Техническое задание:

1. Собрать массив данных
   1. все сообщения и все комментарии
   2. Период 26 января 2014 – 11 февраля 2015
2. Вид данных
   1. Единым текстовым массивом
   2. В виде отдельных юнитов анализа: каждый пост и каждый комментарий – один юнит анализа
3. Проанализировать данные, используя словарь (смотри ниже):

а. Получить долю присутствия элементов словаря в

аа. ~~В общем словарном потоке (тупо все слова (исключая предлоги и местоимения, за исключением стоп-слов - Test The Text)) – из единого текстового массива, на фоне других слов~~

**аб. ~~В сообщениях (в скольких сообщениях присутствует тот или иной элемент словаря)~~**

б. Как «далеко» (в словах) те или иные элементы словаря в среднем находятся друг от друга

в. ~~Как изменяется частота появления элементов словаря в сообщениях в зависимости от времени (по месяцам, если будут пики, то с пиками)~~

4) Визуализировать временные графики, визуализировать графики расстояния

Примечание: этот пункт можно решать в самую последнюю очередь.

Словарь для поиска:

1 группа: глаголы - характеристики умеренно-негативного отношения к сторонникам майдана

скака/ -ли, ть, скачет, доскакались» - '\S\*?[С|с]ка[к|ч]\S+'

+ плакат (в дистанции 1-2х слов)

бега - ть/ли/л

+ плакат (в дистанции 1-2х слов)

прыгать-прыгали

2 группа: существительные - характеристики исторического template №1

банде /ра, ры, овцы '\S\*?[Б|б][аеэ]нд[эе]р\S+'

бендера / ры, ра, овцы

бандерштадт

бандерлог/-и/

3 группа - существительные - характеристики исторического template №2

фашист /ты/тов '\S\*?[Ф|ф][ао]ши[cз]\S+'

укрофашист /ты

нацисты/т '\S\*?[Н|н]аци[cз]\S+'

укронацисты/-т

каратели/ль '\S\*?[К|к]арател\S+'

4 группа: характеристики умеренно-негативного отношения к сторонникам украинцам вообще

укроп/ - пы '\S\*?[У|у]кроп\S+'

укропия

5 группа: содержательные выражения причин негативного отношения

Предатели/ -ль '\S\*[П|п]редател\S+'

убийцы/ ца '\S\*?[У|у]би\S+'

сгореть, сгоревших, сожженых '\S\*?[Г|г]ор[ея]\S+' '\S\*?[С|с]о[жг]\S+'

убить/убивать/убили '\S\*?[У|у]би\S+'

бомбят/бомбить/ разбомбить/бомбили ''\S\*?[Б|б]омб\S+'

Примечание: каждое из этих слов должно учитывать нахождение в радиусе 1-2х слов от него следующих выражений:

Одесса '\S\*?[О|о]дес(с)?\S+'

дети '\S\*?[Д|д]ет\S+'

старики '\S\*?[C|с]тар\S+'

6 группа: прямое обращение к исторической памяти

памятник/ - ки '\S\*?[П|п]амятник\S+'

история/-ю '\S\*?[И|и]стор\S+'

великая отечественная '\S\*?[В|в]елик\S+\sотечественн\S+'

дед/ деды '\S\*?дед\S+'

7 группа - существительные - характеристики умеренно-негативного отношения к сторонникам майдана

майдан '\S\*?майдан\S+'

майдаун '\S\*?даун\S+'

даун '\S\*?даун\S+'

8 группа - имена собственные украинских политиков - элемент персонификации вины.

Порошенко/ Порошенко '\S\*?порошен\S+'

Юля/ Тимошенко/ Тимошенки '\S\*?юл\S{1,5}' '\S\*?тимошен\S+'

Ляшко/ Ляшки '\S\*?ляшк\S+'

Правый Сектор '\S\*?прав\S+\sсектор'

Ярош '\S\*?ярош\S+'

Яценюк/Яйцешлюх/ '\S\*?\яценюS+' '\S\*?яйцешлю\S+'

Турчинов '\S\*?турчин\S+'

Кровавый Пастор '\S\*?кровав\S+\sпастор'

9 группа - максимальный уровень абстракции

Украина 'украин\S+{1,2}'

10 группа - контр дискурс, связанный с основными template

ополченцы + наши 'ополчен\S+'

защитники + наши 'защитни\S+'

герои + наши 'геро[йияе]\S\*'