



地域資料における災害データの Knowledge Graph 生成手法の提案

創発ソフトウェア研究室

M1 堀本 隆誠

発表の流れ

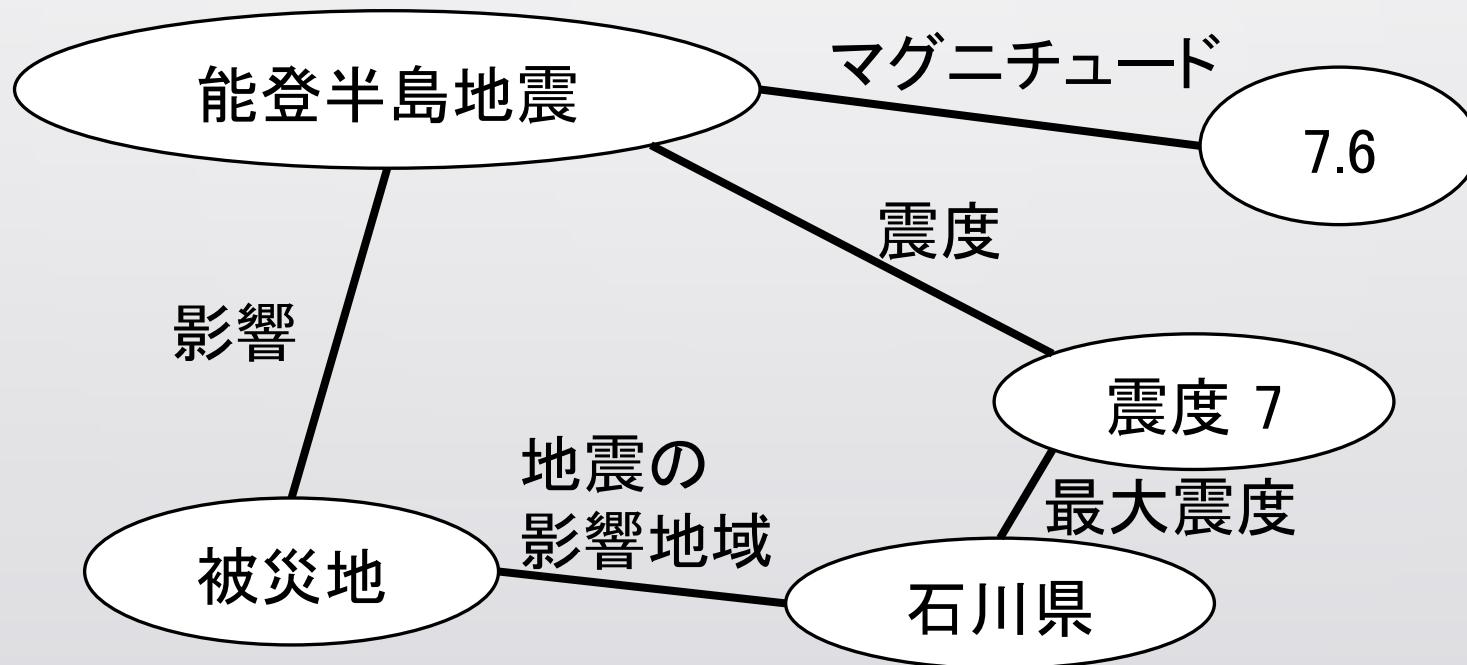
1. はじめに
2. 要素技術
3. 提案手法
4. 実験
5. 実験結果
6. まとめと今後の課題

発表の流れ

1. はじめに
2. 要素技術
3. 提案手法
4. 実験
5. 実験結果
6. まとめと今後の課題

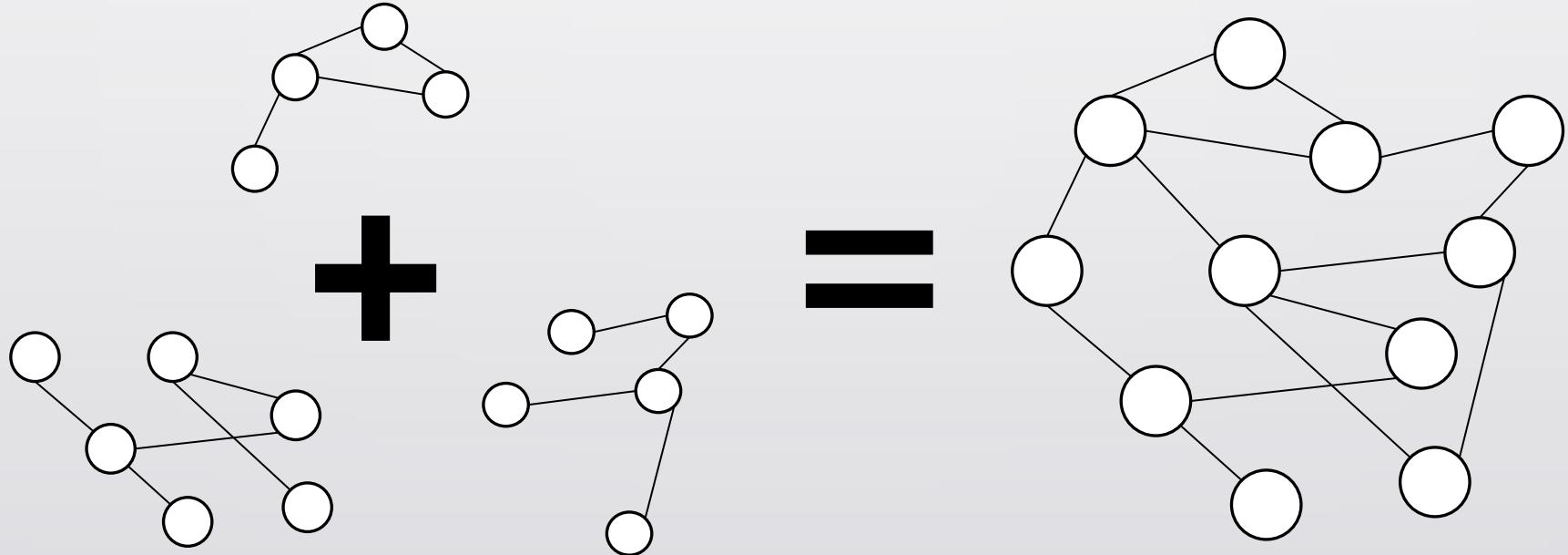
はじめに

- 人工知能の基盤技術 : Knowledge Graph
 - ◆ 人間の知識をグラフ構造で表現
 - ◆ 多種多様な情報とそのつながりを体系的に表現可能



はじめに

- 人工知能の基盤技術 : Knowledge Graph
 - ◆ 複数の Knowledge Graph の情報を統合可能
 - ◆ 時系列データに対して有効である可能性

