

Diese Seite wurde von der Cloud Translation API ([//cloud.google.com/translate/](https://cloud.google.com/translate/)) übersetzt.

cnn_dailymail

- **Beschreibung :**

Nicht anonymisierter CNN / DailyMail-Zusammenfassungsdatensatz.

Es gibt zwei Funktionen: - Artikel: Text des Nachrichtenartikels, der als zusammenzufassendes Dokument verwendet wird. - Hervorhebungen: Verbundener Text der Hervorhebungen mit `and` um jede Hervorhebung, die die Zielzusammenfassung darstellt

- **Homepage :** <https://github.com/abisee/cnn-dailymail>
(<https://github.com/abisee/cnn-dailymail>)
- **Quellcode :** [tfds.summarization.CnnDailymail](https://github.com/tensorflow/datasets/tree/master/tensorflow_datasets/summarization/cnn_dailymail.py)
(https://github.com/tensorflow/datasets/tree/master/tensorflow_datasets/summarization/cnn_dailymail.py)
- **Versionen :**
 - **1.0.0 :** Neue Split-API (<https://tensorflow.org/datasets/splits>)
(<https://tensorflow.org/datasets/splits>)
 - **2.0.0 :** Trennen Sie Zielsätze mit Zeilenumbruch. (Wenn das Modell Zeilenumbruchtrennzeichen vorhersagt, ist die Bewertung mit ROUGE auf Zusammenfassungsebene einfacher.)
 - **3.0.0 :** Verwenden der Gehäuseversion.
 - **3.1.0 (Standard):** BuilderConfig wurde entfernt
- **Download-Größe :** 558.32 MiB
- **Datensatzgröße :** 1.27 GiB
- **Automatisch zwischengespeichert** ([Dokumentation](#))
(<https://www.tensorflow.org/datasets/performances#auto-caching>) : Nein
- **Splits :**

Teilt	Beispiele
'test'	11.490
'train'	287,113
'validation'	13.368

- **Eigenschaften :**

```
FeaturesDict({
    'article': Text(shape=(), dtype=tf.string),
    'highlights': Text(shape=(), dtype=tf.string),
})
```

- **Überwachte Schlüssel** (Siehe [as_supervised doc](https://www.tensorflow.org/datasets/api_docs/python/tfds/load#args) (https://www.tensorflow.org/datasets/api_docs/python/tfds/load#args)): ('article', 'highlights')

- **Zitat :**

```
@article{DBLP:journals/corr/SeeLM17,
  author    = {Abigail See and
               Peter J. Liu and
               Christopher D. Manning},
  title     = {Get To The Point: Summarization with Pointer-Generator Networks},
  journal   = {CoRR},
  volume    = {abs/1704.04368},
  year      = {2017},
  url       = {http://arxiv.org/abs/1704.04368},
  archivePrefix = {arXiv},
  eprint    = {1704.04368},
  timestamp = {Mon, 13 Aug 2018 16:46:08 +0200},
  biburl    = {https://dblp.org/rec/bib/journals/corr/SeeLM17},
  bibsource = {dblp computer science bibliography, https://dblp.org}
}
```

```
@inproceedings{hermann2015teaching,
  title={Teaching machines to read and comprehend},
  author={Hermann, Karl Moritz and Kocisky, Tomas and Grefenstette, Edward and E
  booktitle={Advances in neural information processing systems},
```

```
pages={1693--1701},  
year={2015}  
}
```

- **Abbildung** ([tfds.show_examples](https://www.tensorflow.org/datasets/api_docs/python/tfds/visualization/show_examples)

(https://www.tensorflow.org/datasets/api_docs/python/tfds/visualization/show_examples)): Wird (https://www.tensorflow.org/datasets/api_docs/python/tfds/visualization/show_examples) nicht unterstützt.

- **Beispiele** ([tfds.as_dataframe](https://www.tensorflow.org/datasets/api_docs/python/tfds/as_dataframe)

(https://www.tensorflow.org/datasets/api_docs/python/tfds/as_dataframe)):

[Beispiele anzeigen ...](#)

Except as otherwise noted, the content of this page is licensed under the [Creative Commons Attribution 4.0 License](https://creativecommons.org/licenses/by/4.0/) (<https://creativecommons.org/licenses/by/4.0/>), and code samples are licensed under the [Apache 2.0 License](https://www.apache.org/licenses/LICENSE-2.0) (<https://www.apache.org/licenses/LICENSE-2.0>). For details, see the [Google Developers Site Policies](https://developers.google.com/site-policies) (<https://developers.google.com/site-policies>). Java is a registered trademark of Oracle and/or its affiliates.

Last updated 2021-01-07 UTC.