

## 報告書

### 1 今週の進捗

- ChatGPT を用いた災害データセットのナレッジグラフ生成

### 2 ChatGPT を用いた災害データセットのナレッジグラフ生成

ChatGPT の API を用いて災害年表マップの能登半島地震データからナレッジグラフを生成した。モデルは“gpt-3.5-turbo”を使用しており、今後“gpt-4o”のモデルを使用する予定である。

#### 2.1 入力データ（能登半島地震データ）

災害年表マップから取得した能登半島地震データを以下の URL に示す。

[https://github.com/1g-hub/Horimoto/blob/main/Academic/JSIK2024/dataset/Noto\\_quack\\_1.pdf](https://github.com/1g-hub/Horimoto/blob/main/Academic/JSIK2024/dataset/Noto_quack_1.pdf)

この PDF ファイルをテキストデータに変換した結果を Listing 1 に示す。

Listing 1: pdf2text

```
1 災害事例レポート（概要）登録日：年月日（
2 20210201Ver）.4.0出力日：年月日
3 2022091317:13
4 2020-03-13_N_Exxxxxx_JP16201-062194-20富山県富山市富山県富山市（地方公共団体コード：
5
6 JP16201 ）代表点座標：
7 36.695982, 137.213449出典資料：石川県能登地方を震源とする地震地震災害（地震）震源：石川県能
  登半島（
8
9
10 N ° 37.3, E ° 136.8 ), 深さ：12 kmマグニチュード：, 計測値種類：不明
11 5.5発生日年月日（令和
12 2020313 年月日）2313発生時刻（
13 2:1824）H被害詳細人的被害
14
15
16
17 1 名内閣府（）：石川県能登地方を震源とする地震に係る被害状況等について,
18 2020 p1. /7：（平成）年月日時点発
19 *120132511生時期地震
20
21
22
23 E災害の種類負傷者軽傷自然現象災害発生地現在の自治体名
24
25
26 *1出典資料範囲精度
27 *2災害名称災害事例
```

28 IDレポート作成：国立研究開発法人防災科学技術研究所災害事例レポート（災害現象の詳細）登録日：  
年月日（  
29 20210201Ver）.4.0出力日：年月日  
30 2022091317:13  
31 2020-03-13\_N\_Exxxxx\_JP16201-062194-20石川県能登半島（  
32 N ° 37.3, E ° 136.8 ），深さ：12 km，計測値種類：不明  
33 5.5津波津波なし地震災害事例  
34 ID地震災害震源マグニチュード津波発生レポート作成：国立研究開発法人防災科学技術研究所災害事  
例レポート（概要）：範囲精度の解説  
35  
36  
37  
38  
39  
40  
41  
42 \*2範囲精度事例の地理的な範囲の用例事例範囲のイメージ事例レコードの地理的範囲と災害統計値の  
集計エリアとが  
43  
44 A同一。事例レコードの地理的範囲（現在の市）は、災害発生当時  
45  
46 BAの旧市町村エリアよりも大きいため、実際には災害が発生していない地域も含む。出典資料に事例  
レコードの地理的範囲に関する記述が無  
47  
48  
49 Cく、正確な範囲が不明。（例：非常に古い災害、河川の流域）事例レコードおよびその周辺地域（隣接  
自治体までの範囲）  
50  
51 Dを含む範囲で、かつ、事例レコードの範囲に関する情報を抽出できない。この場合、事例レコード内  
での被害状況は事実よりも過剰な記述となる。事例レコードを含む広域（郡、県、山麓レベル）の  
範囲で、か  
52  
53  
54  
55 Eつ、事例レコードの範囲に関する情報を抽出できない。県・地方単位レポート作成：国立研究開発法  
人防災科学技術研究所

能登半島地震の情報として不要な情報も含まれているため、その情報の扱い方について検討したい。手動で  
不必要なページを取り除く方法が考えられる。

## 2.2 プロンプト

以下に GPT に入力したプロンプトを示す。

あなたは災害の報告書における重要な用語とそれらの関係性を表すナレッジグラフの作成者です。  
このナレッジグラフを用いることで災害データの保存と継承を目的とし、特に災害同士の関係性を明確にし  
たいと考えています。  
あなたのタスクは、与えられた入力文に対して災害において重要な用語と関係性を以下の条件に従って抜き  
出してくることです。