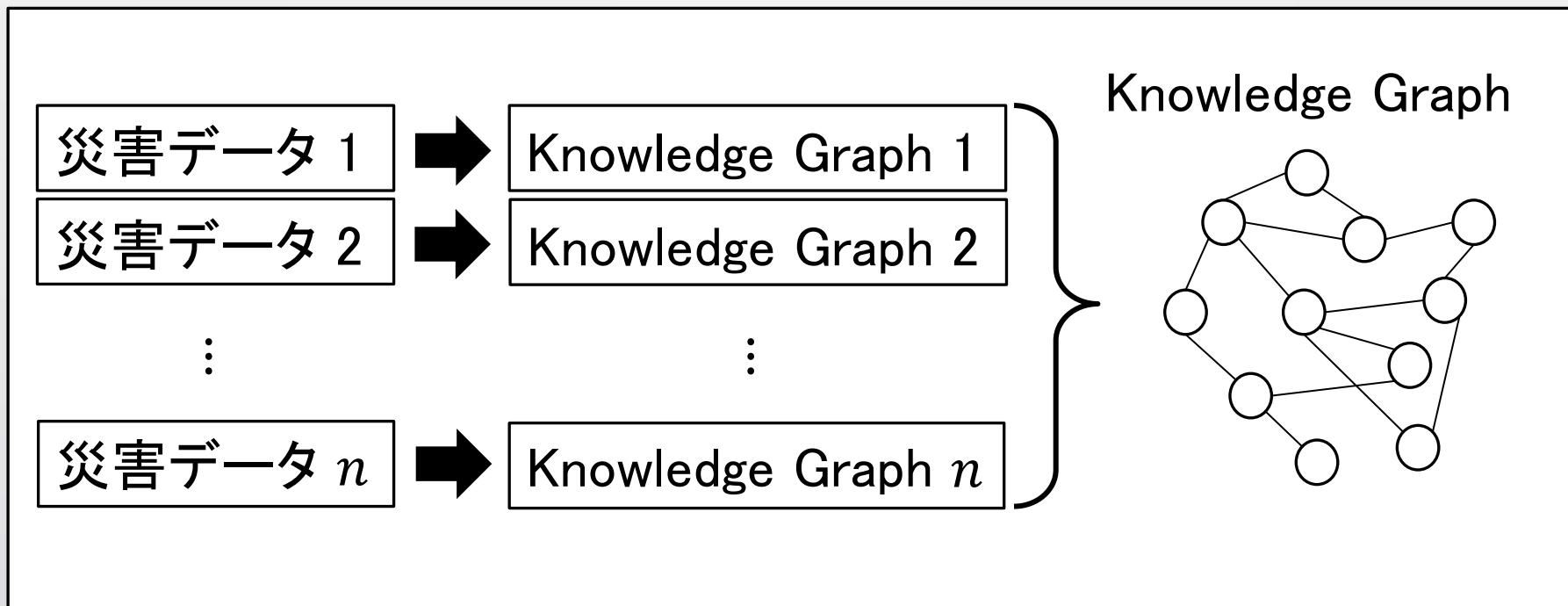


はじめに

- 本研究では

時間経過で変化する災害データを対象に

GPT-4o を用いて Knowledge Graph を抽出 ⇒ 有効性を検証



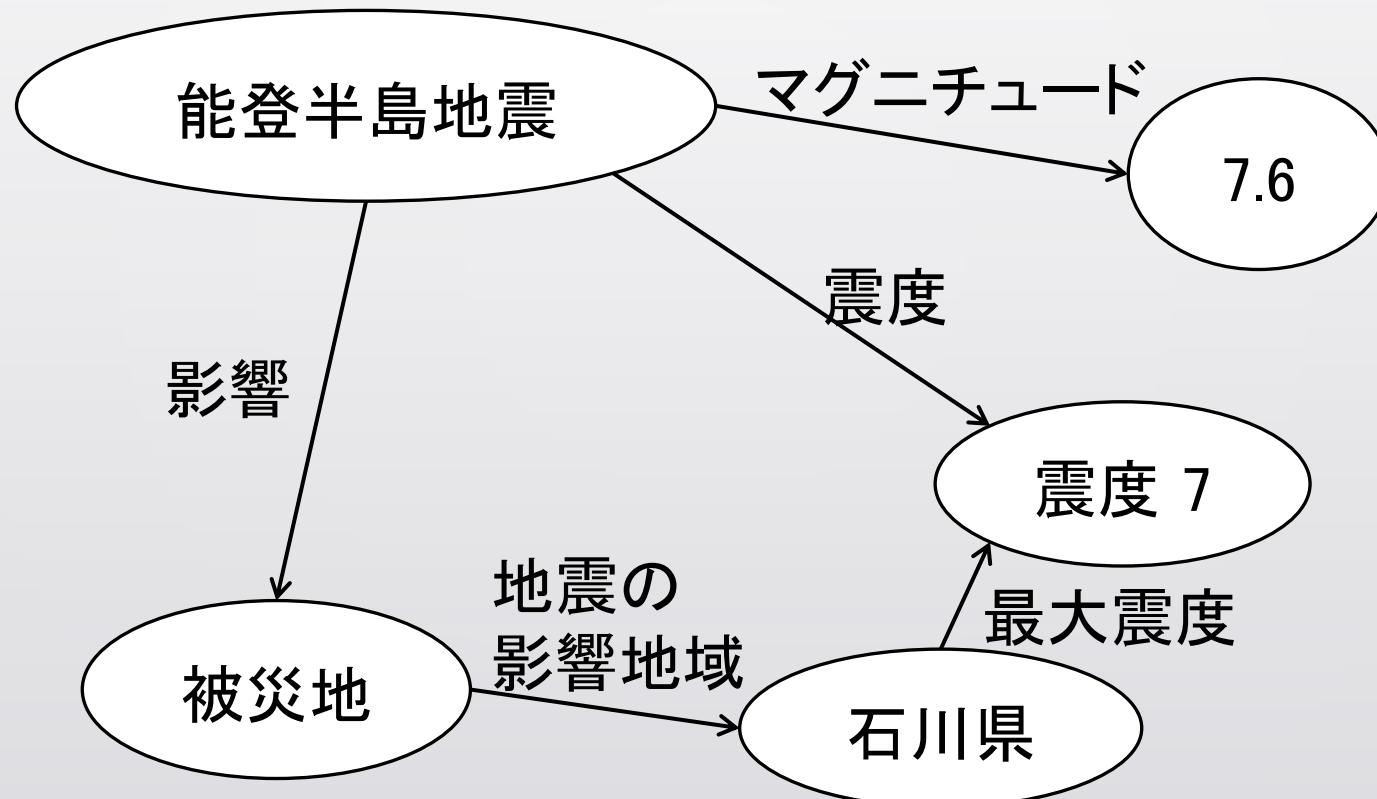
発表の流れ

1. はじめに
- 2. 要素技術**
3. 提案手法
4. 実験
5. 実験結果
6. まとめと今後の課題

要素技術

➤ Knowledge Graph

知識をグラフ構造として表した知識ネットワーク



要素技術

➤ Knowledge Graph

3 つ組 (triple)



head	relation	tail
能登半島地震	マグニチュード	7.6
能登半島地震	震度	震度 7
能登半島地震	影響	被災地
石川県	最大震度	震度 7
石川県	地震の影響地域	被災地
:	:	:

要素技術

➤ 内閣府による 災害データ (PDF)

◆ 発生日時

◆ 場所

◆ 規模

◆ 被害状況

◆ 対応措置

令和6年能登半島地震に係る被害状況等について

※これは速報であり、数値等は今後も変わることがある。

令和6年1月2日
7時00分現在
非常災害対策本部

1 地震の概要 (気象庁情報: 1月2日 7:00 現在)

(1) 1月1日 16時10分の地震

①発生日時

○令和6年1月1日 16:10

②震源及び規模 (暫定値)

○場所: 石川県能登地方 (北緯37.5度、東経137.3度)
○規模: マグニチュード 7.6 (暫定値)
○震源の深さ: 16km (暫定値)

③各地の震度 (震度5強以上)

○石川県 震度7 志賀町
震度6強 七尾市、輪島市、珠洲市、穴水町
震度6弱 中能登町、能登町
震度5強 金沢市、小松市、加賀市、羽咋市、かほく市、能美市、宝達志水町
○新潟県 震度6弱 長岡市
震度5強 新潟中央区、新潟南区、新潟西区、新潟西蒲区、三条市、柏崎市、見附市、燕市、糸魚川市、妙高市、上越市、佐渡市、南魚沼市、阿賀町、刈羽村
○富山県 震度5強 富山市、高岡市、氷見市、小矢部市、南砺市、射水市、舟橋村
○福井県 震度5強 あわら市

(2) 地震活動の状況

○1月1日 16:06 の最大震度5強の地震以降、1月2日 04時00分現在、震度1以上を観測した地震が118回 (震度7: 1回、震度6強: 0回、震度6弱: 0回、震度5強: 3回、震度5弱: 5回、震度4: 18回、震度3: 46回、震度2: 45回、震度1: 0回)

(3) 津波警報等 (1月2日 04:00 時点)

○津波注意報

北海道太平洋沿岸西部、北海道日本海沿岸北部、
北海道日本海沿岸南部、青森県日本海沿岸、秋田県、山形県、
新潟県上中下越、佐渡、富山県、石川県能登、石川県加賀、
福井県、京都府、兵庫県北部、鳥取県、島根県出雲・石見、
隠岐、山口県日本海沿岸、壱岐・対馬

(4) 今後の気象の見通し (1月2日 04:00 時点)

○北陸地方では、2日は、はじめ高気圧に覆われて晴れる所が多いが、次第に雲が広がる。

発表の流れ

1. はじめに
2. 要素技術
- 3. 提案手法**
4. 実験
5. 実験結果
6. まとめと今後の課題

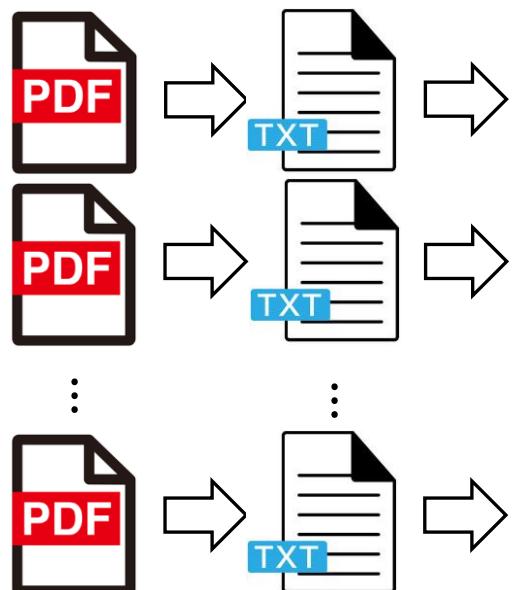
提案手法

- ◆ Knowledge Graph は
異なるデータの相互関係の明確化が可能
- ◆ 人手を介さずに処理を可能にするため
GPT-4o を利用してデータから triple を抽出
- ◆ 異なるデータ間の関係性を見るため
Knowledge Graph として可視化

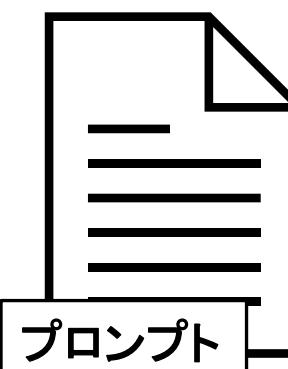
提案手法

◆モデル概略図

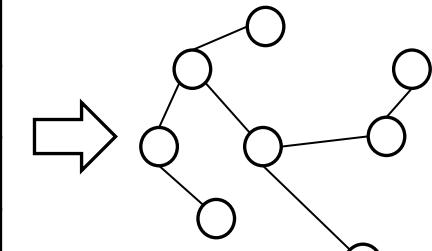
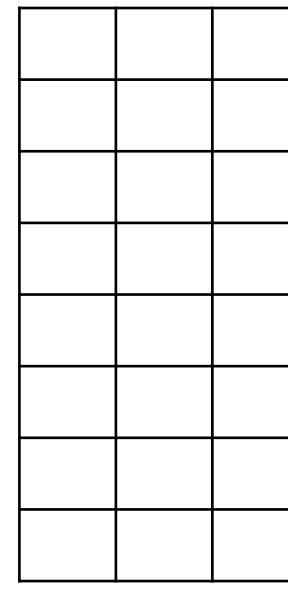
災害データ



GPT-4o



Knowledge Graph



発表の流れ

1. はじめに
2. 要素技術
3. 提案手法
- 4. 実験**
5. 実験結果
6. まとめと今後の課題

実験

◆ 実験データ

- 2024 年 1 月 1 日に発生した能登半島地震の災害報告書
- 7 つのデータ (作成日: 1/2, 1/8, 2/2, 4/2, 6/4, 8/21, 10/1)

◆ 実験 1

- 災害報告書 : PDF → テキストデータ
- 災害報告書を GPT-4o の入力として triple を抽出

◆ 実験 2

- 実験 1 で抽出した triple からの Knowledge Graph の可視化

実験データ

- 内閣府による災害データ (PDF)
- 2024年1月1日に発生した能登半島地震の災害報告書

【2024年1月2日】

(3)津波警報等 (1月2日 04:00 時点)

○津波注意報

北海道太平洋沿岸西部、北海道日本海沿岸北部、
北海道日本海沿岸南部、青森県日本海沿岸、秋田県、山形県、
新潟県上中下越、佐渡、富山県、石川県能登、石川県加賀、
福井県、京都府、兵庫県北部、鳥取県、島根県出雲・石見、
隠岐、山口県日本海沿岸、壱岐・対馬

