ANNEXURE – IV

PAPERS PUBLISHED IN INTERNATIONAL AND NATIONAL CONFERENCES
ANNEXURE – V

PAPERS PUBLISHED IN JOURNALS
ANNEXURE - I
The default list of full-text stopwords as used by MySQL.

a's according able across about actually above after afterwards
according again against ain't all allow
allows almost alone along already
also although always and another anyway
amongst an apart anything appear appreciate
anybody anyhow anyone anything appear appreciate
anyways anywhere aspect around anyway appreciate
appropriate are aren't around as
aside away asking associated at
available away awfully be became
because become becomes becoming been
before beforehand behind being believe
beforehand beside besides best better
below beyond both brief but
between both c's can
can't cannot cant cause causes
certain certainly changes clearly co
come considering contains consequently
consider containing contain containing
considering containing contain containing
couldn't describe current currently
could described contains containing
definitely despite course currently
different does didn't
don't down doing during
done eg either
each earlier besides either
either elsewhere especially ever
even ever every
etc ex
everyone everywhere everything everywhere
except ex
example except
five far
first few
followed few
for following
four following
from forth
further get
given go
given gotten
given
further had
hasn't hard
hadn't hardly
have having
hello hence
<table>
<thead>
<tr>
<th>so</th>
<th>something</th>
<th>soon</th>
<th>still</th>
<th>t's</th>
<th>th</th>
<th>that</th>
<th>theirs</th>
<th>there</th>
<th>therein</th>
<th>they'd</th>
<th>third</th>
<th>though</th>
<th>thus</th>
<th>toward</th>
<th>try</th>
<th>under</th>
<th>unto</th>
<th>used</th>
<th>value</th>
<th>vs</th>
<th>way</th>
<th>we've</th>
<th>weren't</th>
<th>whence</th>
<th>whereas</th>
<th>whether</th>
<th>who's</th>
<th>why</th>
<th>within</th>
<th>wouldn't</th>
<th>you'll</th>
<th>yourself</th>
</tr>
</thead>
<tbody>
<tr>
<td>some</td>
<td>sometime</td>
<td>sorry</td>
<td>sub</td>
<td>take</td>
<td>than</td>
<td>that's</td>
<td>them</td>
<td>there's</td>
<td>theres</td>
<td>thereafter</td>
<td>they'll</td>
<td>this</td>
<td>three</td>
<td>to</td>
<td>towards</td>
<td>trying</td>
<td>unfortunately</td>
<td>up</td>
<td>useful</td>
<td>various</td>
<td>want</td>
<td>welcome</td>
<td>what</td>
<td>whenever</td>
<td>whereby</td>
<td>wherein</td>
<td>which</td>
<td>whoever</td>
<td>will</td>
<td>willing</td>
<td>won't</td>
<td>yet</td>
</tr>
</tbody>
</table>
ANNEXURE -  II

Part-of-Speech Tags

CC - coordinating conjunction: ‘and’, ‘but’, ‘nor’, ‘or’, ‘yet’, plus, minus, less, times (multiplication), over (division). Also ‘for’ (because) and ‘so’ (i.e., ‘so that’).
CD - cardinal number
EX - existential ‘there’: Unstressed ‘there’ that triggers inversion of the inflected verb and the logical subject; ‘There was a party in progress’.
FW - foreign word
IN - preposition or subordinating conjunction
JJ - adjective: Hyphenated compounds that are used as modifiers; happy-go-lucky.
JJR - adjective - comparative: Adjectives with the comparative ending ‘-er’ and a comparative meaning. Sometimes ‘more’ and ‘less’.
JJS - adjective - superlative: Adjectives with the superlative ending ‘-est’ (and ‘worst’). Sometimes ‘most’ and ‘least’.
JJSS - -unknown-, but probably a variant of JJS
-LRB- - -unknown-
LS - list item marker: Numbers and letters used as identifiers of items in a list.
NN - noun - singular or mass
NNP - proper noun - singular: All words in names usually are capitalized but titles might not be.
NNPS - proper noun - plural: All words in names usually are capitalized but titles might not be.
NNS - noun - plural
NP - proper noun - singular
NPS - proper noun - plural
PDT - predeterminer: Determiner like elements preceding an article or possessive pronoun; ‘all/PDT his marbles’, ‘quite/PDT a mess’.
POS - possessive ending: Nouns ending in ‘s’ or ‘’. 
PP - personal pronoun
PRPR$ - unknown-, but probably possessive pronoun
PRP - unknown-, but probably possessive pronoun
PRP$ - unknown, but probably possessive pronoun, such as ‘my’, ‘your’, ‘his’, ‘his’, ‘its’, ‘one’s’, ‘our’, and ‘their’.
RBR - adverb - comparative: adverbs ending with ‘-er’ with a comparative meaning.
RBS - adverb - superlative
RP - particle: Mostly monosyllabic words that also double as directional adverbs.
STAART - start state marker (used internally)
SYM - symbol: technical symbols or expressions that aren’t English words.
TO - literal “to”
VBD - verb - past tense: includes conditional form of the verb ‘to be’; ‘If I were/VBD rich...’.
VBG - verb - gerund or present participle
VBN - verb - past participle
VBZ - verb - non-3rd person singular present
VB - verb - base form: subsumes imperatives, infinitives and subjunctives.
VBP - verb - 3rd person singular present
WDT - ‘wh’-determiner
WP$ - possessive ‘wh’-pronoun: includes ‘whose’
:: - literal colon
, - literal comma
$ - literal dollar sign
- - literal double-dash
“ - literal double quotes
´ - literal grave
( - literal left parenthesis
. - literal period
# - literal pound sign
) - literal right parenthesis
’ - literal single quote or apostrophe
ANNEXURE – III

60 FEATURES FOR MACHINE LEARNING

OPENPARENT
EQUAL
BACKSLASH
INCHEMICALLIST
ASTERISK
NOCAPS
MIXCAPS
LENGTHGROUP
WINDOWSTANDFORDNLPLEMMA
OPENNLPCUNKPARSING
CHARNGRAM
OPENNLPCHUNK
CLEARNLPPOS
CONJUCTCLEARNLPPOS
APOSTROPE
SYMBOLNUMCHAR
CLOSEPARENT
QUOTATIONMARK
INFREQUENTLIST
OPENBRACKET
COMMA
ROMANNUM
WORD

ELEMENTNAMES
CLEARNLPLEMMA
PERCENT
CLOSEBRACKET
CONJUCTOPENNLPPOS
COLON
CONJUCTSTANDFORDNLPLEMMA
4PREFIX
CONJUCTSTANDFORDNLPPOS
WINDOWOPENNLPPOS
DOT
HYPHEN
CONJUCTOPENNLPCUNK
3SUFFIX
3PREFIX
MORPHOLOGYTYPEI
ENDCAPS
MORPHOLOGYTYPEII
MOLECULARFORMULAS
ALLCAPS
2SUFFIX
NUNMDIGITS
STANDFORDNLPLEMMA