

読解型質問応答用データセット

SQuAD

Stanford Question Answering Dataset

1. SQuAD の読解課題と版ごとの違い
2. 質問・回答の作成方法と収録規模
3. 読解の評価指標とモデル性能
4. SQuAD の利用動向と後続データ
5. 質問応答データの入手方法と利用条件

1. SQuAD の読解課題と版ごとの違い

SQuAD の公式サイト

The screenshot shows the SQuAD2.0 website. The header is purple with 'SQuAD' on the left and navigation links 'Home', 'Explore 2.0', and 'Explore 1.1' on the right. The main title 'SQuAD2.0' is in large white font, with the subtitle 'The Stanford Question Answering Dataset' below it. The page is divided into two columns. The left column, titled 'What is SQuAD?', contains a paragraph about the dataset and a button 'Explore SQuAD2.0 and model predictions'. The right column, titled 'Leaderboard', contains a paragraph about the dataset's purpose and a table of top-performing models.

What is SQuAD?

Stanford Question Answering Dataset (SQuAD) is a reading comprehension dataset, consisting of questions posed by crowdworkers on a set of Wikipedia articles, where the answer to every question is a segment of text, or *span*, from the corresponding reading passage, or the question might be unanswerable.

SQuAD2.0 combines the 100,000 questions in SQuAD1.1 with over 50,000 unanswerable questions written adversarially by crowdworkers to look similar to answerable ones. To do well on SQuAD2.0, systems must not only answer questions when possible, but also determine when no answer is supported by the paragraph and abstain from answering.

[Explore SQuAD2.0 and model predictions](#)

[SQuAD2.0 paper \(Rajpurkar & Jia et al. '18\)](#)

Leaderboard

SQuAD2.0 tests the ability of a system to not only answer reading comprehension questions, but also abstain when presented with a question that cannot be answered based on the provided paragraph.

Rank	Model	EM	F1
	Human Performance Stanford University (Rajpurkar & Jia et al. '18)	86.831	89.452
1 Jun 04, 2021	IE-Net (ensemble) RICOH_SRCB_DML	90.939	93.214
2 Feb 21, 2021	FPNet (ensemble) Ant Service Intelligence Team	90.871	93.183
3 May 16, 2021	IE-NetV2 (ensemble) RICOH_SRCB_DML	90.860	93.100
4 Apr 06, 2020	SA-Net on Albert (ensemble) QIANXIN	90.724	93.011

段落と質問を入力し答えの範囲を選ぶ

- 読解型質問応答は、文章を根拠に質問へ答える課題である。SQuAD では、英語 Wikipedia の記事から選んだ 1 段落と、その段落に対する質問を入力する。
- 答えは、段落内にある連続した語句を選ぶ。固有名詞・年・数量に限らず、複数語からなる表現も正解となる。
- バージョン 2.0 には段落から答えられない質問もあり、答えの抽出に加えて回答を控える判断を行う。

[1] Rajpurkar et al. (2016). EMNLP, pp. 2383–2392. SQuAD: 100,000+ Questions for Machine Comprehension of Text. <https://aclanthology.org/D16-1264/>

[2] Rajpurkar et al. (2018). ACL, pp. 784–789. Know What You Don't Know: Unanswerable Questions for SQuAD. <https://aclanthology.org/P18-2124/>

記事の中の段落に複数の質問と正解が付く

- 配布データの全体は記事の集合である。各記事には複数の段落があり、それぞれの段落に質問と正解回答が付く。
- context は入力となる 1 段落、question はその段落への質問、answers は正解の文字列と段落内の開始位置を表す。
- モデルには、その質問に対応する段落を渡す。

実例 1 スポーツ記事の質問

Wikipedia 記事の段落（抜粋）

The American Football Conference (AFC) champion **Denver Broncos** defeated the National Football Conference (NFC) champion Carolina Panthers 24–10 to earn their third Super Bowl title. The game was played on February 7, 2016, at Levi's Stadium in the San Francisco Bay Area at Santa Clara, California.

質問 Which NFL team represented the AFC at Super Bowl 50?

質問の日本語訳 第 50 回スーパーボウルで AFC を代表したチームは？

正解は Denver Broncos

実例 2 人物記事の質問

Wikipedia 記事の段落（抜粋）

Tesla gained experience in telephony and electrical engineering before emigrating to the United States in **1884** to work for Thomas Edison in New York City. He soon struck out on his own with financial backers, setting up laboratories and companies to develop a range of electrical devices.

質問 In what year did Nikola Tesla emigrate to the United States?

質問の日本語訳 ニコラ・テスラがアメリカへ移住したのは何年か？

正解は 1884

実例 3 技術記事の質問

Wikipedia 記事の段落（抜粋）

Steam engines are **external combustion** engines, where the working fluid is separate from the combustion products. Non-combustion heat sources such as solar power, nuclear power or geothermal energy may be used.

質問 What types of engines are steam engines?

質問の日本語訳 蒸気機関はどのような種類の機関か？

正解は external combustion

実例 4 文章から回答できない質問

Wikipedia 記事の段落（抜粋）

They were descended from Norse ("Norman" comes from "Norseman") raiders and pirates from Denmark, Iceland and Norway who, under their leader Rollo, agreed to swear fealty to King Charles III of West Francia.

質問 Who did King Charles III swear fealty to?

質問の日本語訳 シャルル 3 世は誰に忠誠を誓ったか？

正解は 回答不能（文章に根拠がない）

Rollo の側が王に忠誠を誓う。王が誰に誓ったかは記載されていない。

SQuAD のバージョン 1.1 と 2.0

- バージョン 1.1 は、文章に答えがある質問を扱う [1]。
 - 質問に対応する本文中の範囲を回答する
- バージョン 2.0 では、回答不能な質問を追加する [2]。
 - 既存の質問に似た、答えられない質問を人が作成。
 - 答えられるかを判定し、答えがなければ回答を控える。

[1] Rajpurkar et al. (2016). EMNLP, pp. 2383–2392. SQuAD: 100,000+ Questions for Machine Comprehension of Text. <https://aclanthology.org/D16-1264/>

[2] Rajpurkar et al. (2018). ACL, pp. 784–789. Know What You Don't Know: Unanswerable Questions for SQuAD. <https://aclanthology.org/P18-2124/>

2. 質問・回答の作成方法と収録規模

MCTest は物語を読んで 4 択の質問に答える

- MCTest（2013）は、短い物語を読み、4つの選択肢から質問の答えを選ぶ読解データセットである。
- これに対し SQuAD は、百科事典 Wikipedia の段落を読み、答えの表現を本文中から選ぶ。
- 選択肢を選ぶ形式から、段落内の任意の連続した回答範囲を特定する形式へ変わる。

[1] Richardson et al. (2013). EMNLP. MCTest: A Challenge Dataset for the Open-Domain Machine Comprehension of Text. <https://aclanthology.org/D13-1020/>