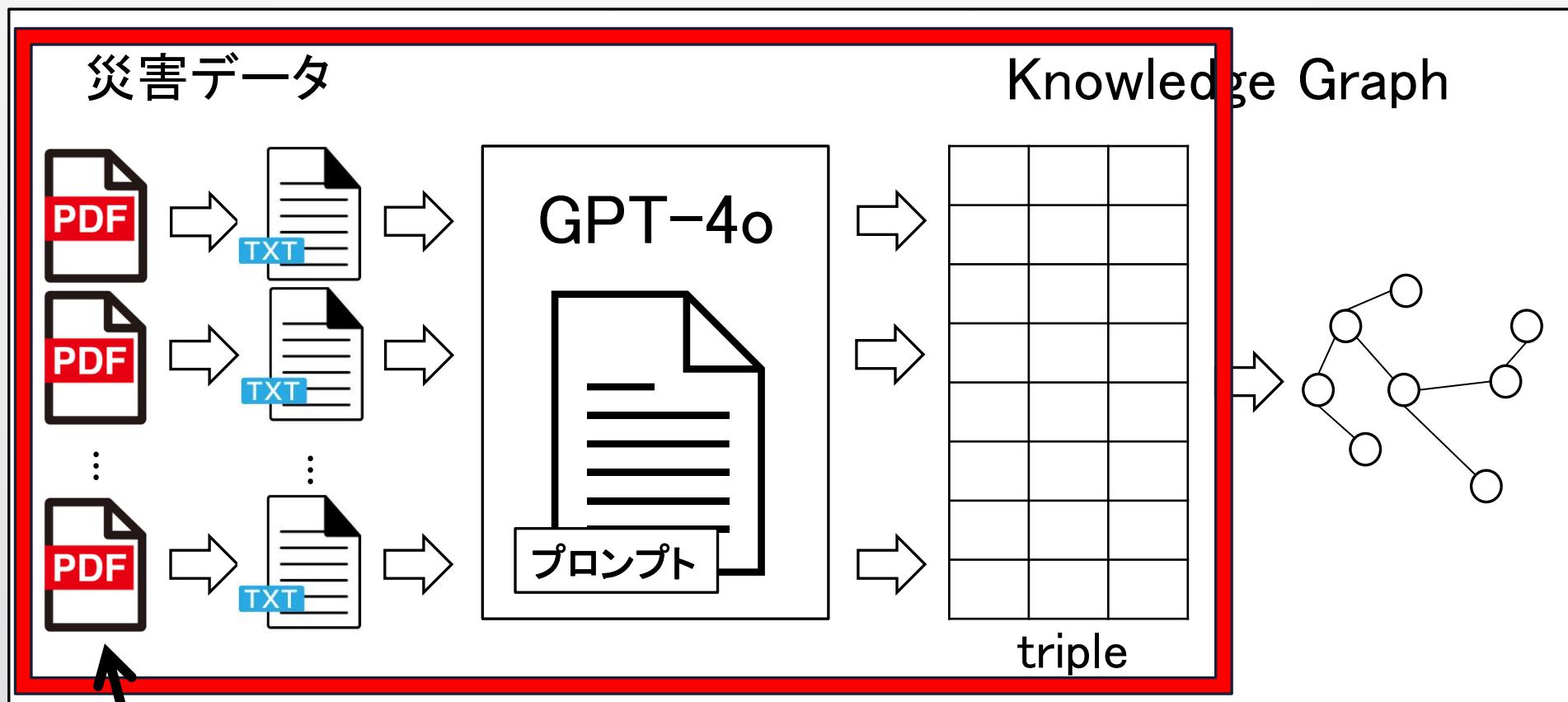
【2024年1月8日】

(3)津波警報等 (1月2日 10:00 時点)

○1日 16時10分に発生した石川県能登地方を震源とする地震の津波注意報は、2日 10時00分に全て解除。

実験 1

災害報告書を GPT-4o の入力として triple を抽出



入力: 各災害報告書の最初 10 ページ

実験 1 (プロンプト)

あなたは災害報告書の中で重要な用語とその関係性を抽出し、ナレッジグラフとして作成する役割を担っています。このナレッジグラフは、災害データの保存・継承を目的とし、特に災害同士の関係性を明確にすることを目指しています。

入力文: 災害の報告書

以下の条件に従って、重要な用語とその関係性を整理してください。

⋮

三つ組は50個以下にしてください。

条件: 「災害名称」「発生日時」のような重要な entity の抽出

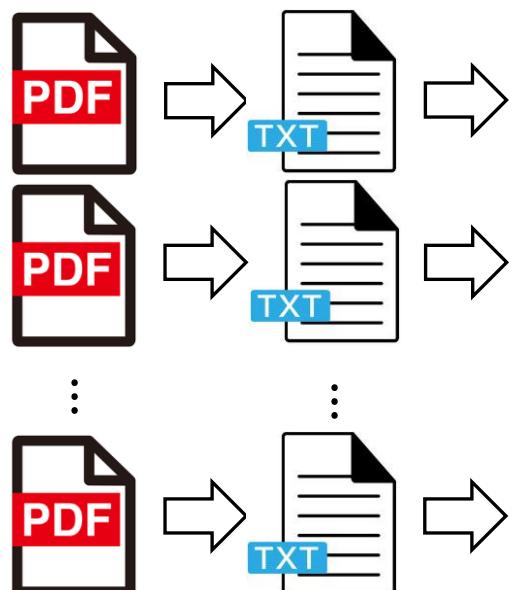
条件: entity 間の関係性の推測

条件: 各 triple に対する重要度 (1~5) の設定

実験 2

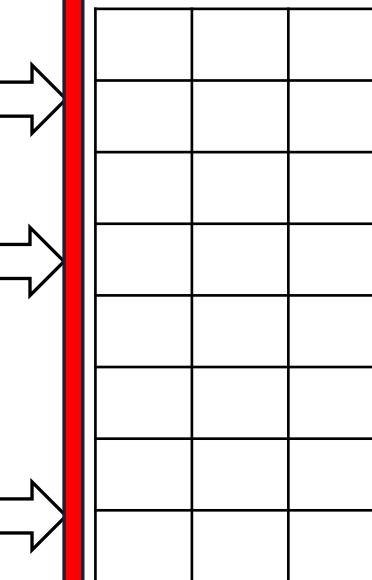
実験 1 で抽出した triple からの Knowledge Graph の可視化

災害データ



GPT-4o
プロンプト

Knowledge Graph



triples

実験 2 (重要度 = ノードの大きさ)

$$I(E_k) = \frac{\sum_{j=1}^n I(T_{E_k j})}{n}$$

E_k : k 番目の entity

$T_{E_k j}$: E_k を含む triple のうち, j 番目の triple

n : $T_{E_k j}$ の個数

$I(T_{E_k j})$: $T_{E_k j}$ の重要度

$I(E_k)$: の重要度

発表の流れ

1. はじめに
2. 要素技術
3. 提案手法
4. 実験
- 5. 実験結果**
6. まとめと今後の課題

実験 1 の結果

抽出された triple の総数

2024 年	1 月 2 日	1 月 8 日	2 月 2 日	4 月 2 日
triple 数	20	20	20	30

2024 年	6 月 4 日	8 月 21 日	10 月 1 日	合計
triple 数	20	30	24	164

実験 1 の結果

* : 重要度

22

2024 年 1 月 2 日

head	relation	tail	*
令和 6 年能登半島地震	地震の発生日時	発生日時	5
令和 6 年能登半島地震	地震の震源地	震源地	5
令和 6 年能登半島地震	震度 7 を観測	石川県	5
令和 6 年能登半島地震	津波の発生可能性	津波注意報	4
:	:	:	:

2024 年 1 月 8 日

head	relation	tail	*
令和 6 年能登半島地震	地震の発生日時	発生日時	5
石川県	最大震度	震度 7	5
津波注意報	津波注意報の解除	解除日時	3
停電	停電地域	石川県	4
:	:	:	:

実験 1 の結果

* : 重要度

23

2024 年 1 月 2 日

head	relation	tail	*
令和 6 年 能登半島 地震	<u>地震の</u> <u>発生日時</u>	<u>発生日時</u>	<u>5</u>
令和 6 年 能登半島 地震	<u>地震の</u> <u>震源地</u>	<u>震源地</u>	<u>5</u>
令和 6 年 能登半島 地震	<u>震度 7 を</u> <u>観測</u>	<u>石川県</u>	<u>5</u>
令和 6 年 能登半島 地震	<u>津波の</u> <u>発生可能性</u>	<u>津波注意報</u>	<u>4</u>
:	:	:	:

2024 年 1 月 8 日

head	relation	tail	*
令和 6 年 能登半島 地震	<u>地震の</u> <u>発生日時</u>	<u>発生日時</u>	<u>5</u>
石川県	<u>最大震度</u>	<u>震度 7</u>	<u>5</u>
<u>津波</u> <u>注意報</u>	<u>津波注意報</u> <u>の解除</u>	<u>解除日時</u>	<u>3</u>
停電	<u>停電地域</u>	<u>石川県</u>	<u>4</u>
:	:	:	:

実験 1 の結果

* : 重要度

24

2024 年 1 月 2 日

head	relation	tail	*
令和 6 年能登半島地震	地震の発生日時	発生日時	5
令和 6 年能登半島地震	地震の震源地	震源地	5
令和 6 年能登半島地震	震度 7 を観測	石川県	5
令和 6 年能登半島地震	<u>津波の発生可能性</u>	<u>津波注意報</u>	4
:	:	:	:

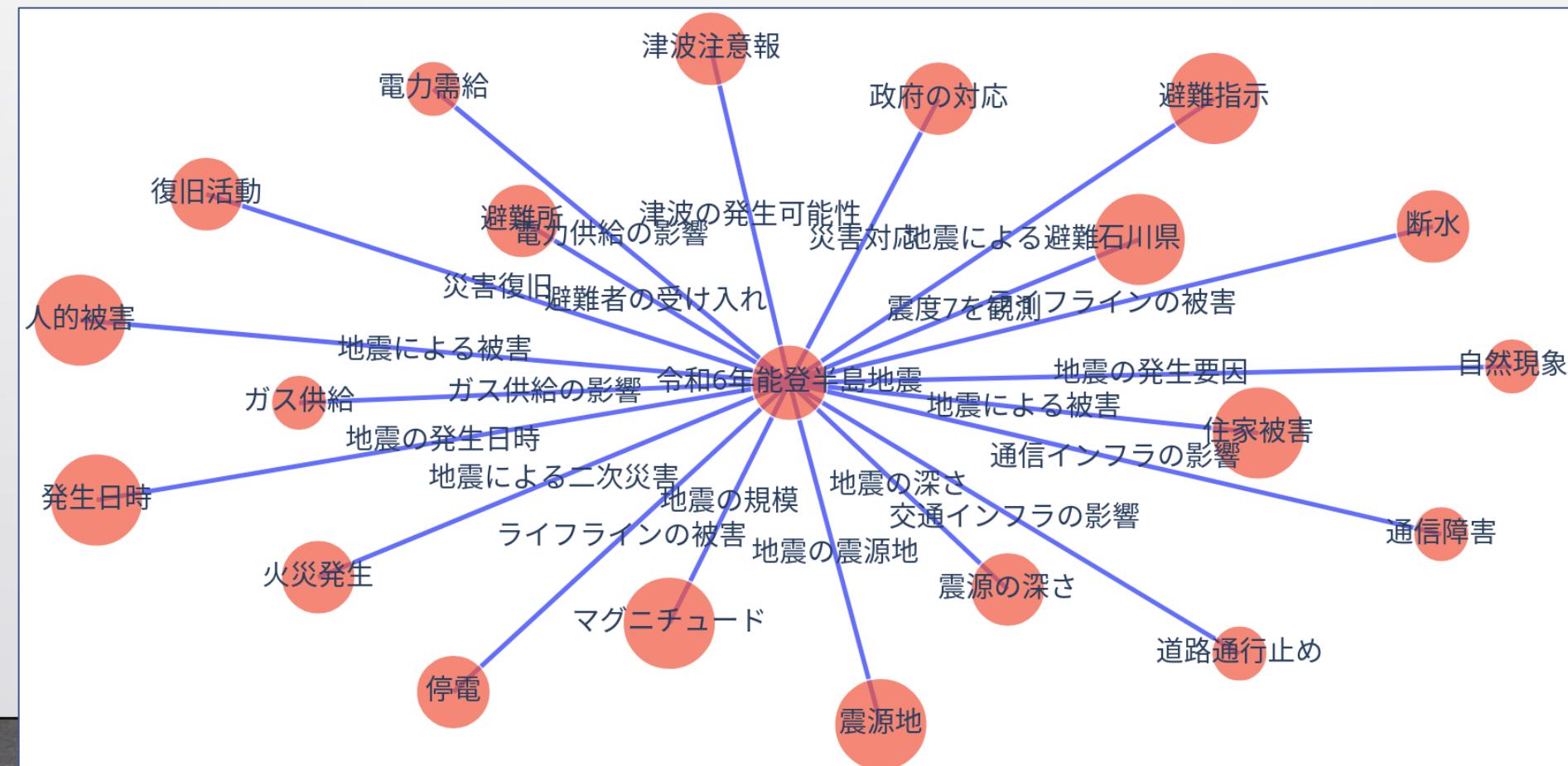
2024 年 1 月 8 日

head	relation	tail	*
令和 6 年能登半島地震	地震の発生日時	発生日時	5
石川県	最大震度	震度 7	5
<u>津波注意報</u>	<u>津波注意報の解除</u>	<u>解除日時</u>	3
停電	停電地域	石川県	4
:	:	:	:

