

# Программа Перезапуск

## Модуль ML.

### Занятие 4: NLP in prod

Преподаватель: Марат Гарафутдинов

Закрываем долги



# Word Embeddings

- **One-hot vectors:**

Rome Paris word V

Rome = [1, 0, 0, 0, 0, 0, ..., 0]

Paris = [0, 1, 0, 0, 0, 0, ..., 0]

Italy = [0, 0, 1, 0, 0, 0, ..., 0]

France = [0, 0, 0, 1, 0, 0, ..., 0]

**Problems:**

- Huge vectors
- VERY sparse
- No semantics or word similarity information included

# Word Embeddings

- **One-hot vectors:**

Rome Paris word V

Rome = [1, 0, 0, 0, 0, 0, ..., 0]

Paris = [0, 1, 0, 0, 0, 0, ..., 0]

Italy = [0, 0, 1, 0, 0, 0, ..., 0]

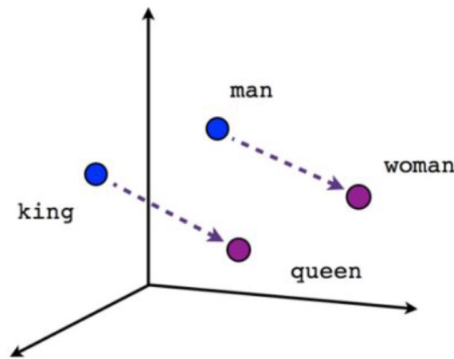
France = [0, 0, 0, 1, 0, 0, ..., 0]

**Problems:**

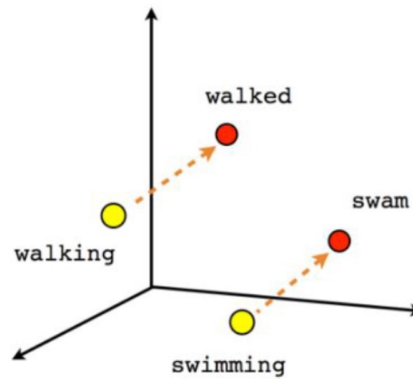
- Huge vectors
- VERY sparse
- No semantics or word similarity information included

# Word Embeddings: word2vec

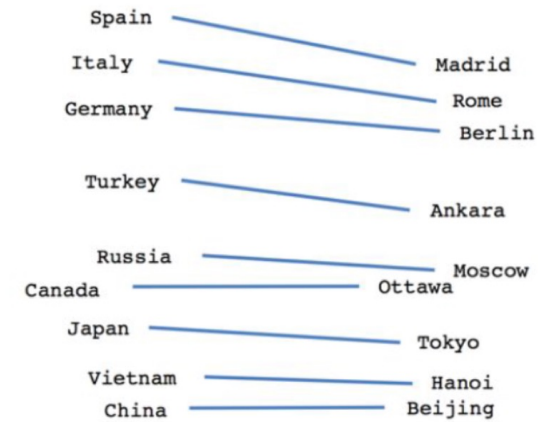
- **Word2vec** (Mikolov et al. 2013) - a framework for learning word embeddings