[2] Rajpurkar et al. (2016). EMNLP, pp. 2383–2392. SQuAD: 100,000+ Questions for Machine Comprehension of Text. <https://aclanthology.org/D16-1264/>

ニュースから穴埋め問題を作る CNN / Daily Mail

- CNN / Daily Mail (2015) は、ニュースの要約文で固有表現を伏せ、元記事を根拠に埋める問題を自動作成した。
- これに対し SQuAD では、人が段落を読んで質問を書き、固有表現に限らない回答範囲を選ぶ。
- 人手による質問の多様性と、ニューラルモデルを学習・評価できる規模の両立を目指した。

人手による多様な読解問題を大規模に収集する

- 先行する読解データは、人手作成では小規模、自動作成では質問や正解の形式が限定されていた。
- SQuAD は、Wikipedia の段落に人が質問と回答範囲を付け、自然な読解問題を大量に用意することを目的とした。
- 2016 年の原著は、500 以上の記事から 10 万問以上を作成したと報告した。

記事を選び段落単位で質問の素材にする

- 英語 Wikipedia の記事を、リンク関係に基づく重要度 PageRank で順位付けし、上位 10,000 記事から 536 記事が無作為に選んだ。
- 原著は、各記事から段落を取り出し、500 文字未満を除いたと説明する。学習・開発・テストは記事単位で分けた。
- 公開バージョン 1.1 の学習用・開発用には、20,963 段落のうち 500 文字未満の段落が 2,874 件ある。

[1] Rajpurkar et al. (2016). EMNLP, pp. 2383–2392. SQuAD: 100,000+ Questions for Machine Comprehension of Text. <https://aclanthology.org/D16-1264/>

[2] <https://github.com/rajpurkar/SQuAD-explorer/tree/master/dataset>

段落を読んで質問と最小限の回答範囲を作る

- オンラインで作業を依頼する Amazon Mechanical Turk を使い、作業者へ段落を提示した。
- 作業者は、その段落だけから答えられる質問を最大 5 問作った。段落の文をコピーせず、自分の言葉で質問を書くよう指示された。
- 答えとなる最小限の連続した範囲を段落内で選んだ。この範囲が、回答位置を予測するモデルの学習用の正解となる。

追加の正解回答で表現の違いを許容する

- 開発用・テスト用では、作成済みの質問へ別の作業者が追加の回答範囲を指定した。
- 同じ質問にも、長さや表現が異なる妥当な回答がありうるため、1問に複数の正解を持たせる。
- 評価では各正解との一致を計算し、その中で最も高い得点を質問の得点とする。

バージョン 2.0 を作った理由と作成方法

- 初版では、どの質問にも段落内に答えがあった [1]。
- 作業者が各段落に対し、回答不能な質問を最大 5 問作成した。
 - 段落の話題に関連し、回答可能な質問に見えるようにする。
 - 段落内に、答えに見える候補がある質問にする。
- 記事・段落ごとの作業を分配し、人手で 50,000 問以上を追加。
- 回答範囲の抽出と、回答保留の判断を評価する。

[1] Rajpurkar et al. (2018). ACL, pp. 784–789. Know What You Don't Know: Unanswerable Questions for SQuAD. §1, §3 <https://aclanthology.org/P18-2124/>

SQuAD 各版の質問数・回答不能数・記事数

版・分割 [1]	質問数	回答不能	Wikipedia 記事数
1.1 ・学習用	87,599	0	442
1.1 ・開発用	10,570	0	48
2.0 ・学習用	130,319	43,498	442
2.0 ・開発用	11,873	5,945	35

- 記事数は、質問の元となった Wikipedia 記事の数。
- バージョン 2.0 は同じ記事に回答不能な質問を追加。

質問に大量の空白が含まれる実例

- 前後に大量の空白が含まれる質問を 1 件確認 [1]。
 - 元文字列 25,651 文字 → 前後空白を除去 59 文字。
- What radiates two lobes perpendicular to the antennas axis?
 - アンテナの軸に垂直な 2 つのローブを放射するものは何か。
- 主要統計は前後空白だけを除去し、本文の文字は保持。

[1] <https://github.com/rajpurkar/SQuAD-explorer/tree/master/dataset> ID 572fdefb947a6a140053cd8d

長い入力段落の例 Sahara

- 段落文字数 3706 ／ 段落単語数 653 [1]。
- 質問 What is the largest hottest continuously large area worldwide?
- 正解は the Great Desert
- 対象の版と分割は 1.1/train 、 2.0/train 。

長い段落の原文抜粋 Sahara

Wikipedia 記事の段落からの抜粋。

A value of 4,300 h or 98% of the time would be recorded in Upper Egypt (Aswan, Luxor) and in the Nubian Desert (Wadi Halfa). The annual average direct solar irradiation is around 2,800 kWh/(m² year) in the Great Desert. The Sahara has a huge potential for solar energy production. The constantly high position of the sun, the extremely low relative humidity, the lack of vegetation and rainfall make the Great Desert the hottest continuously large area worldwide and certainly the hottest place on Earth during summertime in some spots.

入力段落（context）の長さ

context の長さ [1]	平均	中央値	95% 点	最大
文字数	739.6	679.0	1312.0	4063
単語数	117.2	108.0	210.0	653

- バージョン 1.1 train + dev、20,963 段落を各 1 回集計。
- 最大 4,063 文字／653 単語。別々の段落。

[1] 文字数は前後空白を除去／単語数は空白分割。

<https://github.com/rajpurkar/SQuAD-explorer/tree/master/dataset>

長い質問の例 Age of Enlightenment

- 質問文字数 270 / 質問単語数 40 [1]
- 質問 Which century is considered by some as a "drab" period, however saw significant advancements in the practice of medicine, mathematics, and physics, the development of biological taxonomy, a new understanding of magnetism, and the maturation of chemistry as a discipline?
- 正解は the 18th century
- 対象の版と分割は 1.1/train 、 2.0/train 。

長い質問の根拠となる原文抜粋

Wikipedia 記事の段落からの抜粋。

Philosophes introduced the public to many scientific theories, most notably through the Encyclopédie and the popularization of Newtonianism by Voltaire and Émilie du Châtelet. Some historians have marked the 18th century as a drab period in the history of science; however, the century saw significant advancements in the practice of medicine, mathematics, and physics; the development of biological taxonomy; a new understanding of magnetism and electricity; and the maturation of chemistry as a discipline, which established the foundations of modern chemistry.

