

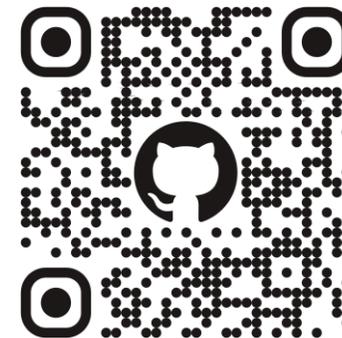
JLR2025

日本語マルチホップQAデータセット
JEMHopQA作成時のエラー分析
および
修正版公開のお知らせ

2025年3月14日（金）

石井 愛 (理研/BIPROGY), 井之上 直也 (JAIST/理研),
鈴木 久美, 関根 聡 (理研)

- Wikipediaをベースとした日本語のマルチホップ QA データセット[1]
- 回答の導出情報がトリプルの形式で付与されている
- Train 1,059問, Dev 120問
- JEMHopQAの質問のタイプとフォーマット:



[JEMHopQA](#)

- **構成問題:**

質問:

ルーヴル美術館が所在する都市の市長の名前は?

導出: (ルーヴル美術館, 所在地, パリ), (パリ, 市長, アンヌ・イダルゴ)

回答: アンヌ・イダルゴ

導出トリプル

- **比較問題:**

質問: 『天空の城ラピュタ』と『となりのトトロ』の公開日が早いのは、『となりのトトロ』ですか?

導出: (天空の城ラピュタ, 公開年, 1986年)

(となりのトトロ, 公開年, 1988年)

回答: NO

JEMHopQAはどのように作られたか

[2] [根拠を説明可能な質問応答システムのための日本語マルチホップQAデータセット構築](#) (石井 et al., NLP2023)

- 3つのタスクに分け、クラウドソーシング※と一部GPT(3.5)にて作成

※ランサーズでタスク形式

● 構成問題



トリプル作成タスク

(ルーヴル美術館, 所在地, パリ)

(パリ, 市長, アンヌ・イダルゴ)

構成問題作成タスク

質問: ルーヴル美術館が所在する都市の市長の名前は?

回答: アンヌ・イダルゴ

根拠: 入力のものから選択

● 比較問題



比較問題作成タスク

回答がYESとなる問題を作成してください。

質問: 『千と千尋の神隠し』と『となりのトトロ』の監督は同じ人物ですか?

根拠: (千と千尋の神隠し, 監督, 宮崎駿), (となりのトトロ, 監督, 宮崎駿)

- 自動的なルールベースでのチェック・修正
 - ・ エンティティの欠落が無いか等の前提チェックをルールベースで実施
- 手動での修正
 - ・ 形式チェックや、回答が一意に定まるかをWikipediaから判別できない可能性も踏まえてチェック

対応	エラー内容	質問 (⇒修正した質問)	導出	回答
修正	ブリッジエンティティが利用されない	檀れいの卒業校はどこですか？ ⇒ 檀れいの元夫が卒業したのは何大学何学部ですか？	(檀れい, 元配偶者, 及川光博), (及川光博, 卒業大学・学部, 成城大学法学部)	成城大学 法学部
削除	ブリッジエンティティが複数	関口宏の高校の同級生である俳優が亡くなった年月日は？ カップヌードル販売会社	(関口宏, 高校の同級生, 林隆三), (林隆三, 没年月日, 2014年6月4日)	2014年 6月4日
修正	回答の異表記が存在する	の東京本社所在地は？ ⇒ …本社所在地の区は？	(カップヌードル, 販売会社, 日清食品), (日清食品, 東京本社所在地, 東京都新宿区新宿)	新宿区
修正	比較問題形式でない	楽曲『おいでシャンプー』と『君の名は希望』の作詞者は？ ⇒ …はどちらも秋元康ですか？	(おいでシャンプー, 作詞者, 秋元康), (君の名は希望, 作詞者, 秋元康)	YES
削除	GPTの捏造	『ダイヤモンドは砕けない』の作者が直木賞を受賞したのは何年ですか？	(ダイヤモンドは砕けない, 作者, 荒木飛呂彦), (荒木飛呂彦, 受賞年, 2013年)	2013年
フラグ	将来的に正解が変わる	ルーヴル美術館が所在する都市の市長の名前は？	(ルーヴル美術館, 所在地, パリ), (パリ, 市長, アンヌ・イダルゴ)	アンヌ・ イダルゴ