入力文：災害の報告書

条件1：災害同士の関係性を明確にするために必要な用語を抜き出してください。具体的には、「災害名称」、「発生日時」、「被害内容」、「被災地」、「被害の程度」などがあります。用語はできる限り詳しく記述してください。

条件2：これらの用語が他の用語とどのような関係性をもつのか推測してください。用語は他の複数の用語と関係していることがあります。

条件3：作成した各3つ組について重要度（1～5）で評価してください。具体的には、他の災害と関係が深いと考えられる3つ組の重要度は5、他の災害と関係が浅いと考えられる3つ組の重要度は1です。

出力形式はJSONのリストにしてください。

リストの各要素は、2つの用語とそれらの関係性、その3つ組の重要度を含み、以下のように記述してください。

```
[
  {
    "用語 A": "抽出された用語 A",
    "用語 B": "抽出された用語 B",
    "関係性": "用語 A と用語 B の関係性",
    "重要度": "（1～5）"
  },
  ...
]
```

3つ組を最低でも10個作成してください。

入力文: {input} 出力:

---

ナレッジグラフの使用目的とともにナレッジグラフの生成ルールを与えている。“input”に入力データを入れる。また、ある3つ組が地震の情報としてどれくらい重要なのかを示す値も出力させている。

現在、プロンプトは1つしかないが、あと2,3個のプロンプトを用意する予定である。

## 2.3 出力結果

出力結果を以下に示す。

```
[
  {
    "用語 A": "地震",
    "用語 B": "石川県能登地方を震源とする地震",
    "関係性": "発生した災害",
    "重要度": 5
  },
  {
    "用語 A": "地震",
    "用語 B": "津波",
```

```

"関係性": "同時に発生した災害",
"重要度": 4
},
{
"用語 A": "地震",
"用語 B": "石川県能登半島",
"関係性": "震源地",
"重要度": 3
},
{
"用語 A": "地震",
"用語 B": "富山県富山市",
"関係性": "被災地",
"重要度": 4
},
{
"用語 A": "地震",
"用語 B": "2020 年 3 月 13 日",
"関係性": "発生日",
"重要度": 4
},
{
"用語 A": "地震",
"用語 B": "2:18",
"関係性": "発生時刻",
"重要度": 3
},
{
"用語 A": "地震",
"用語 B": "マグニチュード",
"関係性": "地震の規模",
"重要度": 4
},
{
"用語 A": "地震",
"用語 B": "深さ",
"関係性": "震源の深さ",
"重要度": 3
},
{
"用語 A": "地震",
"用語 B": "被害内容",
"関係性": "地震による被害",
"重要度": 4
}

```

```

[
  "用語 A": "地震",
  "用語 B": "人的被害",
  "関係性": "地震による人的被害",
  "重要度": 4
]

```

プロンプトで「最低 10 個」と指示したため、出力された 3 つ組の個数が少なく、より多く出力するほうが望ましい。プロンプトに修正を加える必要がある。

出力形式としては、「地震-同時に発生した災害-津波」や「地震-震源地-石川県能登半島」は想定通りの出力であるが、「地震-震源の深さ-深さ」や「地震-地震による人的被害-人的被害」のような単に用語を説明しただけの 3 つ組に関しては好ましくない。また、ヘッド (用語 A) がすべて「地震」となっているが、他のヘッドも出力されてほしい。これらもプロンプトの修正によって改善したい。

重要度に関しては、おおよそ間違いではない値が出力されているが、出力数が少ないことにより 1, 2 の出力がない。これもプロンプトを修正して出力数を増やすことで改善されと考えられる。

### 3 情報知識学会 (JSIK)

#### 3.1 概要

「研究データエコシステム × 地域資料の保存・継承」～災害を乗り越え地域資料継承に貢献する研究データエコシステムの未来～

- 10 月 18 日（金）発表原稿提出期限 12:00
- 11 月 30 日（土）情報知識学フォーラムの開催

### 4 今後

- プロンプトの修正および作成
- 論文の作成
- ナレッジグラフの評価方法の検討

### 参考文献