質問と回答の長さ

長さ [1]	平均	中央値	95% 点	最大
質問の文字数	59.3	56.0	98.0	270
質問の単語数	10.1	10.0	17.0	40
回答の文字数	20.1	14.0	63.0	239
回答の単語数	3.1	2.0	10.0	43

- 公開データ 98,169 問の先頭正解を 1 問 1 件で集計。
- 最大の回答は 43 単語・239 文字。

[1] バージョン 1.1 の train と dev を集計。文字数は前後空白を除去し、単語数は空白で分割。
<https://github.com/rajpurkar/SQuAD-explorer/tree/master/dataset>

3. 読解の評価指標とモデル性能

完全一致と単語の重なりを別々に評価する

- SQuAD の評価には、回答文字列が一致するかを測る Exact Match (EM) と、単語の重なりを測る F1 がある。
- EM は、予測が正解に完全一致すれば 1、それ以外は 0 である。F1 は、一致した単語の割合を予測側と正解側で計算して組み合わせる。
- どちらも、各質問で得た 0 ~ 1 の得点を全質問について平均し、百分率で報告する。

[1] https://github.com/huggingface/evaluate/blob/main/metrics/squad/compute_score.py

Denver だけを答えると部分一致になる

予測（正解は Denver Broncos）	EM	F1
Denver Broncos	1	1
Denver	0	2/3
Carolina Panthers	0	0

Denver だけなら、予測 1 語中 1 語が一致し、正解 2 語中 1 語を含む。適合率 1 と再現率 1/2 の調和平均は 2/3 となる。

[1] https://github.com/huggingface/evaluate/blob/main/metrics/squad/compute_score.py

表記の違いを正規化して複数正解と照合する

- 採点では、小文字化、句読点の除去、英語の冠詞 a ・ an ・ the の除去、空白の正規化を行う。
- F1 の適合率 P は一致語数 ÷ 予測の語数、再現率 R は一致語数 ÷ 正解の語数である。両者の調和平均 $2PR \div (P + R)$ を使う。
- 複数正解がある質問は最良の得点を採用する。バージョン 2.0 では、回答不能な質問に回答を控えた場合を正解として採点する。

[1] https://github.com/huggingface/evaluate/blob/main/metrics/squad/compute_score.py

[2] https://github.com/huggingface/evaluate/blob/main/metrics/squad_v2/squad_v2.py

2016 年の原著のモデルは F1 が 51.0% だった

原著のモデル・人の回答	EM (%)	F1 (%)
単語一致を使う Sliding Window	12.5	19.7
ロジスティック回帰	40.4	51.0
人の回答	77.0	86.8

SQuAD バージョン 1.0 のテスト用で評価した [1]。人の得点は、追加の正解回答を予測として採点した。

バージョン 1.1 のベンチマーク結果

モデル [1]	提出日	EM	F1
BiDAF	2016-11-28	68.0	77.3
BERT	2018-10-05	85.1	91.8
XLNet	2019-05-21	89.9	95.1

- SQuAD 運営者の非公開テストで採点。
- 各行は単一モデルの提出結果。

過去の抽出型モデルの評価結果

モデル [1]	EM (%)	F1 (%)
BERT-Large (2018)	79.0	81.8
RoBERTa-Large (2019)	86.5	89.4
DeBERTa-V3-Large (2021)	89.0	91.5

- SQuAD 2.0 の学習用で微調整し、開発用で評価した公表値。答えの開始・終了位置を予測する抽出型モデルを比較する。

Llama 3.1 でも SQuAD 2 を評価している

事前学習モデル [1]	EM (%)
Llama 3.1 8B	77.0
Llama 3.1 70B	81.8
Llama 3.1 405B	89.3

- 2024-07-23 公開、答え付きの例を 1 件示す 1-shot で生成評価。
- 同じ段落に付いた先行する全質問と正解も入力に含める [2]。

[1] https://github.com/meta-llama/llama-models/blob/main/models/llama3.1/MODEL_CARD.md

[2] https://github.com/meta-llama/llama-models/blob/main/models/llama3.1/eval_details.md

回答範囲を選ぶ方式と文章を生成する方式がある

- 抽出型は、段落中の開始・終了位置を予測する。読解モデル BiDAF や言語モデル BERT を使い、原文中の回答範囲を選ぶ。
- 生成型は、段落と質問を受け取り、答えの文字列を生成する。大規模言語モデル Llama も、この方式で評価される。
- 抽出型は原文中の位置を、生成型は出力する語句を予測する。どちらも、得られた回答文字列を EM ・ F1 で採点する。

[1] Devlin et al. (2019). NAACL-HLT, pp. 4171–4186. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding.

<https://aclanthology.org/N19-1423/>

[2] https://github.com/meta-llama/llama-models/blob/main/models/llama3.1/eval_details.md

SQuAD による微調整と例示付きの生成評価

- 微調整では、SQuAD の学習用の質問と正解を使い、モデルのパラメータを更新する。
- 1-shot では、答え付きの例を 1 件入力に提示してから別の質問へ答えさせる。この提示だけではモデルのパラメータを更新しない。

4. SQuAD の利用動向と後続データ

HotpotQA は複数記事の情報を組み合わせる

- HotpotQA (2018) は、複数の Wikipedia 記事にまたがる推論と、回答の根拠となる文の特定を扱う。
- SQuAD が基本的に 1 段落を根拠とするのに対し、HotpotQA では文書をまたいで情報を組み合わせる。
- 回答文字列の正しさに加え、根拠文を選ぶかも評価対象となる。

Natural Questions は実際の検索質問を使う

- Natural Questions （2019）は、Google 検索に入力された質問と、その回答を含む Wikipedia ページを使う。
- 回答を含む段落などの長い回答と、必要に応じて語句や固有名詞などの短い回答を指定する。
- SQuAD のように段落を見て質問を作るのではなく、実際の情報要求とページ全体を対応付ける。

SQuAD は生成型モデルの読解評価にも使われる

- SQuAD は、回答範囲を選ぶモデルから、答えを生成する大規模言語モデルの評価へも用途が広がった。
- 2024 年の Llama 3.1 の評価にも SQuAD 2 が含まれる。
- 2026-10-15 の確認時点でも、モデル評価ライブラリ Hugging Face Evaluate は SQuAD 2.0 の採点実装を公開している。

[1] https://github.com/meta-llama/llama-models/blob/main/models/llama3.1/MODEL_CARD.md

[2] https://github.com/huggingface/evaluate/blob/main/metrics/squad_v2/squad_v2.py

LongBench v2 はより長い文脈から回答を選ぶ

- LongBench v2 （ 2024 ） は、 8 千～ 200 万語の長い文脈を与え、選択肢から回答を選ぶ評価データである。
- SQuAD の短い段落からの回答範囲抽出と比べ、文脈長と回答形式が異なる。

5. 質問応答データの入手方法と利用条件

バージョンごとのJSON と評価実装を入手する

- 公式サイトは、構造化データ形式JSON で学習用 train と開発用 dev を配布する。バージョン 1.1 と 2.0 のファイルを選んで取得する。
- データ配布サービス Hugging Face Datasets では、1.1 は rajpurkar/squad 、 2.0 は rajpurkar/squad_v2 という名前で取得できる。
- データの利用条件は CC BY-SA 4.0 である。質問への予測は、対応するバージョンの公式採点実装で評価する。