最終セット:
1179件
手動修正実施:
576件 (49%)

今回再度エラーチェックを実施

- ① 検証でいつも解けない問題から発覚 (dev 120件)
 - 7件修正
- ② ChatGPT o3-miniにてシンプルなエラーチェックプロンプト (train1,059件)
 - 15件修正 (21件検出)
- ③ ChatGPT o1 proにて詳細なエラーチェックプロンプト (dev+train1,179件)
 - 12件修正 (14件検出)
- その他改善のための修正
 - 質問の明確化等 19件
 - time_dependentフラグ修正 50件

エラー内容(1/2)

プロンプトによって発見されるエラーが異なった

● クラウドソーサーに起因するエラー

エラータイプ	例	件数	内訳
コピペミス	<ul style="list-style-type: none">『A』と『プロミス・シンデレラ』の作者はどちらも矢沢あいですか？→AではなくNANAアニメ『ニセコイ』と『鬼滅の刃』でテレビ放送が早いのは、『ニセコイ』ですか？→根拠の値が逆に書かれていて反対の回答	9	① : 4 ② : 5 ③ : 2* *②と重複
問題形式の理解不足	<ul style="list-style-type: none">北広島市と江別市はどちらも北海道にありますか？→回答：北海道 (YES/NOを設定してほしい)菅野智之と原辰徳、先に生まれたのは菅野智之ですか？→YES (NOが正しい)	7	② : 7
読解ミス	<ul style="list-style-type: none">豊岡市と伊丹市の管轄区域は、どちらも但馬県民局ですか？ → 管轄県民局 (主語と述語が反対)(カルロス・ゴーン, 高等学校, エコール・ポリテクニーク) → エコール・ポリテクニークは高等教育機関『電影少女』と『るろうに剣心 -明治剣客浪漫譚-』、WJで連載したのは？→両方WJ=週刊少年ジャンプ	6	① : 1 ③ : 5
質問の表現エラー	「…よりも早生まれですか？」→「…よりも早く生まれましたか？」	3	② : 3

エラー内容(2/2)

詳細版プロンプト
でWikipediaの間違
いも発見

● Wikipediaの内容・生成の仕組みによるエラー

エラータイプ	例	件数	内訳
Wikipediaに起因する誤り	<p>諸説あるけれどWikipediaの記載が一般的な内容とは異なるケース</p> <ul style="list-style-type: none">・ 「萩市と豊岡市、日本海側気候に属するのはどちらですか？」 →萩市は一般に日本海側気候には属し、両方同じになってしまう・ 「中華人民共和国の河北省の最大都市の市樹は？」 →根拠では「最大都市=唐山市」としているが、最大都市は省都の石家荘とされることが多い・ （名古屋第二環状自動車道，制定年，2010年） →公式ドキュメントでは1971年と記載	5	③ : 5
生成フレームワークの問題	<ul style="list-style-type: none">・ 本来は回答が複数値になるのに候補のトリプルが1つしかなかった・ 同一表記の異なるエンティティが存在	3	① : 3

チェックの指摘間違い

- ChatGPTによるチェックは指摘の間違いが少数ながら発生

チェック結果タイプ	例	件数	内訳
同義語を間違いと指摘	プロキオンが形成する星座は何の48星座の一つと言われて いるか？→トレミーの48星座（「 プトレマイオスの48星 座 」が正しいと指摘）	1	②:1
根拠の値が間違いと指摘	<ul style="list-style-type: none">・ 3件は社長名や、分社化する前後等、情報が入り組んでい る値への指摘・ 2件はJEMHopQAのWikipediaのバージョンと2021年と 現在の値が異なるためのもの	5	②:5
根拠が不確かと指摘	<ul style="list-style-type: none">・ 「徳川家宣の母が田中保良子」という根拠は誤りと見られ、裏付けが取れないため・ 妻の名「みかみなち」が公式根拠不明でトリプル自体も不確かなため	2	③:2

