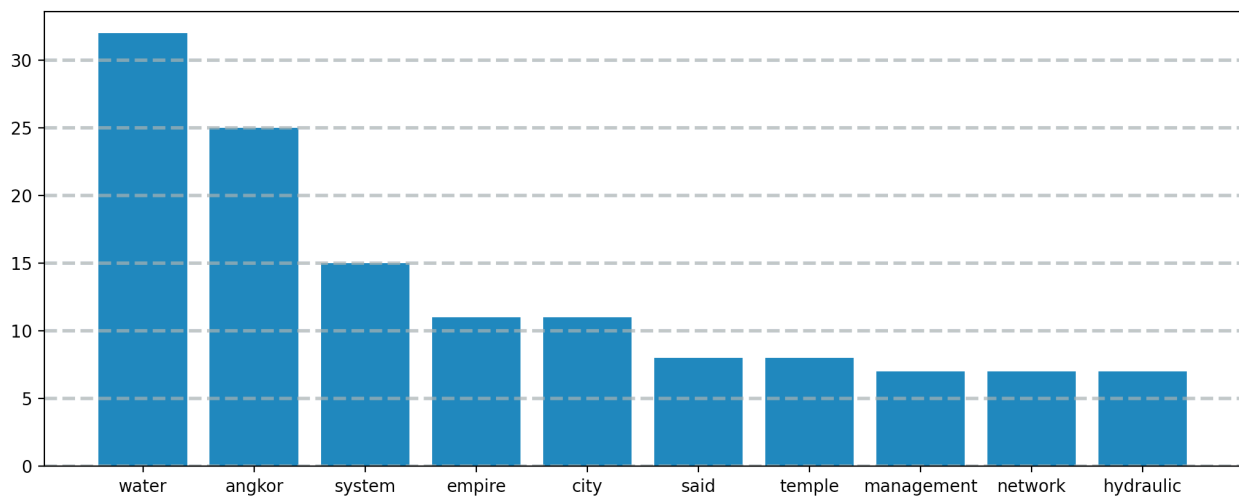


Zadanie nr 1

```
number of tokens: 1649
```

Filtered words (without stop words): 1092

Filtered words (without manual stop words and symbol based words): 861



Zadanie nr 2

```
Document-Term Matrix:
      000  000km  000m  10  100  10th  ...  yet  yilan  young  younger  zero  école
0      2      0      0  1    0    1  ...  0    0    0      0    0    1
1      0      1      2  1   11    0  ...  1    2    1      1    1    0
2      0      0      0  0    0    0  ...  2    0    0      0    0    0

[3 rows x 1377 columns]

Term Frequency Matrix:
      000      000km      000m  ...  younger      zero      école
0  0.001452  0.000000  0.000000  ...  0.000000  0.000000  0.000726
1  0.000000  0.000726  0.001452  ...  0.000726  0.000726  0.000000
2  0.000000  0.000000  0.000000  ...  0.000000  0.000000  0.000000

[3 rows x 1377 columns]

Inverse Data Frequency Matrix:
      Word      IDF
0      000  1.098612
1     000km  1.098612
2      000m  1.098612
3       10  0.405465
4       100  1.098612
...      ...      ...
1372   yilan  1.098612
1373   young  1.098612
1374  younger  1.098612
1375    zero  1.098612
1376   école  1.098612

[1377 rows x 2 columns]

Term Frequency - Inverse Data Frequency Matrix:
      000      000km      000m  ...  younger      zero      école
0  0.001595  0.000000  0.000000  ...  0.000000  0.000000  0.000798
1  0.000000  0.000798  0.001595  ...  0.000798  0.000798  0.000000
2  0.000000  0.000000  0.000000  ...  0.000000  0.000000  0.000000

[3 rows x 1377 columns]

Cosine Similarity Matrix:
      Document pairs  Cosine similarity
0  Angkor-Asias-ancient-Hydraulic-City & Can-taiw...  0.108521
1  Angkor-Asias-ancient-Hydraulic-City & The-inno...  0.245839
2  Can-taiwan-become-asias-next-great-hiking-dest...  0.140931
```

Zadanie nr 3

Odpowiedź: Podczas testu długiej opinii można było odczytać z wyników, która jest dobra, a która zła, ale ta różnica nie była aż tak wyraźną, zwłaszcza w przypadku złej opinii. Natomiast podczas testu pojedynczych zdań różnica była bardzo widoczna.

Bad review scores:

compound: 0.0715,

neg: 0.058,

neu: 0.877,

pos: 0.065,

Good review scores:

compound: 0.923,

neg: 0.0,

neu: 0.653,

pos: 0.347,

Short sentence test:

Positive short sentence:

VADER is smart, handsome, and funny.

compound: 0.8316,

neg: 0.0,

neu: 0.254,

pos: 0.746,

Negative short sentence:

Today SUX!

compound: -0.5461,

neg: 0.779,

neu: 0.221,

pos: 0.0,

Process finished with exit code 0