

# J-KGMHQA: Agentic RAGに向けた知識グラフに基づく日本語マルチホップQAデータセットの自動合成

板井 孝樹<sup>1,2</sup> 松田 耕史<sup>2</sup> 福地 成彦<sup>2</sup> (1. 東京都立大学 2. SB Intuitions)

## 概要

### KG x LLMによる日本語マルチホップQAの自動合成 : J-KGMHQA

(Japanese KG-grounded Multi Hop QA)



データダウンロードはこちら  
(Hugging Face)

## 1. 背景

- Agentic RAGの学習・評価: 高品質な行動軌跡 (trajectory) が不可欠 ⇒ 多段階推論を要するマルチホップQAが重要
- 既存日本語マルチホップQA (MHQA): データ規模に課題 (問題数・ホップ数) ⇒ **課題①. スケーラビリティ**
- 既存MHQA合成手法: 中間ホップの知識の事実性評価がLLMに依存 ⇒ **課題②. 回答根拠の確実性**

- 課題①. スケーラビリティ** : Knowledge Graph (KG) から推論パスの抽出 x LLM による質問合成
- 課題②. 回答根拠の確実性** : KG の知識エンティティに基づく根拠文の合成・挿入

## 提案手法

### Knowledge Base (Wikidata [1])に基づく合成プロセス

- KG構築**: 日本語エンティティ・1対1のリレーション
- 推論パス抽出**: 特異性指標導入し, ハブエンティティへの偏りを防ぎ, 多様かつ任意のホップ数の推論パスを抽出

$$S(h, r, t) = \alpha I_R(r) + \beta I_E(t)$$

- リレーション  $r$  の特異性  $I_R(r) = \log(|T| / (\text{count}(r) + 1))$
- エンティティ  $t$  の特異性  $I_E(t) = \log(|E| / (\text{in-degree}(t) + 1))$

※  $\text{count}(r)$ :  $r$  の出現頻度

※  $|T|$ : 全トリプル数,  $|E|$ : 全エンティティ数

スコア  $S(h, r, t)$  の上位  $K$  個から確率的に次ホップを決定

- 根拠文の合成**: 知識トリプルをLLMで文章に変換  
→ Wikipedia記事に回答根拠文章として挿入
- 質問合成**: 推論パスから質問文を合成
- 品質フィルタ**: LLM-as-a-Judgeによる「推論パスの論理的整合性」と「文書からの回答可能性」の二段階検査

