

Домашняя работа №5

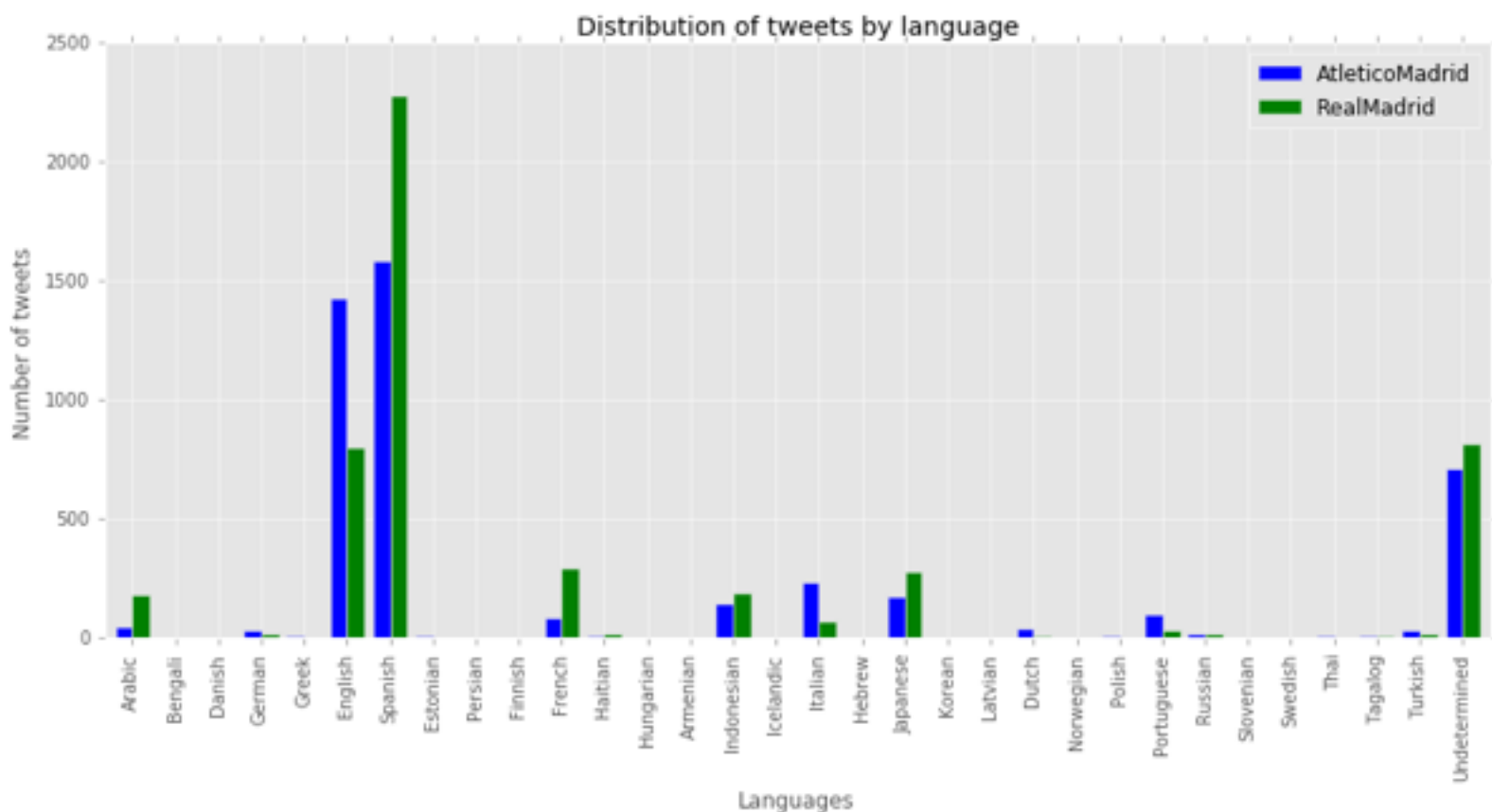
Постников М. М.

Задание:

Получим две поисковые выборки некоторой длины для каждой из команд (считаем 5000 достаточной длиной) по запросам "RealMadrid" и "AtleticoMadrid" за последний месяц. Будем считать, что говорящие на языке L пользователи болеют за команду из первой выборки, если в ней больше твитов на языке L, чем во второй. Необходимо проанализировать предпочтения болельщиков всех языков, на которых были найдены твиты.

Полученные данные должны быть визуализированы (любым средством). Результат в формате pdf должен быть приложен к проекту.

Полученные данные:



Из графика видно, что большинство твитов написаны на испанском и английском языках. Существенная часть твитов имеет неопределенный язык.

Атлетико Мадрид больше упоминают на немецком, английском, итальянском, нидерландском, португальском и турецком языках.
Реал Мадрид - на арабском, испанском, французском, индонезийском, японском.

Посмотрим на облаков тегов на некоторых языках:



Испанский язык, Атлетико Мадрид



Испанский язык, Реал Мадрид



Английский язык, Атлетико Мадрид



Английский язык, Реал Мадрид