Male-Female



Verb tense

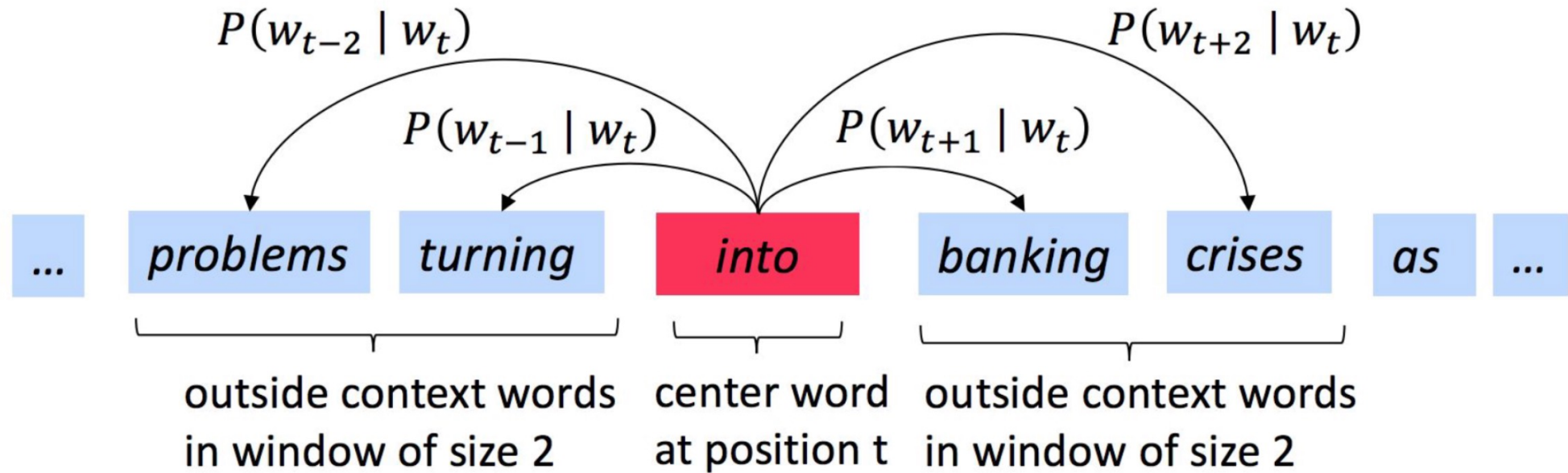


Country-Capital

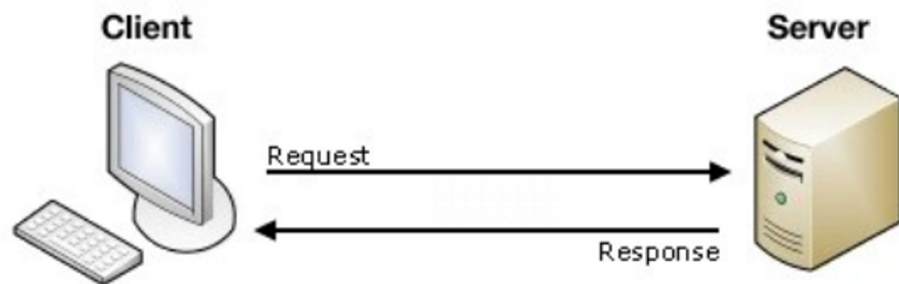
# Word Embeddings: word2vec

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

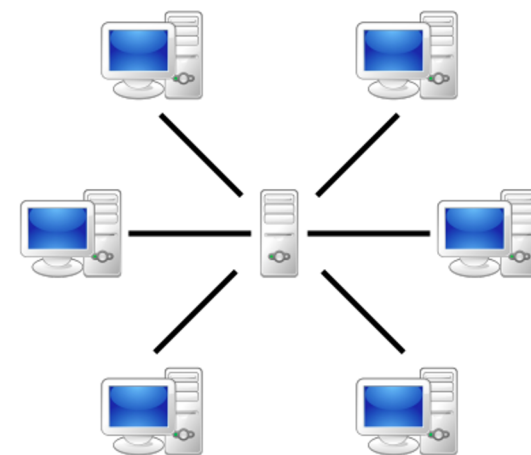
# Word Embeddings: word2vec



# Удаленный сервер



Удаленный сервер – это компьютер, к которому вы можете получить доступ посредством глобальной сети.





# Базовая команда: выгрузка данных из сети

## **wget ссылка**

скачать файл по ссылке и сохранить в текущей директории

## **wget -P путь\_до\_директории ссылка**

скачать файл по ссылке и сохранить в директории заданной путем

# Базовая команда: Ввод / вывод

## **cat файл**

вывести содержимое файла на экран

## **vim файл**

редактировать файл (можно нажать [A] для старта редактирования)

[ESC] + :q - выход без сохранения

[ESC] + :q - выход с сохранением

# Базовая команда: Ввод / вывод

## **cat файл**

вывести содержимое файла на экран

## **vim файл**

редактировать файл (можно нажать [A] для старта редактирования)

[ESC] + :q - выход без сохранения

[ESC] + :q - выход с сохранением

## **mkdir [название папки/путь]**

Создание локальной директории

## **rm -r [путь до файла/папки]**

Удаление файла/папки

# Базовая команда: архивы

## Распаковка архивов:

**unzip архив.zip**

распаковать содержимое архива.zip

**gunzip архив.gz**

распаковать содержимое архива.gz

## Создание архивов:

**zip архив.zip файл1 файл2 ...**

запаковать перечисленные файлы и/или папки в архив.zip

**tar -cvf архив.tar файл1 файл2 ...**

запаковать перечисленные файлы и/или папки в архив.tar (без сжатия)

# Базовая команда: поиск

**find <папка> -name “<имя файла>”**

найти указанный файл в папке

**find ~/ -name “file.txt”**

найти file.txt в домашней директории

**find ~/ -name “\*.txt”**

найти все текстовые файлы в домашней директории

**grep “<строка>” <файл>**

найти строку в файле

**grep -r “<строка>” <папка>**

найти строку во всех файлах в папке

**grep “hello” file.txt**

найти “hello” в файле file.txt

# Базовая команда: копирование файлов

**Копирование файлов:**

**scp -P порт логин@адрес\_сервера:путь1 путь2**  
с сервера (путь1) на клиента (путь2)



# Базовая команда: установка библиотек

Установка программ через терминал:

`sudo apt-get install программа`

Удаление программ через терминал:

`sudo apt-get remove программа`

Обновление ссылок на пакеты:

`sudo apt-get update`

Обновление установленных пакетов:

`sudo apt-get upgrade`

Обновление отдельной программы:

`sudo apt-get install --only-upgrade программа`



# Базовая команда: контроль программ

**Ctrl + C** прервать выполнение

**Ctrl + Z** приостановить выполнение

**jobs**

посмотреть запущенные программы

**ps**

посмотреть ваши процессы

**top**

отслеживать процессы в реальном времени

**kill <номер процесса>**

завершить процесс с этим номером

**kill -9 <номер процесса>**

“убить” процесс с этим номером





# Базовая команда: Менеджер терминалов

## tmux

**Ctrl + C** прервать выполнение

**Ctrl + Z** приостановить выполнение

**jobs**

посмотреть запущенные программы

**ps**

посмотреть ваши процессы

**top**

отслеживать процессы в реальном времени

**kill <номер процесса>**

завершить процесс с этим номером

**kill -9 <номер процесса>**

“убить” процесс с этим номером

# Базовая команда: venv

**virtualenv** является инструментом для создания изолированных сред Python. Эта программа создает папку, которая содержит все необходимые исполняемые файлы для использования пакетов, необходимых для проекта Python.

Создание новой виртуальной среды

**virtualenv foo**

Активация виртуальной среды

**source foo/bin/activate**

Выход из виртуальной среды

**deactivate**



# Программирование на bash

sh (Bourne shell) – стандарт и доступна почти в любом дистрибутиве (1977)

bash (Bourne again shell) – усовершенствованная и модернизированная вариация sh (1989)



оболочка (**shell**) –  
интерпретатор команд

Создаем программу:  
**vim my\_first\_script.sh**

КОД:  
**#!/bin/bash echo "Hello, world!"**