

Zoekmachines

Luca Verhees (10837833)
Sophie Ensing (10751297)
Mitchell Kuin (10764631)

27-10-2017



Dataset

September 21st, 2017

TED Main Dataset

- **name:** The official name of the TED Talk. Includes the title and the speaker.
- **title:** The title of the talk
- **description:** A blurb of what the talk is about.
- **main_speaker:** The first named speaker of the talk.
- **speaker_occupation:** The occupation of the main speaker.
- **num_speaker:** The number of speakers in the talk.
- **duration:** The duration of the talk in seconds.
- **event:** The TED/TEDx event where the talk took place.
- **film_date:** The Unix timestamp of the filming.
- **published_date:** The Unix timestamp for the publication of the talk on TED.com
- **comments:** The number of first level comments made on the talk.
- **tags:** The themes associated with the talk.
- **languages:** The number of languages in which the talk is available.
- **ratings:** A stringified dictionary of the various ratings given to the talk (inspiring, fascinating, jaw dropping, etc.)
- **related_talks:** A list of dictionaries of recommended talks to watch next.
- • **url:** The URL of the talk.
- **views:** The number of views on the talk.

TED Transcripts Dataset

- 1. **url:** The URL of the talk
- 2. **transcript:** The official English transcript of the talk.

Formatting

	comments	description	duration	event	film_date	languages	main_speaker	name	num_speaker	published_date	ratings	speaker_occupation	tags	title	url	views	transcript	tokens	word_cloud
0	4553	Sir Ken Robinson makes an entertaining and pro...	19:24	TED2006	2006-02-25	60	Ken Robinson	Ken Robinson: Do schools kill creativity?	1	2006-06-27	Inspiring, Funny, Persuasive	Author/educator	children, creativity, culture, dance, educatio...	Do schools kill creativity?	https://www.ted.com/talks/ken_robinson_says_sc...	47227110	Good morning. How are you? It's been great, ha...	[secondly, highlytalented, gillian, hears, hierarchy, roy...	highlytalented, gillian, hears, hierarchy, roy...
1	265	With the same humor and humanity he exuded in ...	16:17	TED2006	2006-02-25	43	Al Gore	Al Gore: Averting the climate crisis	1	2006-06-27	Funny, Informative, Inspiring	Climate advocate	alternative energy, cars, climate change, cult...	Averting the climate crisis	https://www.ted.com/talks/al_gore_on_averting_...	3200520	Thank you so much, Chris. And it's truly a gre...	[monterey, consider, chain, global, focus, mon...	branding, calculator, washington, carbonneutra...
2	124	New York Times columnist David Pogue takes aim...	21:26	TED2006	2006-02-24	26	David Pogue	David Pogue: Simplicity sells	1	2006-06-27	Funny, Informative, Inspiring	Technology columnist	computers, entertainment, interface design, me...	Simplicity sells	https://www.ted.com/talks/david_pogue_says_sim...	1636292	Hello voice mail, my old friend. I've called ...	[code, whoever, novelty, magnetic, global, mon...	cupertino, code, features, links, buttons, gad...
3	200	In an emotionally charged talk, MacArthur-winn...	18:36	TED2006	2006-02-26	35	Majora Carter	Majora Carter: Greening the ghetto	1	2006-06-27	Inspiring, Courageous, Persuasive	Activist for environmental justice	MacArthur grant, activism, business, cities, e...	Greening the ghetto	https://www.ted.com/talks/majora_carter_s_tale...	1697550	If you're here today — and I'm very happy that...	[surrounded, grassroots, dollar, atmosphere, f...	poor, neighborhood, grassroots, lowincome, com...
4	593	You've never seen data presented like this. Wi...	19:50	TED2006	2006-02-22	48	Hans Rosling	Hans Rosling: The best stats you've ever seen	1	2006-06-27	Informative, Fascinating, Jaw-dropping	Global health expert; data visionary	Africa, Asia, Google, demo, economics, global ...	The best stats you've ever seen	https://www.ted.com/talks/hans_rosling_shows_t...	12005869	About 10 years ago, I took on the task to teac...	[neighbors, incomewe, concept, searchable, glo...	expectancy, population, statistical, searchabl...

Evaluation

- **Queries**

- *Virtual reality experience*
- *Big data* → “big problems” solved with “data”
- *Why do people value tradition?* → ambiguous
- *New York* → speakers from New York
- *Creativity business*

- **First evaluation**

- Cohen's Kappa between -0,2 and 0,75
- Some queries hard to evaluate
- Differences in judgement → different interpretation

Improvements

- Tokens instead of full text for field *transcript*
- Ignoring fields → *title, main speaker, event (location)*
- Ignoring field length → *event, name, ratings, speaker occupation*

Evaluation

	Cohen's kappa (κ)		Precision	
	EVALUATION 1	EVALUATION 2	EVALUATION 1	EVALUATION 2
Q1	-0,19	1	0,84	1
Q2	0,62	0,62	0,87	0,87
Q3	1	1	0	0,61
Q4	0,74	0,61	0,91	0,95
Q5	0,35	0,55	0,89	0,93



kibana

[live demo](#)