'desk', 'window']

subclass111-texts: ['b6', 'a5']

#### Tree9:

Root-common words: ['hostage', 'dog', 'embassy']

Root-texts: ['a1', 'a0']

#### Tree10:

Root-common words: ['charlestonians']

Root-texts: ['b6', 'b8']

#### Tree11:

Root-common words: ['district', 'ministry', 'homeless', 'dawn', 'word']

Root-texts: ['b3', 'a10']

#### Tree12:

Root-common words: ['second']

Root-texts: ['a4', 'a9']

#### Tree13:

Root-common words: ['nation']

Root-texts: ['a3', 'a2']

#### Tree14:

Root-common words: ['hospital']

Root-texts: ['a1', 'a8', 'a9']

subclass1-common words: ['hospital', 'aftershock', 'us embassy', 'embassy']

subclass1-texts: ['a1', 'a8']

#### Tree15:

Root-common words: ['hour']

Root-texts: ['b2', 'b7', 'b5', 'b4']

subclass1-common words: ['edt', 'hour']

subclass1-texts: ['b4', 'b5', 'b7']

subclass11-common words: ['edt', 'hour', 'tuesday', 'inch']

subclass11-texts: ['b4', 'b5']

subclass2-common words: ['hour']

subclass2-texts: ['b2', 'b7', 'b5']

subclass21-common words: ['hour', 'household', 'home', 'business']

subclass21-texts: ['b2', 'b7']

### Dimension 5 (Dim5)

Number of class trees:14

#### Tree1:

Root-common words: ['community', 'member']

Root-texts: ['b1', 'a1']

#### Tree2:

Root-common words: ['impact', 'watch', 'quarter']

Root-texts: ['b0', 'b3']

#### Tree3:

Root-common words: ['hand']

Root-texts: ['b4', 'a2']

#### Tree4:

Root-common words: ['magnitude']

Root-texts: ['a7', 'a11', 'a6']

subclass1-common words: ['magnitude', 'reading']

subclass1-texts: ['a6', 'a11']

subclass2-common words: ['magnitude', 'collapse', 'colo', 'golden']

subclass2-texts: ['a7', 'a11']

#### Tree5:

Root-common words: ['nepal', 'kingdom']

Root-texts: ['a9', 'a10']

#### Tree6:

Root-common words: ['foot', 'evening']

Root-texts: ['b7', 'a4']

#### Tree7:

Root-common words: ['occupant', 'resident']

Root-texts: ['b7', 'b8', 'a8']

subclass1-common words: ['relief effort', 'flood', 'flooding', 'occupant', 'resident', 'household', 'home', 'effort']

subclass1-texts: ['b7', 'a8']

subclass2-common words: ['fury', 'occupant', 'resident']

subclass2-texts: ['b7', 'b8']

#### Tree8:

Root-common words: ['new york', 'york']

Root-texts: ['a3', 'a5']

#### Tree9:

Root-common words: ['capital']

Root-texts: ['b3', 'a11', 'a9']

subclass1-common words: ['capital', 'impact', 'southeast']

subclass1-texts: ['b3', 'a11']

subclass2-common words: ['am', 'child', 'capital']

subclass2-texts: ['a9', 'a11']

#### Tree10:

Root-common words: ['disaster']

Root-texts: ['b5', 'a8', 'b1']

subclass1-common words: ['disaster', 'threat']

subclass1-texts: ['b5', 'a8']

#### Tree11:

Root-common words: ['war', 'ground']

Root-texts: ['b6', 'a1']

#### Tree12:

Root-common words: ['village']

Root-texts: ['a0', 'a8']

#### Tree13:

Root-common words: ['supply']

Root-texts: ['b7', 'a3', 'b3']

subclass1-common words: ['nurse', 'supply', 'day']

subclass1-texts: ['b7', 'a3']

#### Tree14:

Root-common words: ['lake', 'estimate']

Root-texts: ['b2', 'a7']

### Dimension 6 (Dim6)

Number of class trees:18

#### Tree1:

Root-common words: ['am', 'child']

Root-texts: ['a9', 'a11']

#### Tree2:

Root-common words: ['history', 'person', 'force']

Root-texts: ['b1', 'a7']

