

```
import expand
import re

from patterns import digits_dictionary, contractions_dictionary

def expand_number(m):
    word = []
    for char in m.group(0):
        word.append(digits_dictionary[char])
    return ' '.join(word) + ' '

def expand_contraction(m):
    return contractions_dictionary[m.group(0)]

def simplify(s):
    s = s.lower()
    s = re.sub('<br /><br />', ' ', s)
    s = re.sub('(%s)' % '|'.join(contractions_dictionary.keys()), expand_
d_contraction, s)
    s = re.sub(r'[\w\s]', ' ', s)
    s = re.sub(r'\d+', expand_number, s)
    s = re.sub(r'\s+', ' ', s)
    return s

def tokenize(s):
    return simplify(s).split(' ')
```