[1] <https://rajpurkar.github.io/SQuAD-explorer/>

[2] <https://creativecommons.org/licenses/by-sa/4.0/>

SQuAD の目的・版の違いと評価方法

- SQuAD は、英語の読解型質問応答データセット [1]。
 - Wikipedia の段落から、質問の答えとなる範囲を選ぶ。
- 人手作成の質問を大規模に評価するため、2016 年に発表。
- バージョン 1.1 では、回答可能な質問の答えを抽出。
- バージョン 2.0 は回答不能な質問と回答保留も扱う [2]。
- EM ・ F1 で評価し、大規模言語モデルの読解評価にも用いられる。

Rajpurkar et al. (2016). EMNLP, pp. 2383–2392. SQuAD: 100,000+ Questions for Machine Comprehension of Text. <https://aclanthology.org/D16-1264/>

Rajpurkar et al. (2018). ACL, pp. 784–789. Know What You Don't Know: Unanswerable Questions for SQuAD. <https://aclanthology.org/P18-2124/>

Appendix

1. 開発者の研究と発表先
2. 段落・質問・回答の追加例
3. 段落・質問・回答の統計と極値
4. SQuAD 2.0 のモデル評価結果
5. 収録された Wikipedia 記事の全一覧
6. 公式 JSON の取得コードと関連文献

SQuAD 各版の著者と発表先

- 初版は Pranav Rajpurkar らが EMNLP 2016 で発表した [1]。
- バージョン 2 は、2018 年の ACL で発表された [2]。
- 両論文の第一著者は Pranav Rajpurkar、最終著者は Percy Liang である。
- 2026 年に初版が ACL Test of Time Award を受賞した [1]。

[1] Rajpurkar et al. (2016). EMNLP, pp. 2383–2392. SQuAD: 100,000+ Questions for Machine Comprehension of Text. <https://aclanthology.org/D16-1264/>

[2] Rajpurkar et al. (2018). ACL, pp. 784–789. Know What You Don't Know: Unanswerable Questions for SQuAD. <https://aclanthology.org/P18-2124/>

筆頭・最終著者の研究分野と関連実績

- 第一著者の Pranav Rajpurkar は、読解型 QA と医療 AI を研究している [1,2]。
- SQuAD と、胸部 X 線のデータセット CheXpert に貢献した（2019 年） [5]。
- 最終著者の Percy Liang は、自然言語処理・機械学習を研究している。
- 言語モデルを幅広く評価する HELM にも取り組んだ [3]。

[1] Rajpurkar et al. (2016). EMNLP, pp. 2383–2392. SQuAD: 100,000+ Questions for Machine Comprehension of Text. <https://aclanthology.org/D16-1264/>

[2] Rajpurkar et al. (2018). ACL, pp. 784–789. Know What You Don't Know: Unanswerable Questions for SQuAD. <https://aclanthology.org/P18-2124/>

[3] <https://github.com/stanford-crfm/helm>

[5] <https://github.com/stanfordmlgroup/chexpert-labeler>

補足実例 1 Apollo program

Wikipedia 記事の段落（抜粋）

The first manned flight of Apollo was in **1968**.

質問 What year did the first manned Apollo flight occur?

質問の日本語訳 アポロ計画の最初の有人飛行が行われたのは何年か？

正解は 1968

補足実例 2 Amazon rainforest

Wikipedia 記事の段落（抜粋）

This region includes territory belonging to **nine nations**. The majority of the forest is contained within Brazil, with 60% of the rainforest, followed by Peru with 13%, Colombia with 10%, and with minor amounts in Venezuela, Ecuador, Bolivia, Guyana, Suriname and French Guiana.

質問 How many nations are within the Amazon Basin?

質問の日本語訳 アマゾン盆地には何か国が含まれるか？

正解は nine nations

補足実例 3 Prime number

Wikipedia 記事の段落（抜粋）

A prime number (or a prime) is a natural number greater than 1 that has no positive divisors other than 1 and **itself**. A natural number greater than 1 that is not a prime number is called a composite number.

質問 What is the only divisor besides 1 that a prime number can have?

質問の日本語訳 素数が 1 以外に持つことのできる約数は何か？

正解は itself

補足実例 4 Black Death

Wikipedia 記事の段落（抜粋）

Spreading throughout the Mediterranean and Europe, the Black Death is estimated to have killed **30–60% of Europe's total population**. In total, the plague reduced the world population from an estimated 450 million down to 350–375 million in the 14th century.

質問 How much of the European population did the black death kill?

質問の日本語訳 黒死病でヨーロッパ人口のどのくらいが死亡したか？

正解は 30–60% of Europe's total population

補足実例 5 Warsaw

Wikipedia 記事の段落（抜粋）

One of the most famous people born in Warsaw was Maria Skłodowska-Curie, who achieved international recognition for her research on radioactivity and was the first female recipient of the **Nobel Prize**. Famous musicians include Władysław Szpilman and Frédéric Chopin.

質問 What was Maria Curie the first female recipient of?

質問の日本語訳 マリ・キュリーが女性として初めて受賞した賞は何か？

正解は Nobel Prize

補足実例 6 Genghis Khan

Wikipedia 記事の段落（抜粋）

After founding **the Mongol Empire** and being proclaimed "Genghis Khan", he started the Mongol invasions that resulted in the conquest of most of Eurasia. These included raids or invasions of the Qara Khitai, Caucasus, Khwarezmid Empire, Western Xia and Jin dynasties.

質問 What do we call the empire that Genghis Khan founded?

質問の日本語訳 チングス・ハンが建てた帝国を何と呼ぶか？

正解は the Mongol Empire

補足実例 7 Chloroplast

Wikipedia 記事の段落（抜粋）

Chloroplasts' main role is **to conduct photosynthesis**, where the photosynthetic pigment chlorophyll captures the energy from sunlight and converts it and stores it in the energy-storage molecules ATP and NADPH while freeing oxygen from water.

質問 What is the primary purpose of chloroplasts?

質問の日本語訳 葉緑体の主な役割は何か？

正解は to conduct photosynthesis

補足実例 8 University of Chicago

Wikipedia 記事の段落（抜粋）

The University of Chicago (UChicago, Chicago, or U of C) is **a private research university** in Chicago. The university, established in 1890, consists of The College, various graduate programs, interdisciplinary committees organized into four academic research divisions and seven professional schools.

質問 What kind of university is the University of Chicago?

質問の日本語訳 シカゴ大学はどのような種類の大学か？

正解は a private research university

補足実例 9 Steam engine

Wikipedia 記事の段落（抜粋）

Steam engines are external combustion engines, where the working fluid is separate from the combustion products.

質問 What types of engines are solar engines?

質問の日本語訳 太陽熱機関はどのような種類の機関か？

正解は 回答不能（文章に根拠がない）

原文は蒸気機関を説明する。太陽熱機関の種類は書かれていない。

補足実例 10 Oxygen

Wikipedia 記事の段落（抜粋）

Diatomic oxygen gas constitutes 20.8% of the Earth's atmosphere.

