

Przedmiot wybieralny III - Przetwarzanie języka naturalnego

Etap 1: Web scrapping recenzji użytkowników (imdb)

Prowadzący:
dr inż. Jakub Osuchowski

Autorzy:
Paweł Siemiginowski 101450
Paweł Cedzich 101598

TECHNOLOGIE:



BeautifulSoup



Selenium
WebDriver

11,459 reviews ☐ Hide spoilers

+ Review

Ratings Show all Sort by Featured

★ 10/10

Some birds aren't meant to be caged. >

Spoiler

Helpful · 1.1K 93

hitchcockthelegend · Jul 24, 2010 · Permalink

★ 10/10

An incredible movie. One that lives with you. >

It is no wonder that the film has such a high rating, it is quite literally breathtaking. What can I say that hasn't said before? Not much, it's the story, the acting, the premise, but most of all, this movie is about how it makes you feel. Sometimes you watch a film, and can't remember it days later, this film lives with you, once you've seen it, you don't forget.

The ultimate story of friendship, of hope, and of life, and overcoming adversity.

I understand why so many class this as the best film of all time, it isn't mine, but I get it. If you haven't seen it, or haven't seen it for some time, you need to watch it, it's amazing. 10/10.

More to explore



Before They Were Famous: Actors' Early Roles

See the gallery

Editorial lists

Related lists created

Tanmay Bhat & The OJ Watchlist
updated 4 months ago ·

Top 100 Movies Buck

My favorite movie of all time. >

The Shawshank Redemption has great performances, extremely well written script and story all leading to a deeply emotional climax! One of the best dramas of all time!

Helpful · 127 28

OU · Feb 23, 2020 · Permalink

25 more All

Select Your Preferences

If you agree, we may use your personal information from the IMDb service (such as your past browsing history on IMDb) to personalize the ads we show you on other services. We will also personalize the ads we show you on IMDb using cookies and personal information, including personal information we may receive from third parties.

We use cookies and similar necessary tools to enhance your experiences and to provide our services, as detailed in our [Cookie Notice](#). Your choice applies to using first-party and third-party advertising cookies on this service. Cookies store or access standard device information such as a unique identifier. The 103 third parties who use cookies on Amazon services display and measure personalized ads, generate audience insights, and develop and improve products. Click "Decline" to decline, or "Customize" to make more detailed advertising choices, or learn more. You can change your choices at any time by visiting [cookies and advertising choices](#). To learn more about how and for what purposes Amazon uses personal information (such as Store order history or Prime Video Watch history) and cookies, please visit our [Privacy Notice](#) and Cookie Notice.

Decline Accept Customize

	title	content	rating	helpful_count	not_helpful_count
0	Some birds aren't meant to be caged.	No content	10	1.1K	93
1	An incredible movie. One that lives with you.	It is no wonder that the film has such a high ...	10	415	35
2	Don't Rent Shawshank.	I'm trying to save you money; this is the last...	10	1K	120
3	A classic piece of unforgettable film-making.	No content	10	1.3K	169
4	This is How Movies Should Be Made	This movie is not your ordinary Hollywood flic...	10	1K	133
5	Prepare to be moved	I have never seen such an amazing film since I...	No rating	1.6K	246
6	Shawshank Redeems Hollywood	Can Hollywood, usually creating things for ent...	10	1.7K	286
7	the shawshank redemption	The best movie in history and the best ending ...	10	164	19
8	Eternal Hope	I've lost count of the number of times I have ...	10	221	31
9	The Shawshank Redemption	One of the finest films made in recent years. ...	10	928	181
10	IMDb and the Greatest Film of All Time	Two imprisoned men (Tim Robbins and Morgan Fre...	10	110	15
11	Stephen King's best adapted movie	Misery and Stand By Me were the best adaptatio...	10	844	170
12	Mystery	I still can't quite get the enormous reputatio...	7	126	141
13	!BORING!	I do not know, what to say, I am just bored. M...	3	9	13
14	Enthralling, fantastic, intriguing, truly rema...	No content	10	181	33
15	Two movies in one	No content	No rating	1K	240
16	Simply known as "The Movie"	Whenever I talk about this movie with my frien...	10	629	144
17	Simply amazing. The best film of the 90's.	No content	10	604	170
18	All-time prison film classic	No content	10	61	11
19	The absolute GREATEST movie of all time.	No content	10	120	27
20	Simply the best	There's a reason this is rated the #1 . It is ...	10	22	2
21	"I'm a convicted murderer who provides sound f...	No content	10	52	10