※ 合成器: Qwen3-32B, 評価器: gpt-oss-120bを使用

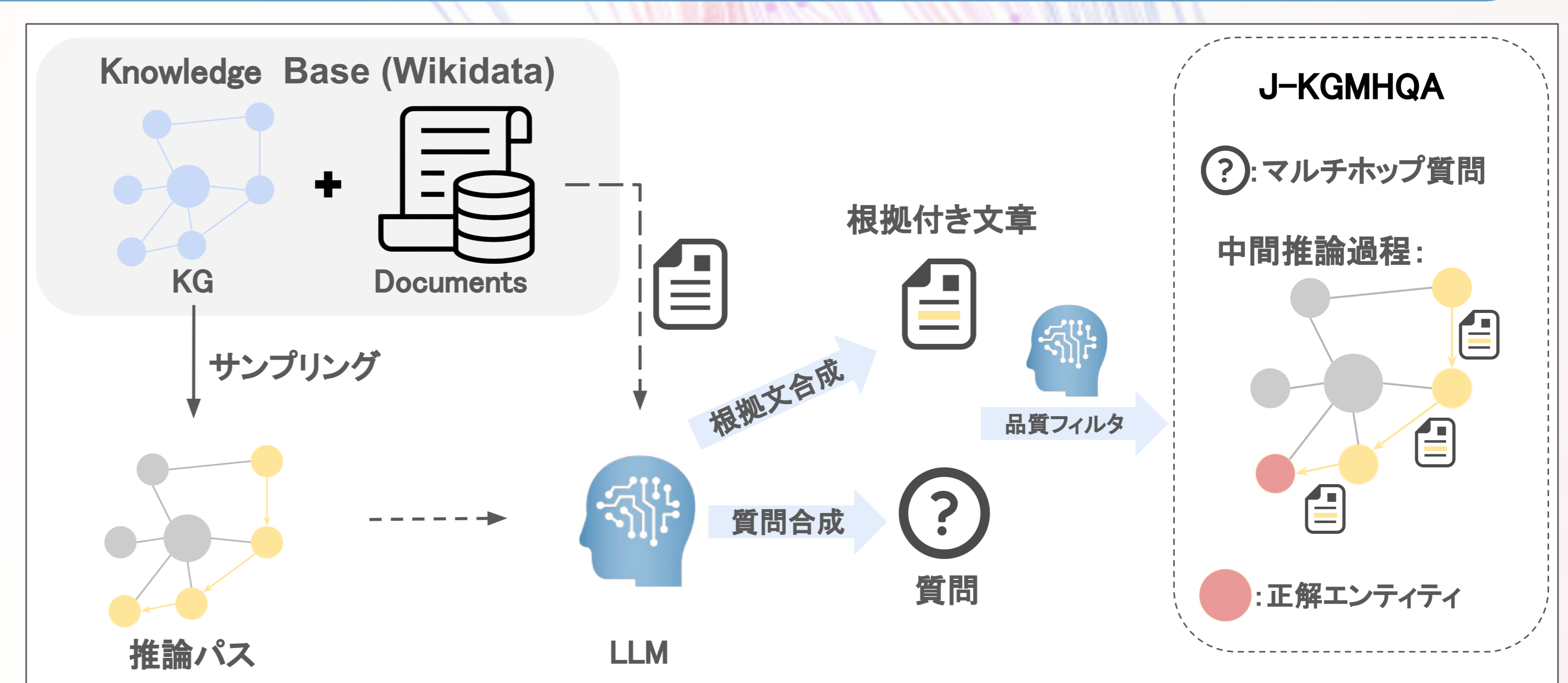


図1. 提案手法のデータ合成の流れ

表1. J-KGMHQAのカテゴリ・ホップ数別の統計量

	2-hop	3-hop	4-hop	5-hop	合計
人物	2,346	1,114	572	279	4,311
場所	1,497	532	198	69	2,296
施設	4,268	2,426	1,553	938	9,185
組織	1,922	910	417	155	3,404
作品	3,922	1,843	875	367	7,007
合計	13,995	6,825	3,615	1,808	26,203

## 評価実験

### 実験設定

- 推論手法**:
  - Closed-book : LLMの内部知識に基づく回答
  - Vanilla RAG : 一度だけ検索を行い回答
  - Agentic RAG : ReAct<sup>[2]</sup>に基づく多段階検索・推論
- 評価モデル**: Qwen3-32B / GPT-5.2
- 検索器**: BM25 (kuromojiを用いて単語分割)
- 評価データ**: 人手品質検査通過済みのサブセット(995件)