質問 What constitutes 28.0% of the Earth's atmosphere?

質問の日本語訳 地球の大気の 28.0% を占めるものは何か？

正解は 回答不能（文章に根拠がない）

原文の割合は 20.8% 。 28.0% に対応する記述はない。

記述統計 バージョン 1.1 ・学習用

質問 87,599 件／回答あり 87,599 件／回答不能 0 件

指標 [1]	平均	標準偏差	最小	25% 点	中央値	75% 点	95% 点	最大
質問文字	59.2	21.2	1	44.0	56.0	71.0	98.0	270
質問単語	10.1	3.6	1	8.0	10.0	12.0	17.0	40
回答文字	20.1	21.6	1	7.0	14.0	23.0	63.0	239
回答単語	3.2	3.4	1	1.0	2.0	3.0	10.0	43

- 文字数は前後の空白だけを除いた文字列の長さ。
- 単語数は空白区切り。回答は先頭の正解を 1 問 1 件、回答不能は除外。

[1] 公式 JSON を全件集計 <https://github.com/rajpurkar/SQuAD-explorer/tree/master/dataset>

記述統計 バージョン 1.1 ・ 開発用

質問 10,570 件／回答あり 10,570 件／回答不能 0 件

指標 [1]	平均	標準偏差	最小	25% 点	中央値	75% 点	95% 点	最大
質問文字	60.0	21.2	12	45.0	57.0	72.0	99.0	197
質問単語	10.2	3.6	3	8.0	10.0	12.0	17.0	33
回答文字	19.4	19.3	1	8.0	14.0	23.0	57.0	160
回答単語	3.0	3.0	1	1.0	2.0	3.0	9.0	29

- 文字数は前後の空白だけを除いた文字列の長さ。
- 単語数は空白区切り。回答は先頭の正解を 1 問 1 件、回答不能は除外。

[1] 公式 JSON を全件集計 <https://github.com/rajpurkar/SQuAD-explorer/tree/master/dataset>

記述統計 バージョン 2.0 ・ 学習用

質問 130,319 件／回答あり 86,821 件／回答不能 43,498 件

指標 [1]	平均	標準偏差	最小	25% 点	中央値	75% 点	95% 点	最大
質問文字	58.3	20.4	1	44.0	55.0	69.0	96.0	270
質問単語	9.9	3.4	1	7.0	9.0	12.0	16.0	40
回答文字	20.1	21.6	1	7.0	14.0	23.0	63.0	239
回答単語	3.2	3.4	1	1.0	2.0	4.0	10.0	43

- 文字数は前後の空白だけを除いた文字列の長さ。
- 単語数は空白区切り。回答は先頭の正解を 1 問 1 件、回答不能は除外。

[1] 公式 JSON を全件集計 <https://github.com/rajpurkar/SQuAD-explorer/tree/master/dataset>

記述統計 バージョン 2.0 ・ 開発用

質問 11,873 件／回答あり 5,928 件／回答不能 5,945 件

指標 [1]	平均	標準偏差	最小	25% 点	中央値	75% 点	95% 点	最大
質問文字	59.4	20.8	11	45.0	57.0	71.0	98.0	197
質問単語	10.0	3.5	3	8.0	10.0	12.0	16.0	31
回答文字	20.9	21.4	1	8.0	14.0	24.0	64.0	160
回答単語	3.2	3.3	1	1.0	2.0	4.0	10.0	29

- 文字数は前後の空白だけを除いた文字列の長さ。
- 単語数は空白区切り。回答は先頭の正解を 1 問 1 件、回答不能は除外。

[1] 公式 JSON を全件集計 <https://github.com/rajpurkar/SQuAD-explorer/tree/master/dataset>

回答長の集計単位 バージョン 1.1

対象・指標 [1]	件数	平均	中央値	95% 点	最大
1 問 1 件・文字	98,169	20.1	14.0	63.0	239
1 問 1 件・単語	98,169	3.1	2.0	10.0	43
全正解注釈・文字	122,325	19.7	14.0	61.0	239
全正解注釈・単語	122,325	3.1	2.0	9.0	43

- 学習用＋開発用を集計。回答不能の質問は除外。
- 全正解注釈では、同じ質問の複数正解・重複注釈も数える。

回答長の集計単位 バージョン 2.0

対象・指標 [1]	件数	平均	中央値	95% 点	最大
1 問 1 件・文字	92,749	20.2	14.0	63.0	239
1 問 1 件・単語	92,749	3.2	2.0	10.0	43
全正解注釈・文字	107,123	20.1	14.0	63.0	239
全正解注釈・単語	107,123	3.1	2.0	10.0	43

- 学習用＋開発用を集計。回答不能の質問は除外。
- 全正解注釈では、同じ質問の複数正解・重複注釈も数える。

入力段落の記述統計 バージョン 1.1 ・ train

context [1]	件数	平均	SD	中央値	95% 点	最大
文字数	18896	735.8	310.0	678.0	1311.0	3706
単語数	18896	116.6	49.7	107.0	210.0	653

- 各段落を 1 回ずつ集計。
- 文字数は前後空白除去／単語数は空白分割。

[1] SD は母標準偏差。95% 点は 95 パーセンタイル。

<https://github.com/rajpurkar/SQuAD-explorer/tree/master/dataset>

入力段落の記述統計 バージョン 1.1 ・ dev

context [1]	件数	平均	SD	中央値	95% 点	最大
文字数	2067	774.3	342.9	700.0	1338.0	4063
単語数	2067	122.8	54.8	111.0	213.0	629

- 各段落を 1 回ずつ集計。
- 文字数は前後空白除去／単語数は空白分割。

入力段落の記述統計 バージョン 2.0 ・ train

context [1]	件数	平均	SD	中央値	95% 点	最大
文字数	19035	735.5	309.8	677.0	1310.3	3706
単語数	19035	116.6	49.7	107.0	209.0	653

- 各段落を 1 回ずつ集計。
- 文字数は前後空白除去／単語数は空白分割。

[1] SD は母標準偏差。95% 点は 95 パーセンタイル。

<https://github.com/rajpurkar/SQuAD-explorer/tree/master/dataset>

入力段落の記述統計 バージョン 2.0 ・ dev

context [1]	件数	平均	SD	中央値	95% 点	最大
文字数	1204	802.6	359.6	711.5	1425.9	4063
単語数	1204	126.6	57.4	113.0	223.8	629

- 各段落を 1 回ずつ集計。
- 文字数は前後空白除去／単語数は空白分割。

[1] SD は母標準偏差。95% 点は 95 パーセンタイル。

<https://github.com/rajpurkar/SQuAD-explorer/tree/master/dataset>

長い回答の例 2008 Sichuan earthquake

- 回答文字数 239 ／ 回答単語数 43 [1]。
- 質問 What was concluded about the construction?
- 正解は that the sudden shift of a huge quantity of water into the region could have relaxed the tension between the two sides of the fault, allowing them to move apart, and could have increased the direct pressure on it, causing a violent rupture
- 対象の版と分割は 1.1/train 、 2.0/train 。

長い回答の原文抜粋 2008 Sichuan earthquake

Wikipedia 記事の段落からの抜粋。

An article in Science suggested that the construction and filling of the Zipingpu Dam may have triggered the earthquake. The chief engineer of the Sichuan Geology and Mineral Bureau said that the sudden shift of a huge quantity of water into the region could have relaxed the tension between the two sides of the fault, allowing them to move apart, and could have increased the direct pressure on it, causing a violent rupture. The effect was "25 times more" than a year's worth of natural stress from tectonic movement.

