

Problem13

Нестерова Ирина

October 2018

1 Задача

В задаче требуется выделить темы на коллекции документов при помощи сингулярного разложения (SVD метода).

2 Код

Определим рабочую папку, импортируем документы и словарь.

```
mydir = 'C:\ProgramFiles\MATLAB\problem13\';  
fileID_1 = [mydir,'ball.txt'];  
fileID_2 = [mydir,'bull.txt'];  
fileID_3 = [mydir,'rabbit.txt'];  
dicID = [mydir,'dictionary.txt'];  
doc_1 = importdata(fileID_1);  
doc_2 = importdata(fileID_2);  
doc_3 = importdata(fileID_3);  
word = importdata(dicID);
```

m - число документов, n - число слов в словаре, mat - рабочая матрица.

```
m = 3;  
n = length(word);  
mat = zeros(m,n);
```

Циклы в которых мы записываем значения в матрицу документ-слово, а точнее сколько раз слово из словаря встречается в документе.

```
for j = 1:4  
    for i = 1:n  
        k = findstr(doc_1j,1, wordi,1);  
        x = isempty(k);  
        y = 1 - x;  
        z = mat(1,i);
```

```

mat(1,i) = z + y;
end
end

```

```

for j =1:4
    for i=1:n
        k = findstr(doc_2j,1, wordi,1);
        x = isempty(k);
        y = 1 - x;
        z = mat(2,i);
        mat(2,i) = z + y;
    end
end

```

```

for j =1:4
    for i=1:n
        k = findstr(doc_3j,1, wordi,1);
        x = isempty(k);
        y = 1 - x;
        z = mat(3,i);
        mat(3,i) = z + y;
    end
end

```

Применяем метод сингулярного разложения.

```

[U, S, V] = svd(M);

```