



LREC-COLING 2024

Tell Me Again! A Large-Scale Dataset of Multiple Summaries for the Same Story

Hans Ole Hatzel, Chris Biemann

Language Technology Group, Department of Informatics {hans.ole.hatzel, chris.biemann}@uni-hamburg.de

Extracting Plot Summaries from Wikipedia

文A 18 languages ~			
Read Edit View history Tools			

Day of the Dead is a 2008 American horror film directed by Steve Miner and written by Jeffrey Reddick. It is a remake of George A. Romero's 1985 film of the same name, the third in Romero's Dead series,^[3] and it is the first of two remakes of the original 1985 film; the other is *Day of the Dead: Bloodline* (2017). The film sees a virus outbreak that causes people to turn into violent zombie-like creatures. The project was principally shot in Bulgaria, with limited shooting in Los Angeles, California.

Day of the Dead was released direct-to-video on April 8, 2008, by First Look Studios. It was panned by critics.

Plot [edit]

Minors Trevor Bowman and Nina meet in an abandoned warehouse in Colorado. At the same time, military roadblocks seal the city off for a 24-hour guarantine exercise. Corporal Sarah Bowman leaves her barricade and drives with Private Bud Crain to visit her sick mother. There, Trevor and Nina reveal that the local populace got infected by a influenza-like virus and that their friend Kyle gushed blood from the



Multiple Summaries for "Day of the Dead"

A zombie apocalypse has ravaged the entire world. A handful of surviving humans live within a secure underground bunker housing scientists and soldiers in the Everglades. [...]

English Original

Following the global invasion of the undead, a small group arrives by helicopter to search for possible survivors. Disappointed, they return to their base, a fortified military camp. [...]

Translation from French

The end of the human race seems to be upon us. The only thing left in the cities are maneating undead. [...]

The dead have awakened and have been ruling the globe for some time now. The film opens with Jamaican pilot John leading a group of people made up of Sarah, [...]

Translation from Italian

Cannibal zombies have taken over the world and now the ratio is one living human for every 400,000 zombies leaving humanity on the brink of extinction. The few remnants of the U.S. government and military are hiding in shelters and small colonies as scientists race against the clock to find a solution to the zombie pandemic. [...]

Translation from Spanish

nose that morning. Sarah and Bud head to Kyle's house to investigate and uncover his parents' mauled corpses. She radios Captain Rhodes about the situation and heads homeward to bring her family and Nina to the Medical Center.

As the CDC's Dr. Logan questions Sara in the crowded hospital, the infected become catatonic and reanimate as zombies. As carnage ensues, Nina and Trevor seek refuge at the local radio station, and Captain Rhodes is mauled. Dr. Logan, Sarah and Bud rush to a storage room, but Bud inadvertently drops the car keys before entering. Sarah and Bud resolve to reach the room Captain Rhodes was mauled in through the air ducts to retrieve his Humvee keys. When they land on the floor, Private Salazar appears. During their return to the storeroom, Rhodes rises to pursue them and bites Bud's hand as he replaces the ceiling grille. The group jumps from the window into the undeadinfested parking lot. Dr. Logan deliberately pushes a woman toward a zombie and departs in a vehicle.

Motivation

- What constitutes similar stories?
- Narrative similarity is not solvable with today's models
- Summaries as a surrogate for longer texts

How do we get multiple summaries?

- Starting from Wikidata -> metadata
- Consider multiple Wikipedia language version

Based on Produced by	Day of the Dead by George A. Romero Boaz Davidson James Dudelson Randall Emmett George Furla
Starring	Mena Suvari Nick Cannon Michael Welch AnnaLynne McCord Stark Sands Matt Rippy Pat Kilbane Taylor Hoover Christa Campbell Ving Rhames
Cinematography	Patrick Cady
Edited by	Nathan Easterling

Translation from German

Dataset Statistics

- **Number of works**: 29,505 of that: \bullet
 - Movies: 25,860
 - Books: 3644 \bullet
- Number of unique summaries: 96,831



Experiment: Entity Matching

Observation: All similarity models heavily focuses on entities!

We approach this by anonymizing entities in texts, using a coreference model, NER and rule-based replacements.

Model Name	Regular	Anonymized
sentence-T5-large all-monet-base-v2	0.920 0.920	0.575 0.248
LaBSE-narrative	0.882 0.919	0.228 0.405
Entities-BoW	0.877	_

Result:P@1 retrieval scores are severely affected by anonymization.

• (Optionally) Translate all summaries to English using a model by NLLB-Team (2022)

20 60 40

80

Length in Sentences

	Movie Remakes		Translated Summaries	
Model Name	Text	Anonymized	Text	Anonymized
sentence-T5-large	0.644	0.461	0.905	0.686
all-mpnet-base-v2	0.609	0.300	0.925	0.382
LaBSE	0.476	0.204	0.865	0.329
LaBSE-narrative	0.622	0.339	0.925	0.528
Entities Bag of Words	0.410	-	0.850	-
Bag of Words Chaturvedi et al. (2018)	0.558	-	-	-
Full Model Chaturvedi et al. (2018)	0.637	-	-	-

P@1 retrieval scores on our dataset and the one by Chaturvedi et al. (2018). We sub-sample our dataset to replicate their distribution.



H.O. HATZEL, C. BIEMANN

INFORMATIK

LANGUAGE TECHNOLOGY GROUP