開発用で長い質問の例 Construction

- 質問文字数 197 [1]
- 質問 What kind of contract is given when the contractor is given a performance specification and must undertake the project from design to construction, while adhering to the performance specifications?
- 正解は "design build" contract
- 対象の版と分割は 1.1/dev 、 2.0/dev 。

長い質問の原文抜粋 Construction

Wikipedia 記事の段落からの抜粋。

Presently, a firm that is nominally an "architecture" or "construction management" firm may have experts from all related fields as employees, or to have an associated company that provides each necessary skill. Thus, each such firm may offer itself as "one-stop shopping" for a construction project, from beginning to end. This is designated as a "design build" contract where the contractor is given a performance specification and must undertake the project from design to construction, while adhering to the performance specifications.

開発用で単語数が多い質問の例

- 質問単語数 33 [1]
- 質問 What was the name of the event at the The Embarcadero that was held prior to Super Bowl 50 to help show off some of the things that San Francisco has to offer?
- 正解は Super Bowl City
- 対象の版と分割は 1.1/dev。

質問の原文抜粋 Super Bowl 50

Wikipedia 記事の段落からの抜粋。

The annual NFL Experience was held at the Moscone Center in San Francisco. In addition, "Super Bowl City" opened on January 30 at Justin Herman Plaza on The Embarcadero, featuring games and activities that will highlight the Bay Area's technology, culinary creations, and cultural diversity. More than 1 million people are expected to attend the festivities in San Francisco during Super Bowl Week. San Francisco mayor Ed Lee said of the highly visible homeless presence in this area "they are going to have to leave".

文字数が最も多い入力段落の例

- 段落文字数 4063 ／段落単語数 629 [1]。
- 質問 The freedom to provide services under TFEU article 56 applies to who?
- 正解は to people who give services "for remuneration"
- 対象の版と分割は 1.1/dev 、 2.0/dev 。

最長段落からの原文抜粋

Wikipedia 記事の段落からの抜粋。

The "freedom to provide services" under TFEU article 56 applies to people who give services "for remuneration", especially commercial or professional activity. For example, in *Van Binsbergen v Bestuur van de Bedrijfvereniging voor de Metaalnijverheid* a Dutch lawyer moved to Belgium while advising a client in a social security case, and was told he could not continue because Dutch law said only people established in the Netherlands could give legal advice.

開発で長い回答の例 French and Indian War

- 回答文字数 160 [1]。
- 質問 What was normal British defense?
- 正解は mustered local militia companies, generally ill trained and available only for short periods, to deal with native threats, but did not have any standing forces.
- 対象の版と分割は 1.1/dev 、 2.0/dev 。

長い回答の原文抜粋 French and Indian War

Wikipedia 記事の段落からの抜粋。

New France was defended by about 3,000 troupes de la marine, companies of colonial regulars (some of whom had significant woodland combat experience). The colonial government recruited militia support when needed. Most British colonies mustered local militia companies, generally ill trained and available only for short periods, to deal with native threats, but did not have any standing forces.

開発用で単語数が多い回答の例

- 回答単語数 29 [1]。
- 質問 What does the internal cavity contain?
- 正解は a mouth that can usually be closed by muscles; a pharynx ("throat"); a wider area in the center that acts as a stomach; and a system of internal canals.
- 対象の版と分割は 1.1/dev 、 2.0/dev 。

回答の原文抜粋 Ctenophora

Wikipedia 記事の段落からの抜粋。

The internal cavity forms: a mouth that can usually be closed by muscles; a pharynx ("throat"); a wider area in the center that acts as a stomach; and a system of internal canals. These branch through the mesoglea to the most active parts of the animal: the mouth and pharynx; the roots of the tentacles, if present; all along the underside of each comb row; and four branches round the sensory complex at the far end from the mouth – two of these four branches terminate in anal pores.

バージョン 2.0 開発用の長い質問

- 質問単語数 31 [1]
- 質問 In which case did the Court of Justice state that refusal to admit a lawyer to the Belgian bar because he did not have Belgian heritage wasn't able to be justified?
- 正解は Reyners v Belgium
- 対象の版と分割は 2.0/dev。

質問の原文抜粋 European Union law

Wikipedia 記事の段落からの抜粋。

The Court of Justice has held that both a member state government and a private party can hinder freedom of establishment, so article 49 has both "vertical" and "horizontal" direct effect. In *Reyners v Belgium* the Court of Justice held that a refusal to admit a lawyer to the Belgian bar because he lacked Belgian nationality was unjustified. TFEU article 49 says states are exempt from infringing others' freedom of establishment when they exercise "official authority", but this did an advocate's work (as opposed to a court's) was not official.

バージョン 2.0 のベンチマーク結果

モデル [1]	提出日	EM	F1
BERT	2018-11-08	80.0	83.1
RoBERTa	2019-07-20	86.8	89.8
XLNet	2019-11-15	87.9	90.7

- SQuAD 運営者の非公開テストで採点。
- 各行は単一モデルの提出結果。

[1] 掲載された提出結果（EM・F1 は %）。
<https://rajpurkar.github.io/SQuAD-explorer/>

開発用 Wikipedia 記事一覧 1/2

バージョン 1.1 では、48 記事／ 2.0 このうち 35 記事。

1973_oil_crisis

Amazon_rainforest

American_Broadcasting_Company *

Apollo_program *

Black_Death

Chloroplast *

Civil_disobedience

Computational_complexity_theory

Construction

Ctenophora

Doctor_Who *

Economic_inequality

European_Union_law

Force

French_and_Indian_War

Fresno,_California

Genghis_Khan *

Geology

Harvard_University

Huguenot

Immune_system

Imperialism

Intergovernmental_Panel_on_Climate_Change

Islamism

* はバージョン 1.1 のみに含まれる開発用記事。

<https://github.com/rajpurkar/SQuAD-explorer/tree/master/dataset>

開発用 Wikipedia 記事一覧 2/2

バージョン 1.1 では、48 記事／ 2.0 このうち 35 記事。

Jacksonville,_Florida

Kenya *

Martin_Luther *

Newcastle_upon_Tyne *

Nikola_Tesla *

Normans

Oxygen

Packet_switching

Pharmacy

Prime_number

Private_school

Rhine

Scottish_Parliament

Sky_(United_Kingdom)