#### Tree3:

Root-common words: ['tornado', 'speed', 'afternoon']

Root-texts: ['b1', 'b7']

#### Tree4:

Root-common words: ['reading']

Root-texts: ['a6', 'a4', 'a11']

subclass1-common words: ['reading']

subclass1-texts: ['a6', 'a4']

#### Tree5:

Root-common words: ['geology', 'addition']

Root-texts: ['a6', 'a9']

#### Tree6:

Root-common words: ['force', 'sea', 'airport']

Root-texts: ['b4', 'a7']

#### Tree7:

Root-common words: ['radio', 'associated press', 'spokeswoman', 'press']

Root-texts: ['b3', 'a1']

#### Tree8:

Root-common words: ['operation', 'precaution']

Root-texts: ['b0', 'b7']

#### Tree9:

Root-common words: ['household', 'home']

Root-texts: ['b2', 'a8', 'b7', 'b3']

subclass1-common words: ['household', 'home']

subclass1-texts: ['b2', 'a8', 'b7']

subclass11-common words: ['household', 'home']

subclass11-texts: ['b2', 'a8']

#### Tree10:

Root-common words: ['tuesday']

Root-texts: ['b4', 'b5', 'a4']

subclass1-common words: ['tuesday', 'inch']

subclass1-texts: ['b4', 'b5']

#### Tree11:

Root-common words: ['respect']

Root-texts: ['b1', 'b8']

#### Tree12:

Root-common words: ['assessment']

Root-texts: ['b6', 'a7']

#### Tree13:

Root-common words: ['crack']

Root-texts: ['b6', 'a10']

#### Tree14:

Root-common words: ['aftershock']

Root-texts: ['a7', 'a8', 'a1']

subclass1-common words: ['aftershock', 'airport', 'period']

subclass1-texts: ['a1', 'a7']

subclass2-common words: ['aftershock', 'rupture', 'several']

subclass2-texts: ['a7', 'a8']

#### Tree15:

Root-common words: ['dead', 'case']

Root-texts: ['a3', 'a7']

#### Tree16:

Root-common words: ['bridge']

Root-texts: ['a7', 'a5', 'a10']

subclass1-common words: ['south', 'bridge']

subclass1-texts: ['a7', 'a5']

#### Tree17:

Root-common words: ['man']

Root-texts: ['a0', 'a5']

#### Tree18:

Root-common words: ['resource']

Root-texts: ['b7', 'a2']

### Dimension 7 (Dim7)

Number of class trees:17

#### Tree1:

Root-common words: ['government', 'urgent appeal', 'appeal']

Root-texts: ['a1', 'a11']

#### Tree2:

Root-common words: ['risk', 'concern']

Root-texts: ['b0', 'a7']

#### Tree3:

Root-common words: ['kind']

Root-texts: ['a1', 'a6']

#### Tree4:

Root-common words: ['fury']

Root-texts: ['b7', 'b8']

#### Tree5:

Root-common words: ['show', 'tide']

Root-texts: ['b7', 'b4']

#### Tree6:

Root-common words: ['injury', 'cbs-tv']

Root-texts: ['b3', 'a5']

#### Tree7:

Root-common words: ['view']

Root-texts: ['b6', 'a6']

#### Tree8:

Root-common words: ['mountain', 'mountains']

Root-texts: ['a6', 'a10']

#### Tree9:

Root-common words: ['doctor']

Root-texts: ['a3', 'a10']

#### Tree10:

Root-common words: ['department']

Root-texts: ['a5', 'a9']

#### Tree11:

Root-common words: ['eastern india', 'india']

Root-texts: ['a8', 'a9']

#### Tree12:

Root-common words: ['body']

Root-texts: ['a0', 'a10']

#### Tree13:

Root-common words: ['action']

Root-texts: ['a1', 'a2']

#### Tree14:

Root-common words: ['destruction']

Root-texts: ['b1', 'a10']

#### Tree15:

Root-common words: ['sunday']

Root-texts: ['b5', 'a11']

#### Tree16:

Root-common words: ['base']

Root-texts: ['b2', 'a10']

#### Tree17:

Root-common words: ['east']

Root-texts: ['a4', 'a7']