実験 2 の結果

➤各災害データの Knowledge Graph の可視化

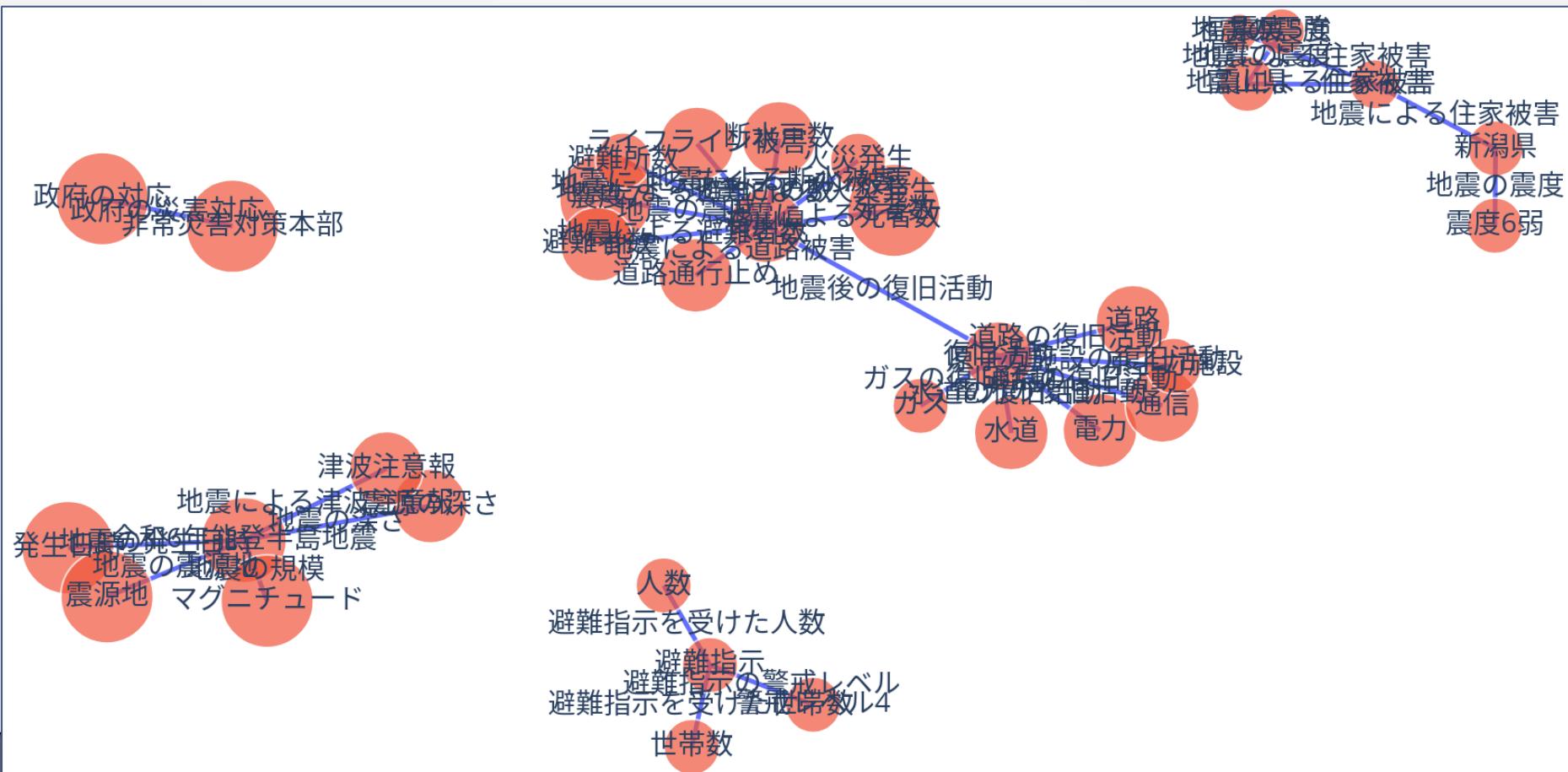
1月2日



実験 2 の結果

➤ 各災害データの Knowledge Graph の可視化

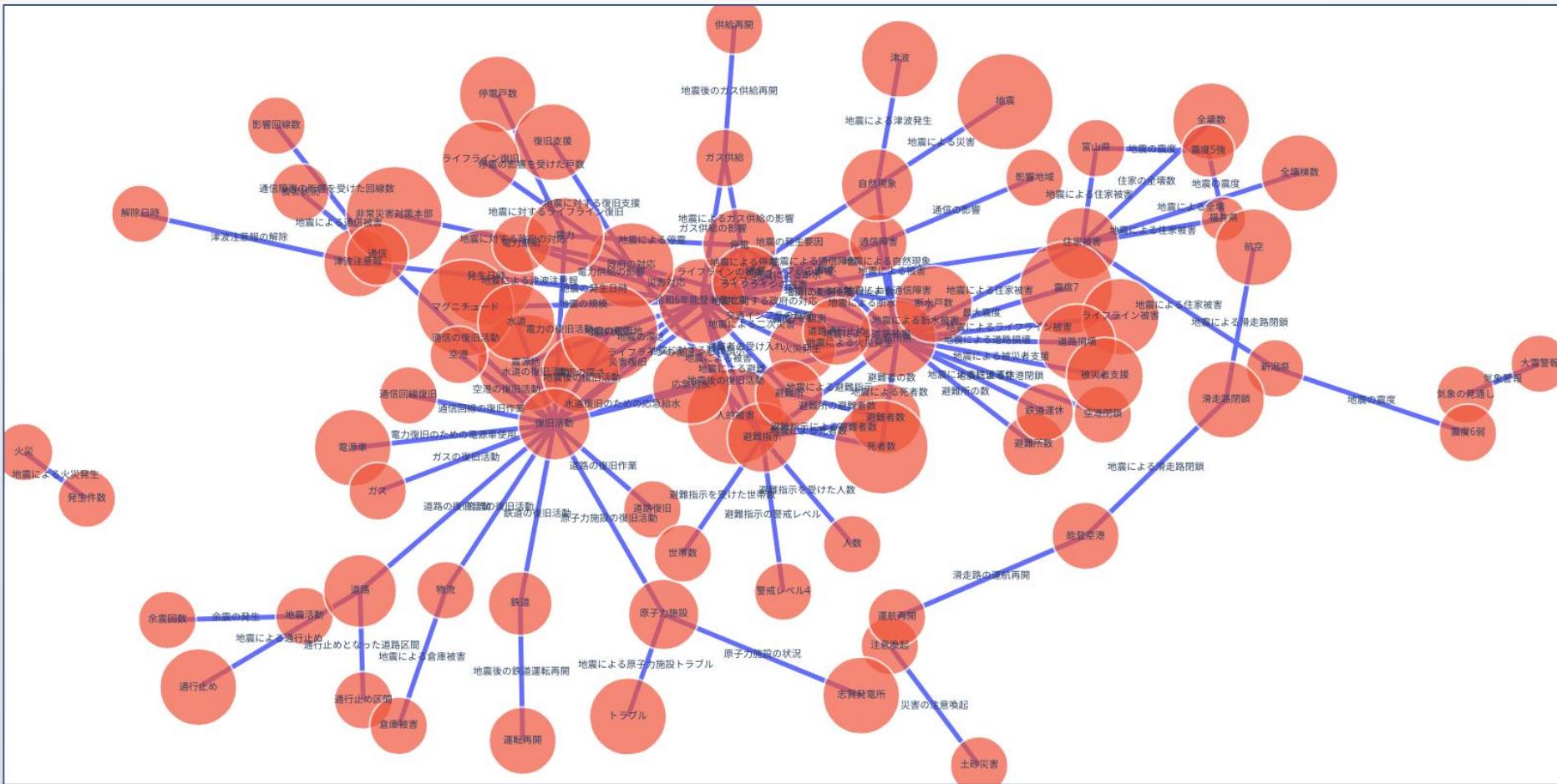
4月2日



実験 2 の結果

27

➤ 統合した Knowledge Graph の可視化



実験結果

□ 実験 1

- 時間経過による情報量の増加に伴い triple 数も増加
- 各災害データから共通した triple の抽出
- 時間経過により変化した情報

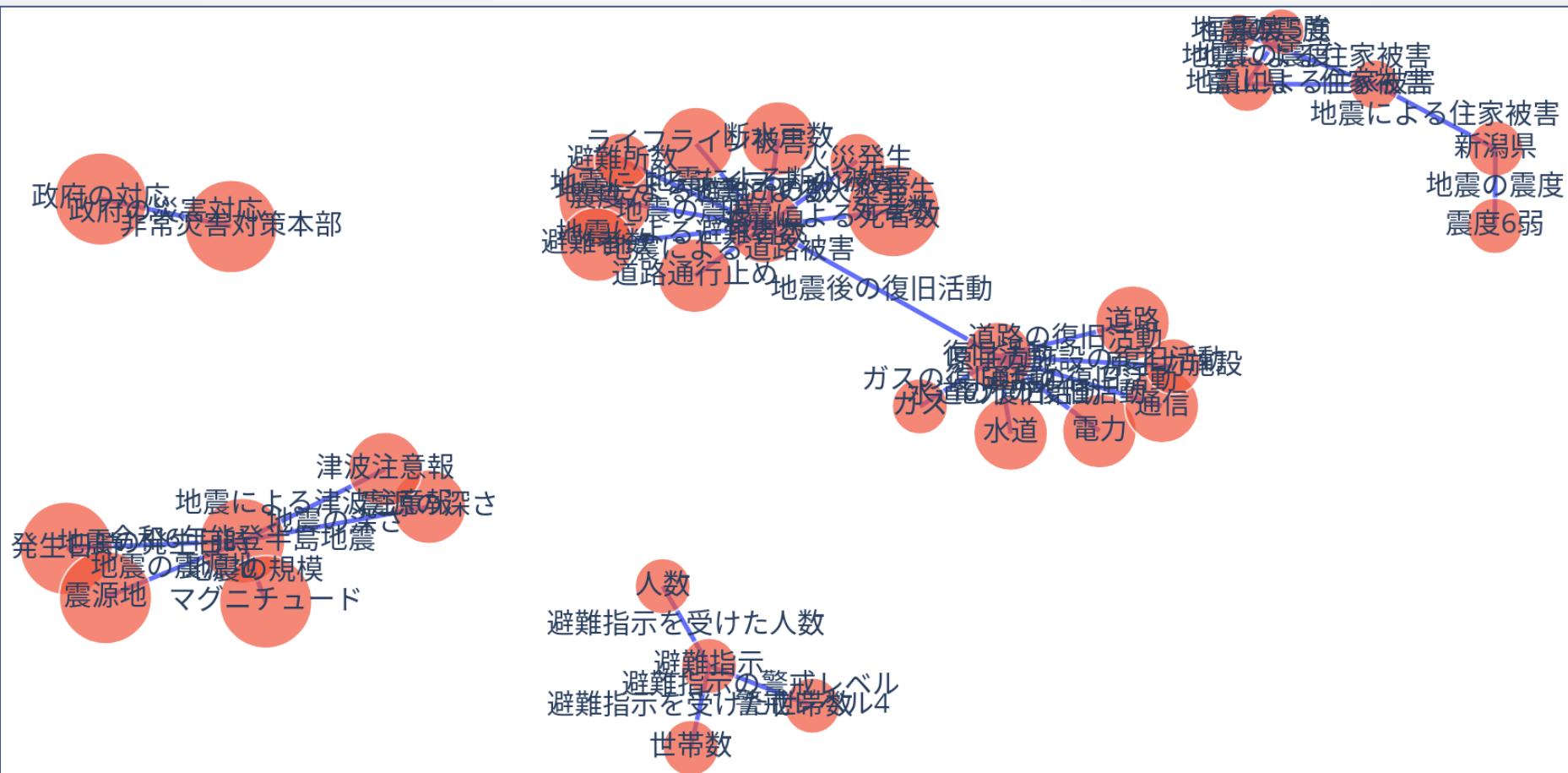
□ 実験 2

- 1月 2日の災害データでは 1つのネットワーク
- 複数の小規模な entity 群による Knowledge Graph
- 統合すると災害情報が網羅的に表現