Southern_California

Steam_engine

Super_Bowl_50 *

Teacher *

United_Methodist_Church *

University_of_Chicago

Victoria_(Australia)

Victoria_and_Albert_Museum *

Warsaw

Yuan_dynasty

* はバージョン 1.1 のみに含まれる開発用記事。

<https://github.com/rajpurkar/SQuAD-explorer/tree/master/dataset>

学習用 Wikipedia 記事一覧 1/19

バージョン 1.1 と 2.0 に共通する 442 記事。

2008_Sichuan_earthquake

2008_Summer_Olympics_torch_relay

51st_state

A_cappella

Adolescence

Adult_contemporary_music

Affirmative_action_in_the_United_States

Age_of_Enlightenment

Aircraft_carrier

Airport

Alaska

Alexander_Graham_Bell

Alfred_North_Whitehead

Alloy

Alps

Alsace

American_Idol

Animal

Ann_Arbor,_Michigan

Annelid

Antarctica

Antenna_(radio)

Anthropology

Anti-aircraft_warfare

学習用 Wikipedia 記事一覧 2/19

バージョン 1.1 と 2.0 に共通する 442 記事。

Antibiotics

Apollo

Appalachian_Mountains

Architecture

Arena_Football_League

Armenia

Armenians

Arnold_Schwarzenegger

Arsenal_F.C.

ASCII

Ashkenazi_Jews

Asphalt

Aspirated_consonant

Association_football

Asthma

Athanasius_of_Alexandria

Atlantic_City,_New_Jersey

Avicenna

Bacteria

Baptists

BBC_Television

Beer

BeiDou_Navigation_Satellite_System

Bermuda

学習用 Wikipedia 記事一覧 3/19

バージョン 1.1 と 2.0 に共通する 442 記事。

Bern

Beyoncé

Bill_%26_Melinda_Gates_Foundation

Biodiversity

Bird

Bird_migration

Black_people

Boston

Botany

Brain

Bras%C3%ADlia

Brigham_Young_University

British_Empire

British_Isles

Buckingham_Palace

Buddhism

Canadian_Armed_Forces

Canadian_football

Canon_law

Capacitor

Capital_punishment_in_the_United_States

Cardinal_(Catholicism)

Carnival

Catalan_language

学習用 Wikipedia 記事一覧 4/19

バージョン 1.1 と 2.0 に共通する 442 記事。

CBC_Television

Central_African_Republic

Central_Intelligence_Agency

Charleston,_South_Carolina

Chicago_Cubs

Chihuahua_(state)

Child_labour

Chinese_characters

Christian

Circadian_rhythm

Classical_music

Clothing

Comcast

Comics

Communication

Communications_in_Somalia

Compact_disc

Comprehensive_school

Computer

Computer_security

Copper

Copyright_infringement

Cork_(city)

Cotton

学習用 Wikipedia 記事一覧 5/19

バージョン 1.1 と 2.0 に共通する 442 記事。

Crimean_War

Crucifixion_of_Jesus

Cubism

Culture

Cyprus

Czech_language

Data_compression

Database

Daylight_saving_time

Dell

Department_store

Detroit

Dialect

Diarrhea

Digestion

Digimon

Dissolution_of_the_Soviet_Union

Dog

Dominican_Order

Dutch_language

Dutch_Republic

Dwight_D._Eisenhower

East_India_Company

East_Prussia

学習用 Wikipedia 記事一覧 6/19

バージョン 1.1 と 2.0 に共通する 442 記事。

Economy_of_Greece

Edmund_Burke

Education

Egypt

Electric_motor

Elevator

Elizabeth_II

Emotion

Empiricism

Endangered_Species_Act

Energy

England_national_football_team

Eritrea

Estonia

Estonian_language

Eton_College

European_Central_Bank

Everton_F.C.

Exhibition_game

FA_Cup

FC_Barcelona

Federal_Aviation_Administration

Federal_Bureau_of_Investigation

Federalism

学習用 Wikipedia 記事一覧 7/19

バージョン 1.1 と 2.0 に共通する 442 記事。

Film_speed

Financial_crisis_of_2007%E2%80%9308

Florida

Flowering_plant

Franco-Prussian_War

Freemasonry

Friedrich_Hayek

Frédéric_Chopin

Galicia_(Spain)

Gamal_Abel_Nasser

Gene

General_Electric

Genocide

Genome

Geography_of_the_United_States

Geological_history_of_Earth

George_VI

Georgian_architecture

Germans

Glacier

Glass

God

Gothic_architecture

Gramophone_record

学習用 Wikipedia 記事一覧 8/19

バージョン 1.1 と 2.0 に共通する 442 記事。

Grape

Great_Plains

Great_power

Greece

Greeks

Green

Gregorian_calendar

Group_(mathematics)

Guam

Guinea-Bissau

Gymnastics

Han_dynasty

Hanover

Hard_rock

Heian_period

Hellenistic_period

Heresy

High-definition_television

Himachal_Pradesh

Hindu_philosophy

History_of_India

History_of_science

Hokkien

House_music

学習用 Wikipedia 記事一覧 9/19

バージョン 1.1 と 2.0 に共通する 442 記事。

Houston

Human_Development_Index

Humanism

Hunter-gatherer

Hunting

Hyderabad

Hydrogen

IBM

Idealism

Identity_(social_science)

Imamah_(Shia_doctrine)

Immaculate_Conception

Immunology

Imperial_College_London

Incandescent_light_bulb

Indigenous_peoples_of_the_Americas

Infection

Infrared

Insect

Institute_of_technology

Intellectual_property

Internet_service_provider

iPod

Iran

学習用 Wikipedia 記事一覧 10/19

バージョン 1.1 と 2.0 に共通する 442 記事。

Iranian_languages

Israel

Jehovah%27s_Witnesses

Jews

John,_King_of_England

John_Kerry

John_von_Neumann

Josip_Broz_Tito

Kanye_West

Karl_Popper

Kathmandu

Kievan_Rus%27

Korean_War

Labour_Party_(UK)

Lancashire

LaserDisc

Late_Middle_Ages

Law_of_the_United_States

Letter_case

Liberal_Party_of_Australia

Liberia

Light-emitting_diode

Lighting

List_of_numbered_streets_in_Manhattan

学習用 Wikipedia 記事一覧 11/19

バージョン 1.1 と 2.0 に共通する 442 記事。

Literature

London

Macintosh

Madonna_(entertainer)

Madrasa

Mali

Mammal

Mandolin

Marshall_Islands

Marvel_Comics

Mary_(mother_of_Jesus)

Materialism

Matter

Melbourne

Memory

Mesozoic

Mexico_City

Miami

Middle_Ages

Military_history_of_the_United_States

Ministry_of_Defence_(United_Kingdom)

