1강 연습문제

1. 다음은 전처리 코드이다. 출력될 6개를 순서대로 쓰시오.

```
text = 'The sky is very blue and the sky is very beautiful today.'
text = text.lower()
print(text)

text = text.replace('.', ' .')
print(text)

words = text.split(' ')
print(words)

word_to_id = {}
id_to_word = {}

for word in words:
    if word not in word_to_id:
        new_id = len(word_to_id)
        word_to_id[word] = new_id
        id_to_word[new_id] = word

print(word_to_id)
print(id_to_word)

corpus = np.array([word_to_id[w] for w in words])
print(corpus)
```

2. (i) 어휘들의 희소 표현(sparse representation)을 통해 말뭉치(corpus)를 행렬로 표현하시오.
   (ii) 코드를 작성하여 출력하시오.

3. (i) 윈도우 크기를 1로 잡았을 때 말뭉치(corpus)의 동시발생 행렬(co-occurrence matrix)을 구하시오.
   (ii) 윈도우 크기를 2로 잡았을 때 말뭉치(corpus)의 동시발생 행렬(co-occurrence matrix)을 구하시오.
   (iii) `create_co_matrix` 함수를 이용하여 확인하시오.

4. (i) 위 희소표현과 동시발생 행렬 각각에 대해 blue와 beautiful의 코사인 유사도(cosine similarity) 값을 구하시오.
   (ii) `cos_similarity` 함수를 이용하여 확인하시오.