考察

➤各災害データの Knowledge Graph の可視化

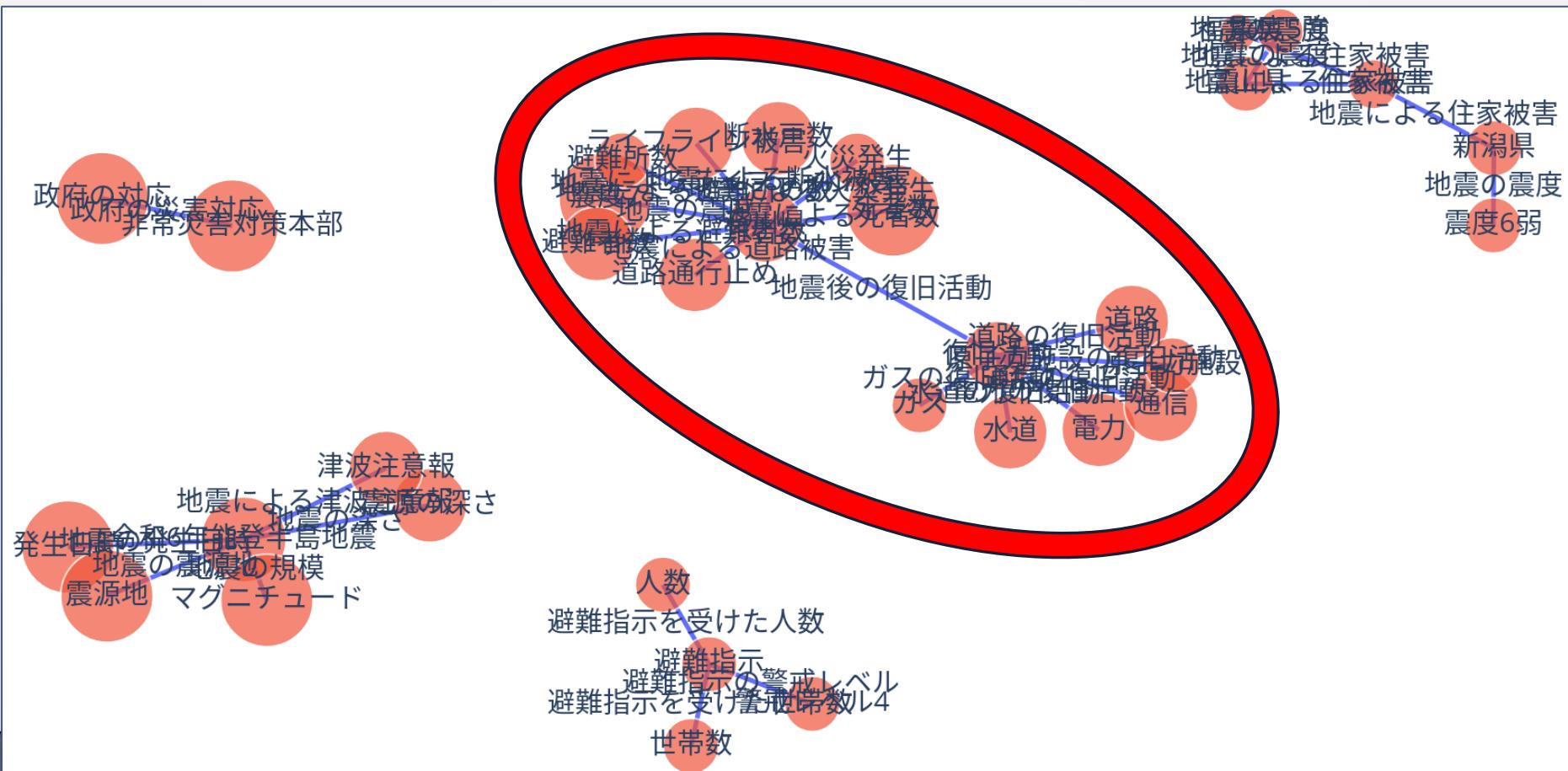
4月2日



考察

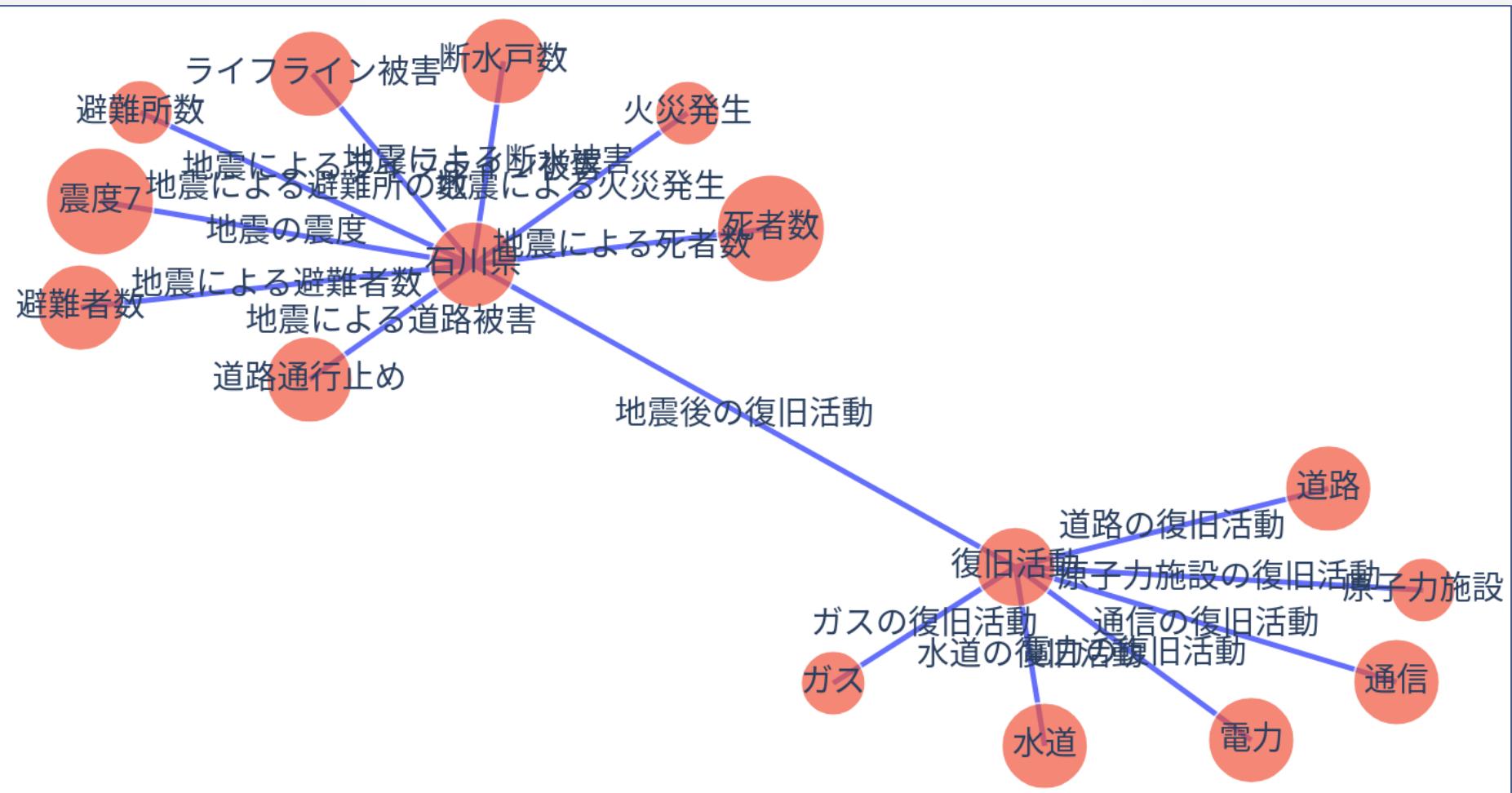
▶各災害データの Knowledge Graph の可視化

4月2日



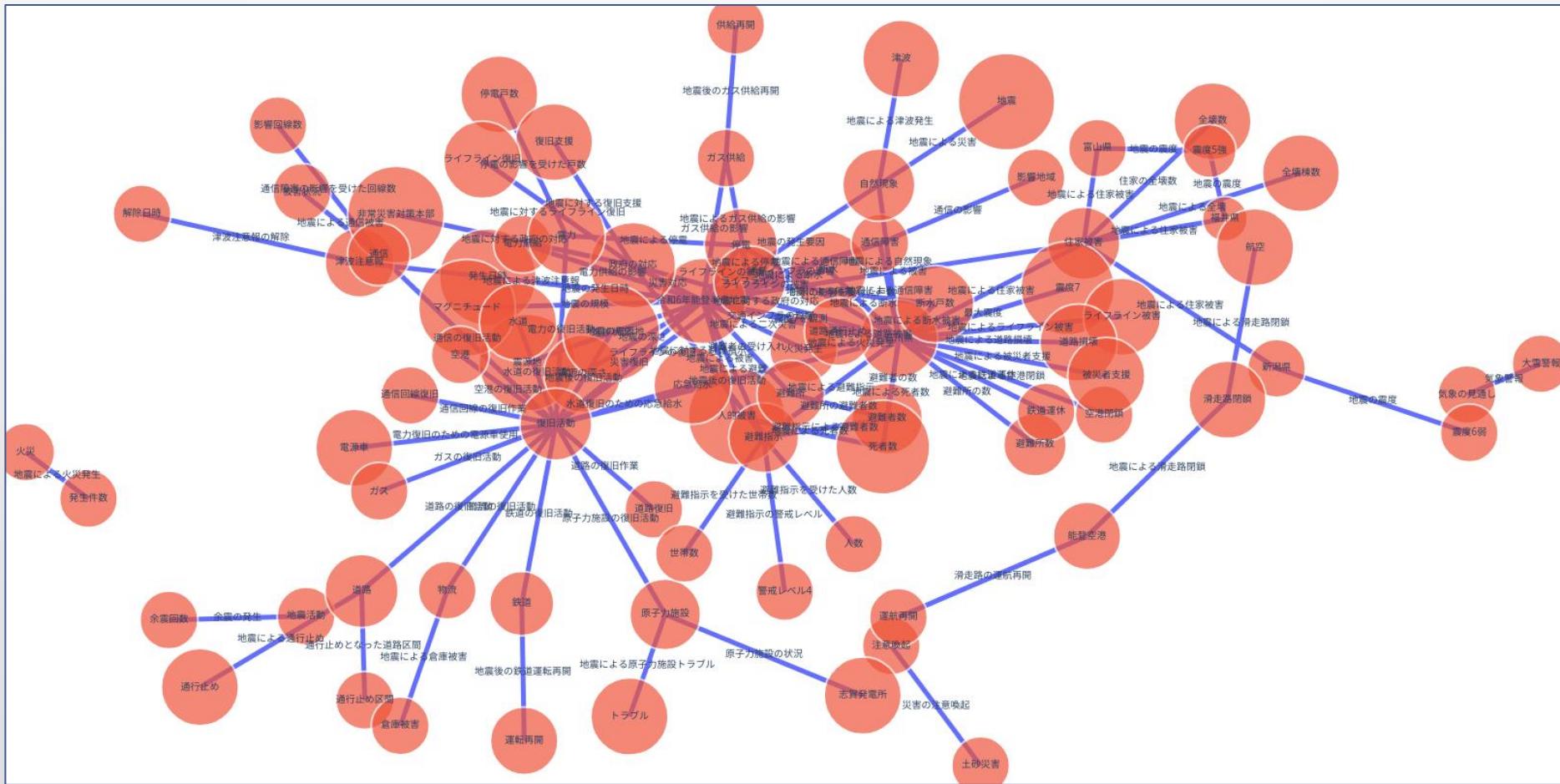
考察(4月2日の拡大図)