Modern_history

Molotov%E2%80%93Ribbentrop_Pact

Montana

学習用 Wikipedia 記事一覧 12/19

バージョン 1.1 と 2.0 に共通する 442 記事。

Montevideo

Mosaic

MP3

Muammar_Gaddafi

Multiracial_American

Muslim_world

Myanmar

Myocardial_infarction

Namibia

Nanjing

Napoleon

National_Archives_and_Records_Administration

Near_East

Neoclassical_architecture

Neolithic

Neptune

New_Delhi

New_Haven,_Connecticut

New_York_City

Nigeria

Nintendo_Entertainment_System

Nonprofit_organization

Norfolk_Island

North_Carolina

学習用 Wikipedia 記事一覧 13/19

バージョン 1.1 と 2.0 に共通する 442 記事。

Northwestern_University

Nutrition

Oklahoma

Oklahoma_City

Old_English

On_the_Origin_of_Species

Order_of_the_British_Empire

Orthodox_Judaism

Ottoman_Empire

Pacific_War

Pain

Palermo

Paper

Paris

Party_leaders_of_the_United_States_House_of_Representatives

Pesticide

Pharmaceutical_industry

Philadelphia

Philosophy_of_space_and_time

Phonology

Pitch_(music)

Planck_constant

PlayStation_3

学習用 Wikipedia 記事一覧 14/19

バージョン 1.1 と 2.0 に共通する 442 記事。

Plymouth

Police

Political_corruption

Political_party

Political_philosophy

Pope_John_XXIII

Pope_Paul_VI

Portugal

Post-punk

Poultry

Predation

Premier_League

Presbyterianism

Prime_minister

Printed_circuit_board

Professional_wrestling

Protestantism

Pub

Punjab,_Pakistan

Qing_dynasty

Queen_(band)

Queen_Victoria

Quran

Race_(human_categorization)

学習用 Wikipedia 記事一覧 15/19

バージョン 1.1 と 2.0 に共通する 442 記事。

Race_and_ethnicity_in_the_United_States_Census

Railway_electrification_system

Rajasthan

Raleigh,_North_Carolina

Red

Religion_in_ancient_Rome

Renewable_energy_commercialization

Republic_of_the_Congo

Richard_Feynman

Richmond,_Virginia

Roman_Republic

Royal_assent

Royal_Dutch_Shell

Royal_Institute_of_British_Architects

Rule_of_law

Russian_language

Russian_Soviet_Federative_Socialist_Republic

Sahara

Saint_Barth%C3%A9lemy

Saint_Helena

Samoa

Samurai

San_Diego

Sanskrit

学習用 Wikipedia 記事一覧 16/19

バージョン 1.1 と 2.0 に共通する 442 記事。

Santa_Monica,_California

Seattle

Separation_of_church_and_state_in_the_United_States

Separation_of_powers_under_the_United_States_Constitution

Serbo-Croatian

Seven_Years%27_War

Sexual_orientation

Sichuan

Sino-Tibetan_relations_during_the_Ming_dynasty

Slavs

Software_testing

Solar_energy

Somalis

Somerset

Sony_Music_Entertainment

Southampton

Southeast_Asia

Southern_Europe

Space_Race

Spanish_language_in_the_United_States

Spectre_(2015_film)

St._John%27s,_Newfoundland_and_Labrador

学習用 Wikipedia 記事一覧 17/19

バージョン 1.1 と 2.0 に共通する 442 記事。

States_of_Germany

Steven_Spielberg

Strasbourg

Sumer

Super_Nintendo_Entertainment_System

Supreme_court

Swaziland

Switzerland

Symbiosis

Szlachta

Tajikistan

Tennessee

Textual_criticism

The_Blitz

The_Bronx

The_Legend_of_Zelda:_Twilight_Princess

The_Sun_(United_Kingdom)

The_Times

Thuringia

Tibet

Time

To_Kill_a_Mockingbird

Transistor

Translation

学習用 Wikipedia 記事一覧 18/19

バージョン 1.1 と 2.0 に共通する 442 記事。

Treaty	United_States_presidential_election,_2004
Tristan_da_Cunha	Universal_Studios
Tuberculosis	University
Tucson,_Arizona	University_of_Kansas
Turner_Classic_Movies	University_of_Notre_Dame
Tuvalu	Uranium
Umayyad_Caliphate	USB
Unicode	Utrecht
United_Nations_Population_Fund	Vacuum
United_States_Air_Force	Valencia
United_States_Army	Videoconferencing
United_States_dollar	Virgil

学習用 Wikipedia 記事一覧 19/19

バージョン 1.1 と 2.0 に共通する 442 記事。

War_on_Terror

Zinc

Warsaw_Pact

Washington_University_in_St._Louis

Wayback_Machine

Web_browser

Westminster_Abbey

Windows_8

Wood

Xbox_360

Yale_University

YouTube

Zhejiang

各版の公式 JSON を取得する Python コード

```
import json
from urllib.request import urlopen

base = "https://rajpurkar.github.io/SQuAD-explorer/dataset/"
for version in ("1.1", "2.0"):
    for split in ("train", "dev"):
        name = f"{split}-v{version}.json"
        with urlopen(base + name) as response:
            data = json.load(response)
        with open(name, "w", encoding="utf-8") as output:
            json.dump(data, output, ensure_ascii=False)
```

SQuAD: 100,000+ Questions for Machine Comprehension of Text

Rajpurkar, P., Zhang, J., Lopyrev, K., & Liang, P. (2016). EMNLP, pp. 2383–2392.

SQuAD の原著論文。データセットの収集方法、回答範囲の注釈、EM ・ F1 評価を記載する。

<https://aclanthology.org/D16-1264/>

Know What You Don't Know: Unanswerable Questions for SQuAD

Rajpurkar, P., Jia, R., & Liang, P. (2018). ACL, pp. 784–789.

SQuAD 2.0 の原著論文。回答不能な質問の作成方法と、回答保留を含む評価を報告する。

<https://aclanthology.org/P18-2124/>

Question answering

Hugging Face (n.d.). Hugging Face Course, Chapter 7, Section 7. 参照 2026-10-15.

質問応答の公式教材。長い段落の分割、回答位置の変換、モデルの微調整と評価を扱う。

<https://github.com/huggingface/course/blob/main/chapters/en/chapter7/7.mdx>

BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). NAACL-HLT, pp. 4171–4186.

BERT の原著論文。双方向の事前学習と微調整を提案し、SQuADなどで評価する。

<https://aclanthology.org/N19-1423/>

Adversarial Examples for Evaluating Reading Comprehension Systems

Jia, R., & Liang, P. (2017). EMNLP, pp. 2021–2031.

読解モデルの頑健性を調べた研究。SQuAD の段落に紛らわしい文を追加して性能を検証する。

<https://aclanthology.org/D17-1215/>

SQuAD-explorer

Rajpurkar, P., et al. (n.d.). GitHub repository. 参照 2026-10-15.

SQuAD の公式配布リポジトリ。公開JSON、EM・F1の評価スクリプトと提出形式を提供する。

<https://github.com/rajpurkar/SQuAD-explorer>