A1				title		
	A	B	C	D	E	F
1	title	content	rating	helpful_cc	not_helpful_count	
2	Some bird	No conten	10	1.1K		93
3	An incredi	It is no wo	10	415		35
4	Don't Rent	I'm trying	10	1K		120
5	A classic	No conten	10	1.3K		169
6	This is Ho	This movi	10	1K		133
7	Prepare to	I have nev	No rating	1.6K		246
8	Shawshar	Can Holly	10	1.7K		286
9	the shaws	The best n	10	164		19
10	Eternal Hc	I've lost cc	10	221		31
11	The Shaw	One of the	10	928		181
12	IMDb and	Two impri	10	110		15
13	Stephen K	Misery an	10	844		170
14	Mystery	I still can't	7	126		141
15	!BORING!	I do not kr	3	9		13
16	Enthrallin	No conten	10	181		33
17	Two movi	No conten	No rating	1K		240
18	Simply kn	Whenever	10	629		144
19	Simply an	No conten	10	604		170
20	All-time p	No conten	10	61		11
21	The absol	No conten	10	120		27
22	Simply the	There's a	10	22		2
23	"I'm a con	No conten	10	52		10
24	Good mov	I was a litt	7	154		178
25	The Prisor	A fair num	3	9		13
26	My favorit	The Shaw	No rating	127		28
27	A Timeses	It is a time	10	10		0
28	My #1	I was relui	10	10		0
29	The best s	No conten	8	943		379
30	Finding B	It is a cine	10	8		0
31	The Great	Could a m	10	29		6
32	Masterpie	This for m	10	12		1
33	From my f	It is one of	10	63		19
34	Undeserv	An	7	29		28
35	Nothing s	This is jus	4	9		13
36	Hope can	One of my	10	50		15

Przedmiot wybieralny III - Przetwarzanie języka naturalnego

Etap 2: Korekcja błędów

Prowadzący:
dr inż. Jakub Osuchowski

Autorzy:
Paweł Siemiginowski 101450
Paweł Cedzich 101598

TECHNOLOGIE:



Transformers



Cornell University We gratefully acknowledge support from

arXiv > cs > arXiv:1702.04066 Search

Computer Science > Computation and Language

[Submitted on 14 Feb 2017]

JFLEG: A Fluency Corpus and Benchmark for Grammatical Error Correction

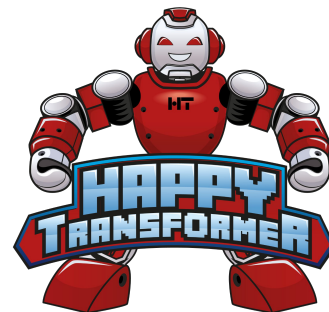
Courtney Napoles, Keisuke Sakaguchi, Joel Tetreault

We present a new parallel corpus, JHU Fluency-Extended GUG corpus (JFLEG) for developing and evaluating grammatical error correction (GEC). Unlike other corpora, it represents a broad range of language proficiency levels and uses holistic fluency edits to not only correct grammatical errors but also make the original text more native sounding. We describe the types of corrections made and benchmark four leading GEC systems on this corpus, identifying specific areas in which they do well and how they can improve. JFLEG fulfills the need for a new gold standard to properly assess the current state of GEC.

Comments: To appear in EACL 2017 (short papers)
Subjects: Computation and Language (cs.CL)
Cite as: arXiv:1702.04066 [cs.CL]
(or arXiv:1702.04066v1 [cs.CL] for this version)
<https://doi.org/10.48550/arXiv.1702.04066>

Submission history

From: Courtney Napoles [view email]
[v1] Tue, 14 Feb 2017 03:47:34 UTC (56 KB)



vennify/t5-base-grammar-correction

```
def correct_review(review):
    review = fix_spacing(review)
    sentences = sent_tokenize(review)
    corrected_sentences = []

    for sentence in sentences:
        inputs = tokenizer.encode("grammar: " + sentence, return_tensors="pt", max_length=512, truncation=True)
        outputs = model.generate(inputs, max_new_tokens=128)
        corrected_sentence = tokenizer.decode(outputs[0], skip_special_tokens=True)
        corrected_sentences.append(corrected_sentence)

    corrected_review = ' '.join(corrected_sentences)
    return corrected_review
```

content	corrected_content
this is a wierd text. So what can i see, if you see.there it is.	This is a strange text. So what can I see, if you see? There it is.
Hi, my name is Josh.i really... cant fix *****;) it.	Hi, my name is Josh. I really... can't fix it.

Text Compare!

↑ i watch this movie for minimum 5-6 times first to last. So every time i enjoyed as i watching it first time. Its a piece of art. and we now know how to entertain with no action or fighting scenes. Its a pure drama that anyone can enjoy with family in silence. So i rate this movie 10 out of 10 because of amazing screen play, acting, casting, background scoring editing etc. this is my first movie review so pardon me if i did any mistake.

↑ I watch this movie for minimum 5-6 times first to last. So every time I enjoyed watching it first time. It's a piece of art. And we now know how to entertain with no action or fighting scenes. It's a pure drama that anyone can enjoy with family in silence. So I rate this movie 10 out of 10 because of amazing screenplay, acting, casting, background scoring editing, etc. This is my first movie review, so forgive me if I made any mistakes.

Edit texts ...

Switch texts

Compare!

Clear all

↑ 1 My top film. Hope is good, and it never dies.

2 dīzī .dīzī .dīzī .

3

↑ 1 My top film. Hope is good, and it never dies.

2

Kolejny etap:
Analiza sentymentu recenzji użytkowników.

Dziękujemy za uwagę!

Pytania?