

Для решения этой задачи, я первую очередь зашёл на MyShows. Не найдя топа на состояние 2014 года, я пошёл в webarchive и взял топ ровно с той даты.

Скачал в пару кликов субтитры с сайта(<http://subs.com.ru/list.php?c=serials>), получалось брать сериал либо по сезонам, либо целиком. После скачивания занялся написанием парсера на языке python.

Посчитал для каждого сериала кол-во одинаковых слов, кол-во уникальных слов и самое длинное слово.

Название	Кол-во слов	кол-во уникальных	Коэф лекс разнообраз.	Самое длинное слово
Lost	46595	6979	15%	Пространственно-временной
Sherlock	24459	5987	24.4%	поймайменяпокаянеубилснова
The Big Bang Theory	51569	10535	20.4%	гейлесбобисексуальнотрансвеститского
House	92278	17486	18.9%	Гиперинсулинемичный-эугликемичный
Game of Thrones	14858	3913	26.35%	Самосовершенствования
Supernatural	60986	10492	17.2%	Мэри-фрэнк-шесть-ноль-три-семь
How I Met Your Mother	54870	11034	20.1%	свободныйбракетоестественно
Dexter	61314	9334	15.2%	Учеников-десятиклассников
Breaking Bad	47259	9269	19.6	высококвалифицированный
Misfits	25689	5368	21%	Интернет-провайдером

Некоторые слова получились из двух или трех, участвующих в одной фразе, но от этого только интереснее. Так же была сделана автоматическая оценка по пяти параметрам для каждого сериала:

- Оценка положительного настроения сериала
- Оценка грубости, жесткости и насилия
- Оценка похабности и развязности речи героев
- Оценка древности
- Оценка научных терминов

Название	Положитель.	Жестокость	Развязность	Древность	Научность
Lost	16.73	2.918	17.59	7.296	2.575
Sherlock	16.55	3.924	4.906	8.994	5.315
The Big Bang Theory	20.53	2.831	5.429	4.653	6.593
House	12.64	2.871	7.260	5.526	12.67
Game of Thrones	11.44	3.903	0.0	32.30	2.019
Supernatural	15.10	4.624	9.182	14.42	2.623
How I Met Your Mother	17.84	2.624	21.14	4.920	1.093
Dexter	20.15	5.365	5.382	6.686	3.098
Breaking Bad	19.61	2.835	13.33	7.405	1.904
Misfits	19.07	5.177	56.83	4.281	4.281

Данные оценки строятся по проценту встречающихся слов, принадлежащих какой-то группе. Так, например, можно сказать, что сериал, где чаще всего говорят о крови, смерти и убийствах – это Dexter и Misfits, за ним идет Supernatural и Sherlock.

Пошлость и сленг на подобию “чувак, фигня, грудь” и так далее, часто фигурируют в Misfits, How I Met Your Mother и Lost. Самым научным является House, а сериал с большим кол-вом упоминания древности это Game of Thrones и Supernatural. Оценка положительности у всех достаточно высока, ибо в сериалах, как и в жизни, персонажи достаточно часто произносят слова “привет, спасибо, дети, легко, хороший...”. Но, как пример, Game of Thrones сам по себе, как видно по таблице, серьезный сериал, поэтому эти слова там произносятся реже, чем в других (слово привет было использовано в Thrones 0 раз), так же как и коэффициент развязности у этого сериала (он равен нулю), хоть и в самом сериале есть сцены эротического характера, но упоминаний этого в сказанных словах почти нет.

Эти оценки очень приблизительны, и их подсчёт это всего лишь поиск слова в той или иной группе составленной вручную, однако по ним уже можно уверенно сравнивать этот топ10 не смотря самих сериалов.