



НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ  
УНИВЕРСИТЕТ

Кафедра прикладной математики и  
информатики

# Компьютерная лингвистика. Практика.

#7

Нижний Новгород, 2019



# Лабораторная работа #8.

Тематическое моделирование

## Задача.

1. Выполнить тематическое моделирование в собственном корпусе. Выбрать лучший алгоритм, подобрать оптимальные параметры, в т.ч. количество латентных тем
2. Оценить результат



# Лекция #7.

- Векторное представление слов
- Введение в нейронные сети



- Skip-Gram

Source Text	Training Samples			
<table><tr><td>The</td><td>quick</td><td>brown</td></tr></table> fox jumps over the lazy dog. ➡	The	quick	brown	(the, quick) (the, brown)
The	quick	brown		
The <table><tr><td>quick</td><td>brown</td><td>fox</td></tr></table> jumps over the lazy dog. ➡	quick	brown	fox	(quick, the) (quick, brown) (quick, fox)
quick	brown	fox		
The quick <table><tr><td>brown</td><td>fox</td><td>jumps</td></tr></table> over the lazy dog. ➡	brown	fox	jumps	(brown, the) (brown, quick) (brown, fox) (brown, jumps)
brown	fox	jumps		
The quick brown <table><tr><td>fox</td><td>jumps</td><td>over</td></tr></table> the lazy dog. ➡	fox	jumps	over	(fox, quick) (fox, brown) (fox, jumps) (fox, over)
fox	jumps	over		



- CBOW (Continuous Bag Of Words)

context word    center words

I like playing football with my friends →

[0, 1, 0, 0, 0, 0, 0]  
[0, 0, 1, 0, 0, 0, 0]

[1, 0, 0, 0, 0, 0, 0]

I like playing football with my friends →

[1, 0, 0, 0, 0, 0, 0]  
[0, 0, 1, 0, 0, 0, 0]  
[0, 0, 0, 1, 0, 0, 0]

[0, 1, 0, 0, 0, 0, 0]

I like playing football with my friends →

[1, 0, 0, 0, 0, 0, 0]  
[0, 1, 0, 0, 0, 0, 0]  
[0, 0, 0, 1, 0, 0, 0]  
[0, 0, 0, 0, 1, 0, 0]

[0, 0, 1, 0, 0, 0, 0]

I like playing football with my friends →

[0, 1, 0, 0, 0, 0, 0]  
[0, 0, 1, 0, 0, 0, 0]  
[0, 0, 0, 0, 1, 0, 0]  
[0, 0, 0, 0, 0, 1, 0]

[0, 0, 0, 1, 0, 0, 0]

I like playing football with my friends →

[0, 0, 1, 0, 0, 0, 0]  
[0, 0, 0, 1, 0, 0, 0]  
[0, 0, 0, 0, 0, 1, 0]  
[0, 0, 0, 0, 0, 0, 1]

[0, 0, 0, 0, 1, 0, 0]

I like playing football with my friends →

[0, 0, 0, 1, 0, 0, 0]  
[0, 0, 0, 0, 1, 0, 0]  
[0, 0, 0, 0, 0, 0, 1]

[0, 0, 0, 0, 0, 1, 0]

I like playing football with my friends →

[0, 0, 0, 0, 1, 0, 0]  
[0, 0, 0, 0, 0, 1, 0]

[0, 0, 0, 0, 0, 0, 1]



НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ  
УНИВЕРСИТЕТ