### 結果

- Agentic RAG >> Vanilla RAG > Closed-book
- ホップ数の増加に伴うAccuracyの低下

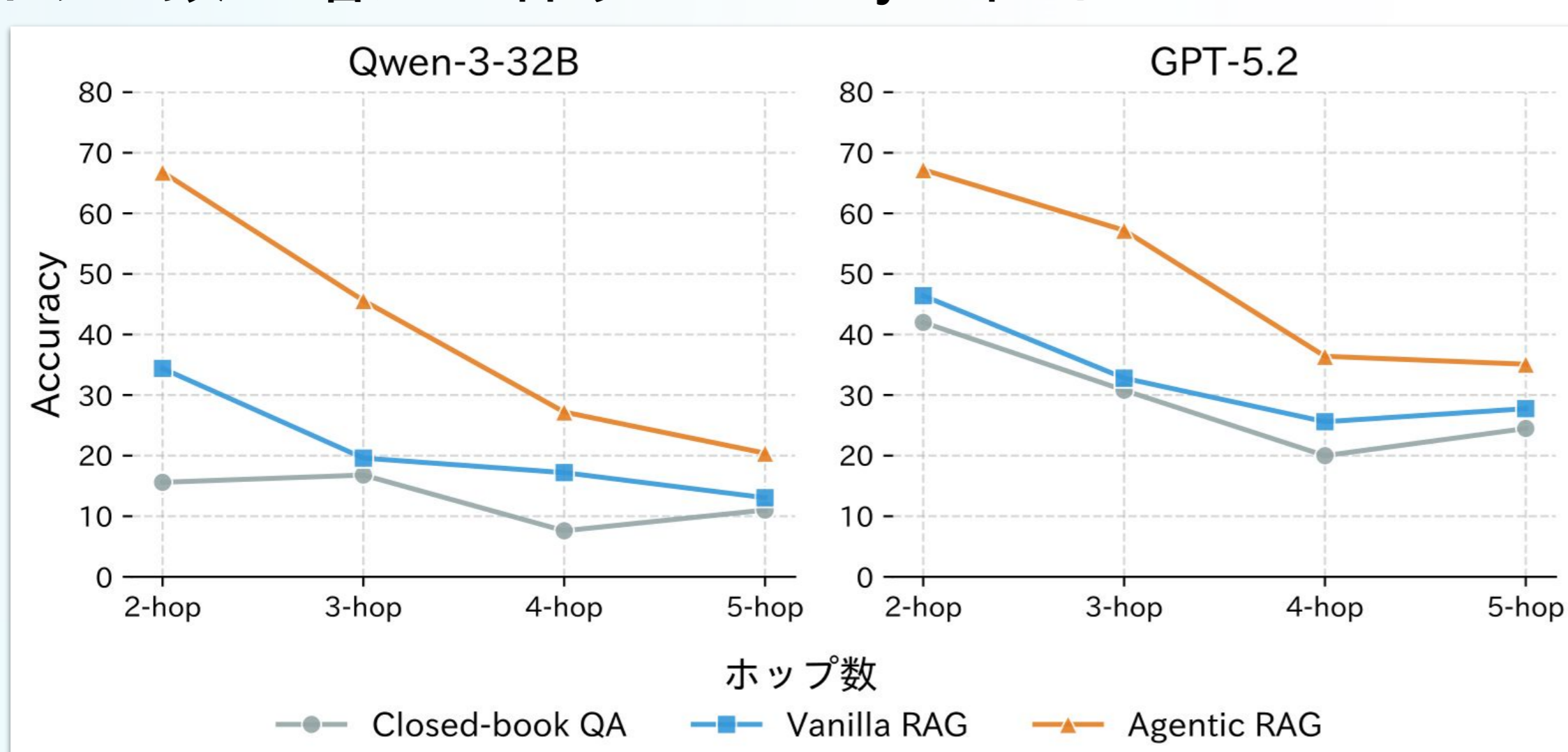


図2. 各推論手法におけるQAの正解率. ホップ数の増加にしたがって, 減少する傾向を確認.

## 分析

- KGと実世界知識の競合**: モデルが導き出した実世界の正解がデータセットの想定解と矛盾する「閉世界仮定」の限界
- 内部知識に基づく誤謬**: 検索で根拠が得られない場合, モデル自身の知識で補完を行い, 論理飛躍・事実誤認

Q.	中村計の出生地のマスコットキャラクターは何ですか?	Q.	「常山紀駿」のまえがきを書いた人物が勤務していた藩の創設者の父は誰ですか?
推論パス	(中村計, 出生地, 船橋市), (船橋市, マスコットキャラクター-ふなっしー)	推論パス	(常山紀駿, 創設者, 松井康重), (松井康重, 父, 松平康政)
A.	ふなっしー	A.	松平康政
Trajectory	<search>船橋市公式マスコットキャラクター</search> <information>ふなっしーは...市や県からは公認されていない...</information> <thinking>検索結果のふなっしーは非公認とあるため不適。船橋市公式のマスコットは「目利き番頭船えもん」である</thinking> 予測回答: 目利き番頭船えもん	Trajectory	<search>常山紀駿創設者</search> <information>(根拠の記述が存在しない)</information> <thinking>常山紀駿の創設者は松井松平康重である。松井松平康重は、江戸時代初期の武将・大名。常陸笠間藩主を経て、1609年より丹波篠山藩初代藩主を務めた。康重は、松井康政の子として生まれた。松井松平康重の父は松井康政である。</thinking> 予測回答: 松井康政

図3. 左: KGの想定解と実世界知識の競合, 右: 検索失敗時に内部知識に基づく誤謬

## 本研究の貢献・展望

### 貢献

『KGに基づく知識の論理的整合性』・『根拠文の確実性』を保証する日本語MHQAデータセットの自動合成手法の確立

### 展望

- 知識網羅性の拡張・他ドメインKGでの適用
- 非構造化データからの動的なKG構築・知識抽出への拡張

### 参考文献

- [1] Denny Vrandečić et al., Wikidata: A free collaborative knowledge base. Communications of the ACM, 2014.  
[2] Shunyu Yao et al., React: Synergizing reasoning and acting in language models. ICLR 2023.

多段階推論・検索が有効

Agentic RAG 学習・評価に有用な難易度