

# Энтузиаст нашёл самую грустную песню Radiohead с помощью анализа данных

 [tjournal.ru/41346-entuziast-nashyol-samuu-grustnuu-pesnu-radiohead-s-pomoshu-analiza-dannih](https://tjournal.ru/41346-entuziast-nashyol-samuu-grustnuu-pesnu-radiohead-s-pomoshu-analiza-dannih)

И построил график «депрессивности» их композиций.



Состав Radiohead

Исследователь в области науки данных Чарли Томпсон (Charlie Thompson) использовал язык программирования R, который в том числе предназначен для статистической обработки данных, чтобы найти «самую депрессивную песню на основе аналитики». Подробный процесс поисков Томпсон [опубликовал](#) в своём блоге.

Изначально учёный воспользовался API (*интерфейс для программирования приложений*) музыкального сервиса Spotify, чтобы получить информацию о каждом треке из девяти альбомов группы. Томпсон выяснил, что Spotify оценивает песни по параметру «валентность», который сервис описывает как «значение от 0.0 до 1.0, показывающее позитивность музыки в зависимости от трека». По этому показателю песни Radiohead «True Love Waits» и «We Suck Young Blood» оказались антирекордсменами с рейтингом в районе 0.0378.

На этом исследователь не остановился и изучил тексты песен. Томпсон использовал краудсорсинговый

проект Канадского национального исследовательского совета, в котором учёные разделили все английские слова по принадлежности к восьми базовым эмоциям (злость, страх, ожидание, доверие, удивление, печаль, радость и отвращение) и двум настроениям (отрицательные или положительные).

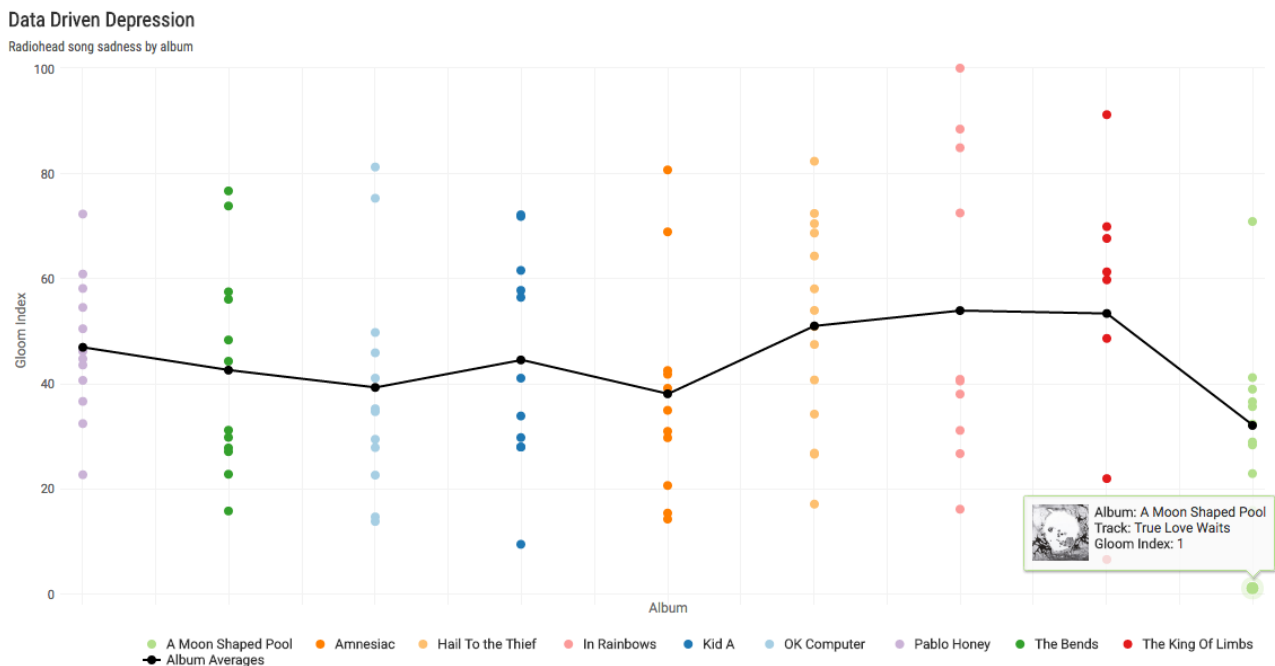
Затем с помощью API сервиса Genius.com, Томпсон собрал тексты песен Radiohead и сравнил их с заранее заготовленным словарём «грустных» слов. По «текстовому» анализу самой депрессивной стала композиция «High and Dry», которая содержала 36% слов из словаря.

Однако Томпсону также нужно было определить, какие слова были самыми важными в каждой песне. Поэтому «вишенкой на торте» исследования стала формула, которая одновременно определяла валентность каждого трека и процент грустных слов в нём.

$$gloomIndex = \frac{(1 - valence) + pctSad * (1 + lyricalDensity)}{2}$$

Индекс мрака = (1-валентность) + процент грусти \* (1+плотность Текста) / 2

Результаты были увеличены до размера 1 к 100, чтобы обеспечить большую наглядность. Полученные данные исследователь использовал для построения интерактивного [графика](#).



Самой депрессивной песней по «индексу мрака» Томпсона стала «True Love Waits», выпущенная в 2001 году. Можно сделать вывод, что именно эта композиция содержит больше всего слов, определяемых как «грустные», и при этом наименее «валентна».

+69

Нашли опечатку? Выделите фрагмент и отправьте нажатием Ctrl+Enter.