- 修正版Ver1.2および評価用スクリプトを公開します

根拠を説明可能な質問応答システムのための日本語マルチホップQAデータセット

JEMHopQA: <https://github.com/aiishii/JEMHopQA>



研究・開発等にご活用いただければ幸いです。フィードバックもお待ちしております。

ご参考) ChatGPTを用いたエラーチェックのプロンプト

② シンプル版

以下のマルチホップQAデータセットの根拠であるderivationsのエンティティが実在し、正しい情報が記載されているか、その根拠を用いたquestionとanswerが適切かを詳細にチェックしてください。
また、この問題は2021年のWikipediaを元に作成されていますが、社長や市長等、時間がたつと回答が変化する問題については、time_dependentをtrueとしています。その設定が適切かどうかあわせてチェックしてください。
修正箇所のレコードのみを同じフォーマットのJSONで出力してください。

③ 詳細版

以下の根拠情報を含むマルチホップQAデータセットの詳細なチェックをお願いします。

【チェック項目】

- ・質問文として成り立っているか(...である。等質問文ではなく通常の文の場合はNG)
- ・根拠であるderivationsのエンティティが実在し、正しい情報が記載されているか
- ・その根拠を用いたquestionとanswerが適切か
- ・質問文の意図が明確か
- ・質問に対する回答が一意になるか（様々な回答があり得る場合NG)
- ・根拠は（主語、関係、目的語）のトリプルの形式になっているか

【Compositionタイプのチェック項目】

あるエンティティの属性情報であるブリッジエンティティ介した2つめのエンティティの属性値を問う質問形式です。

例) 「フランスパリのセーヌ河岸にある美術館の最寄り駅は？」のように2つのトリプル(パリ,美術館,ルーヴル美術館) (ルーヴル美術館,最寄り駅,パレ・ロワイヤル=ミュゼ・デュ・ルーヴル駅)で構成され、ブリッジエンティティと呼ばれる「ルーブル美術館」が質問では隠されています。

上記の例ではルーブル美術館以外にも美術館が存在し、特定できない可能性があるため、NGとなります。

- ・ブリッジエンティティと回答が一意に定まるか
- ・ブリッジエンティティが質問文から隠されているか
- ・ブリッジエンティティでリンクされた2つ以上のトリプルが根拠として設定されているか
- ・ブリッジエンティティでリンクされた2つ以上のトリプルが、質問に回答するために必要か

【Comparisonタイプのチェック項目】

2つのエンティティの属性の比較や、両方が指定された属性値を持つかどうか、指定された属性値を持つのはどちらかを問う質問形式です。

例) 「神田沙也加と加護亜依で年齢が上なのは、神田沙也加ですか？」といった2つのエンティティの属性を比較する質問、
「川崎重工業と本田技研工業はどちらも自動車を製造していますか？」といった2つのエンティティの属性が同じかどうかを判定する質問、
「武漢市と咸陽市では中国中部に位置するのはどちらでしょうか？」といったその属性値を持つのが2つのエンティティのどちらかを回答する問題です。
「東北新幹線と北海道新幹線は高速鉄道路線ですか？」という質問は不完全なのでNGで、「東北新幹線と北海道新幹線はどちらも／両方高速鉄道路線ですか？」という質問にするのが正しいです。
「『見上げてごらん夜の星を』と『明日があるさ』を歌っている人は？」という質問は、どちらかのエンティティの情報のみで回答できてしまうためNGです。

- ・comparisonの形式は正しいか
- ・比較しているエンティティ同士は妥当か
- ・数値の比較結果は正しいか
- ・どちらか一方を選ぶ問題で適切に回答を選んでいるか

上記の結果でNGのもののみ、入力のTSVにcheck_result列、check_reason列を追加し、それぞれOK/NG、その理由を出力してください。