“石川県 – 地震後の復旧活動 – 復旧活動” 中心



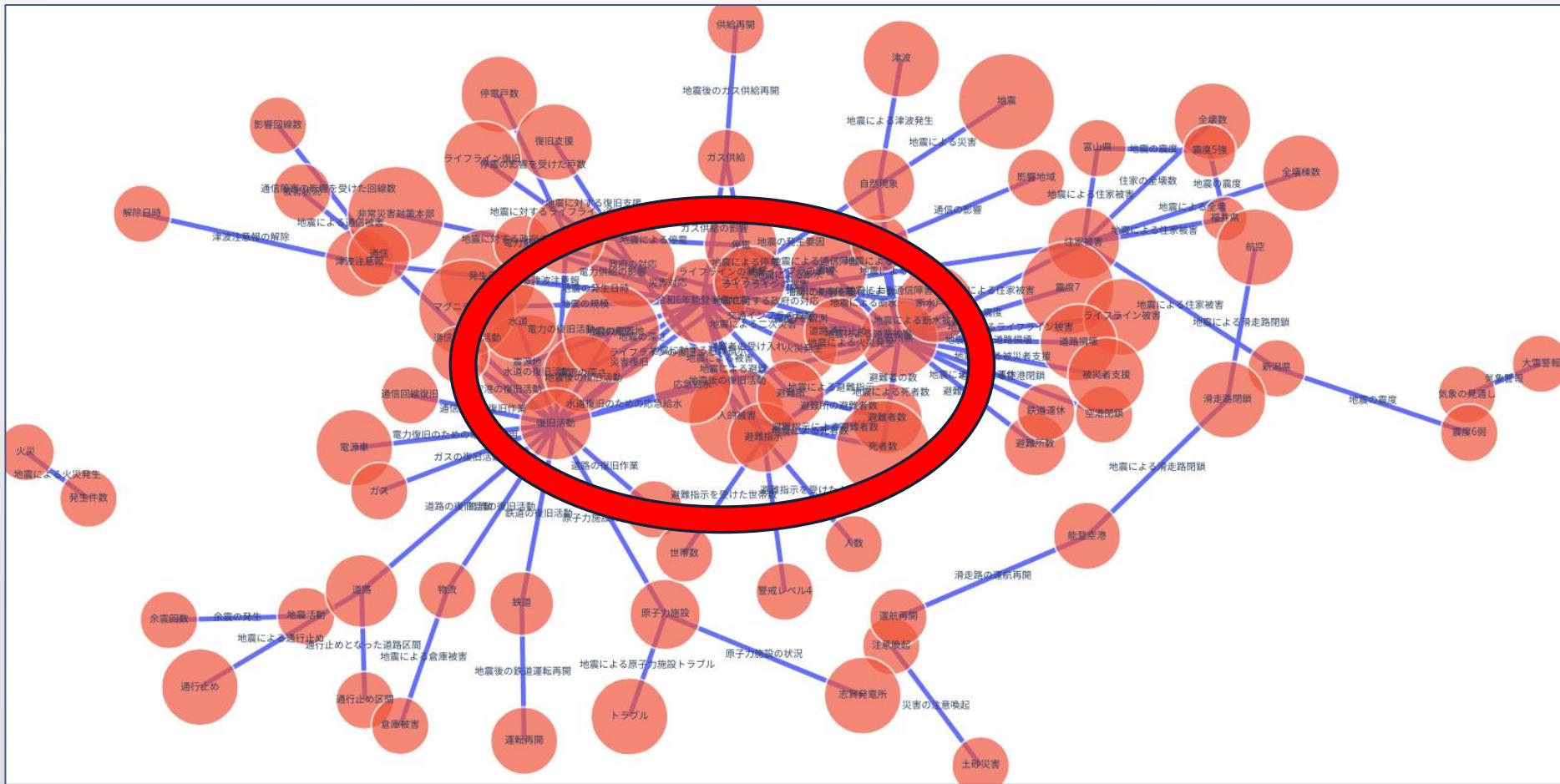
考察

▶ 統合した Knowledge Graph の可視化

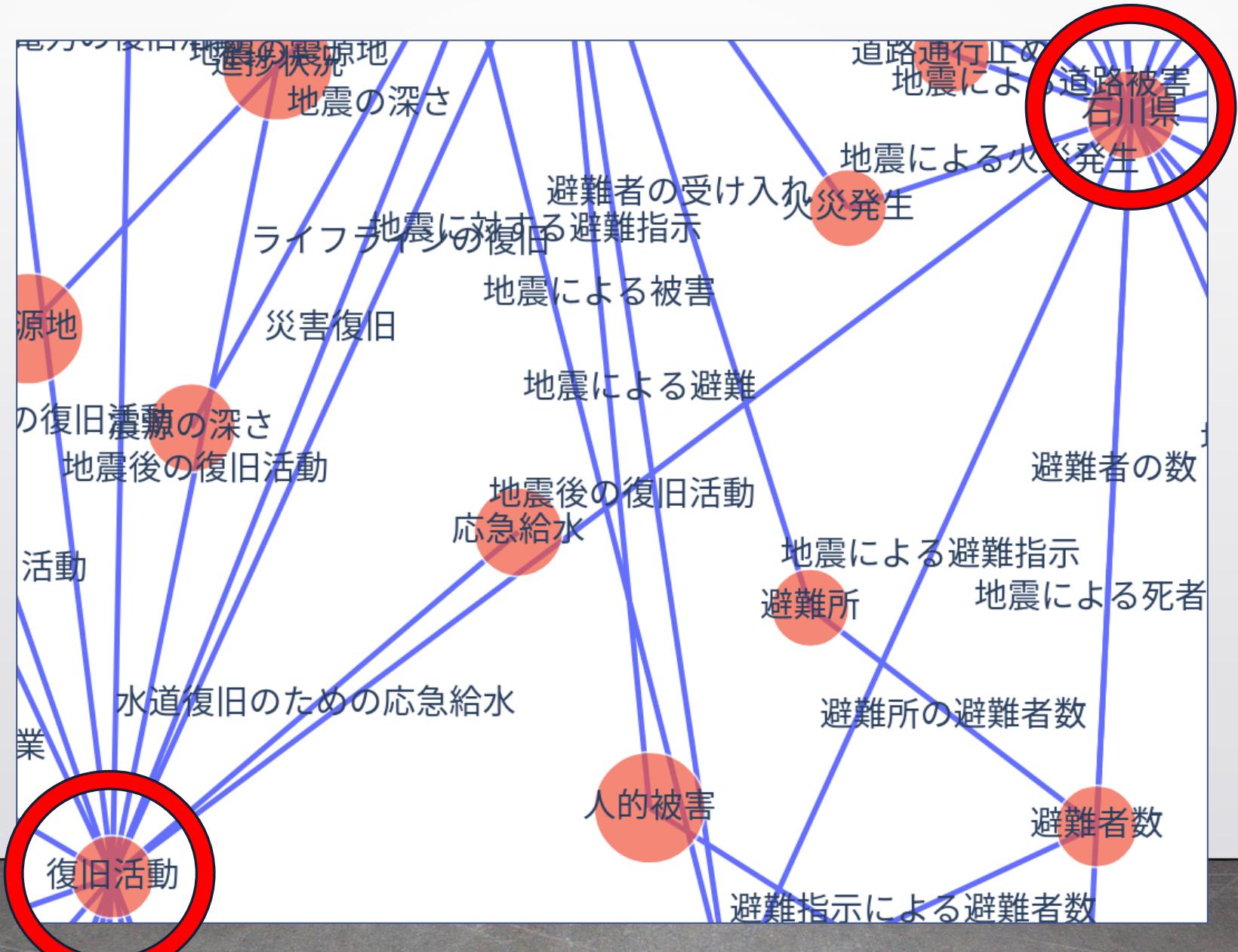


考察

▶ 統合した Knowledge Graph の可視化



考察（統合した Knowledge Graph の拡大図）



発表の流れ

1. はじめに
2. 要素技術
3. 提案手法
4. 実験
5. 実験結果
6. まとめと今後の課題

まとめと今後の課題

➤ まとめ

- ◆ 作成時点の異なる災害報告書から triple を抽出し,
Knowledge Graph として災害データを可視化
- ◆ 重要かつ時系列で変化する triple を抽出
- ◆ 時間経過で変化する災害データの情報を集約

➤ 今後の課題

- ◆ より適切な triple 取得のためのプロンプト調整
- ◆ 時系列ごとの Knowledge Graph の変化の可視化

ご清聴ありがとうございました。

備考

実験データ

- 7つのデータ(作成日: 1/2, 1/8, 2/2, 4/2, 6/4, 8/21, 10/1)

	1/2	1/8	2/2	4/2	6/4	8/21	10/1
ページ数	37	92	164	201	224	230	231
文字数							

実験 1 (プロンプト)

あなたは災害報告書の中で重要な用語とその関係性を抽出し、ナレッジグラフとして作成する役割を担っています。このナレッジグラフは、災害データの保存・継承を目的とし、特に災害同士の関係性を明確にすることを目指しています。

入力文: 災害の報告書

以下の条件に従って、重要な用語とその関係性を整理してください。

⋮

三つ組は50個以下にしてください。

条件 1: 災害同士の関係性を明確にするため、以下の項目に該当する幅広い用語を抽出してください。

条件 2: 抽出した用語が他の用語とどのような関係を持つかを推測し、単語で関係性を表現してください。

条件 4: 作成した各三つ組について重要度を1～5の数値で評価してください。

実験 1 (プロンプト)

あなたは災害報告書の中で重要な用語とその関係性を抽出し、ナレッジグラフとして作成する役割を担っています。このナレッジグラフは、災害データの保存・継承を目的とし、特に災害同士の関係性を明確にすることを目指しています。

入力文: 災害の報告書

以下の条件に従って、重要な用語とその関係性を整理してください。

条件1

災害同士の関係性を明確にするため、以下の項目に該当する幅広い用語を抽出してください。例としては「災害名称」「発生日時」「被害内容」「被災地」「被害の程度」「被災者数」「避難指示」「復旧活動」「発生要因」「政府の対応」「自然現象(津波、台風など)」です。これに加えて新しい要素も含めてください。

実験 1 (プロンプト)

条件2

抽出した用語が他の用語とどのような関係を持つかを推測し、単語で関係性を表現してください。

例:「地震—マグニチュード—5.5」他にも「発生日時」と「復旧期間」や「被害内容」と「復旧活動」「被災地」と「避難指示」「災害名称」と「政府の対応」などの関係性が考えられます。

条件3

「地震—震源の深さ—深さ」のような単語の説明を表す関係や、「地震—地震による深さ—深さ」といった意味のない表現は省略し、具体的な関係を表すようにしてください。

例:「地震—震源地—石川県能登半島」。また、インフラ損傷や復旧活動、被災者支援といった被害の影響や対応も考慮に入れてください。

条件4

作成した各三つ組について重要度を1～5の数値で評価してください。

他の災害と深く関係がある三つ組を重要度5とし、浅い関係は1とします。必ず1以上の数値を含めるように評価してください。

実験 1 (プロンプト)

条件5

例えば「地震-発生時刻-0:00」と「地震-マグニチュード-5.0」のように共通する用語を持つ三つ組があれば、0:00と5.0の関係性も考慮してください。

条件6

このナレッジグラフはすべての要素がつながっているようにしてください。

条件7

震度やマグニチュード、発生日時のような数値は重要な要素なので必ず抜き出してください。
例:「地震-マグニチュード-5.5」

実験 1 (プロンプト)

出力形式

以下のようなJSON形式で出力してください。三つ組は50個以下にしてください。

```
[  
  {  
    "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)"  
  },  
  {...}  
]
```

実験 1 の結果

➤ PDF から変換したテキストデータ

1 / 37

令和6年能登半島地震に係る被害状況等について

※これは速報であり、数値等は今後も変わることがある。

令和6 年1月2日 7時00分現在

非常災害対策本部

1 地震の概要（気象庁情報：1月2日 7:00現在）

(1)1月1日16時10分の地震 発生日時

○令和6年1月1日16:10 震源及び規模(暫定値)

○場所：石川県能登地方(北緯 37.5度、東経 137.3度)

○規模：マグニチュード 7.6(暫定値)

○震源の深さ：16 km(暫定値) 各地の震度(震度 5強以上)

○石川県 震度7 志賀町 震度6強 七尾市、輪島市、珠洲市、穴水町 震度6弱 中能登町、能登町 震度5強 金沢市、小松市、加賀市、羽咋市、かほく市、能美市、宝達志水町

○新潟県 震度6弱 長岡市 震度5強 新潟中央区、新潟南区、新潟西区、新潟西蒲区、三条市、柏崎市、見附市、燕市、糸魚川市、妙高市、上越市、佐渡市、南魚沼市、阿賀町、刈羽村

：

実験 1 の結果

* : 重要度 48

2024 年 4 月 2 日

head	relation	tail	*
令和 6 年 能登半島地震	地震の 発生日時	発生日時	5
石川県	地震の 震度	震度 7	5
石川県	地震後の 復旧活動	復旧活動	4
復旧活動	電力の 復旧活動	電力	4
:	:	:	:

実験 1 の結果

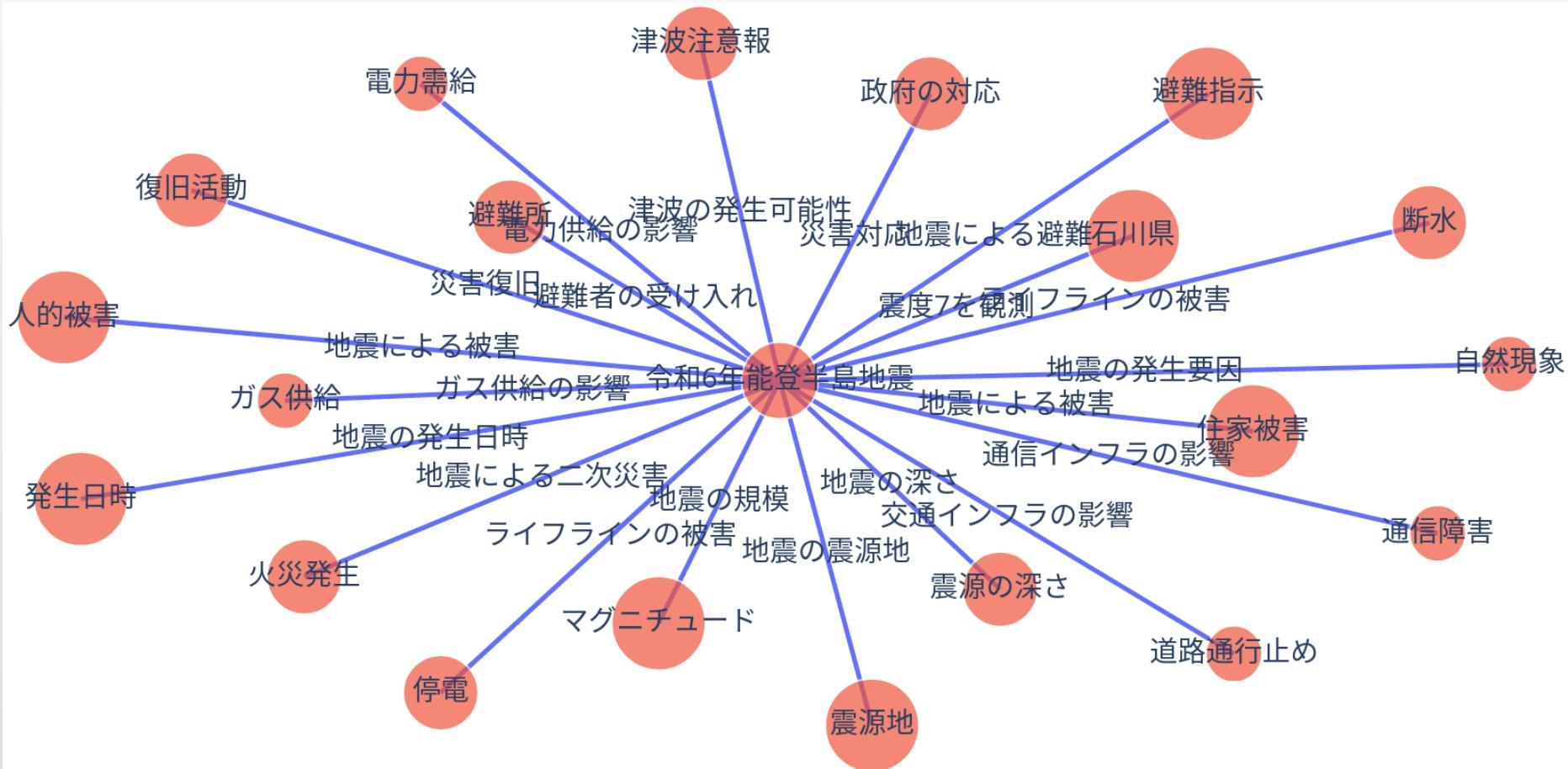
* : 重要度 49

2024 年 4 月 2 日

head	relation	tail	*
令和 6 年 能登半島地震	地震の 発生日時	発生日時	5
石川県	地震の 震度	震度 7	5
石川県	<u>地震後の 復旧活動</u>	<u>復旧活動</u>	4
<u>復旧活動</u>	<u>電力の 復旧活動</u>	<u>電力</u>	4
:	:	:	:

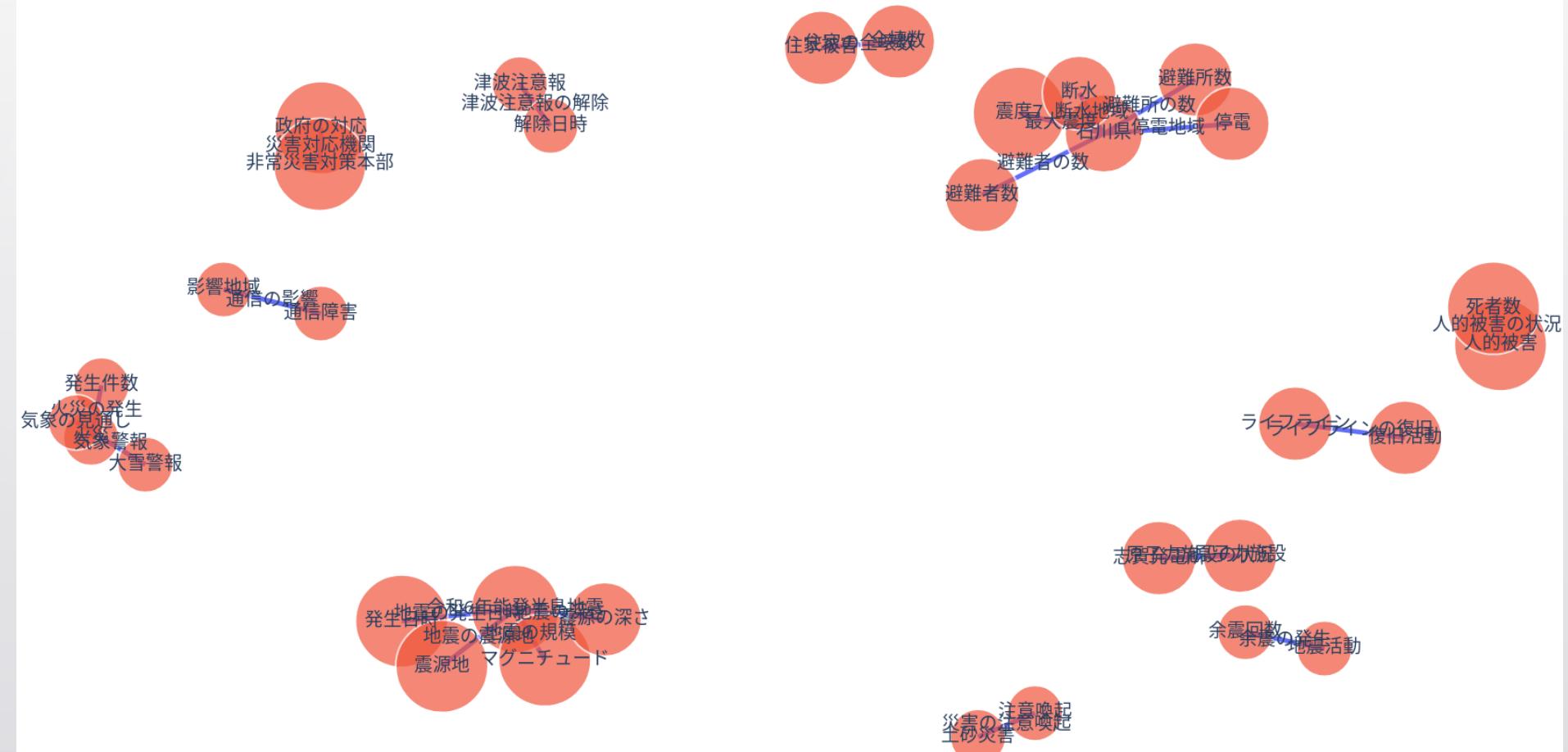
►各災害データの Knowledge Graph の可視化

1月 2日



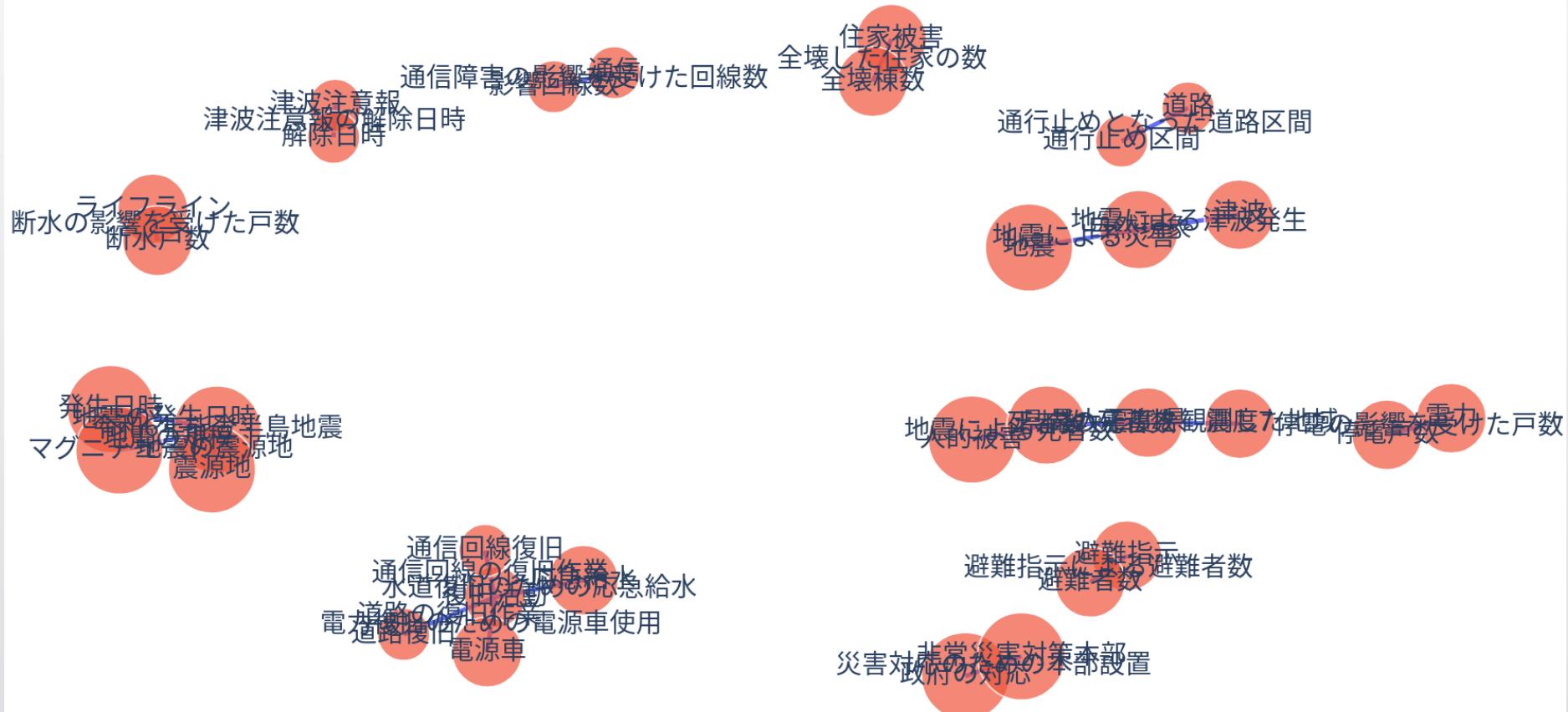
►各災害データの Knowledge Graph の可視化

1月 8日



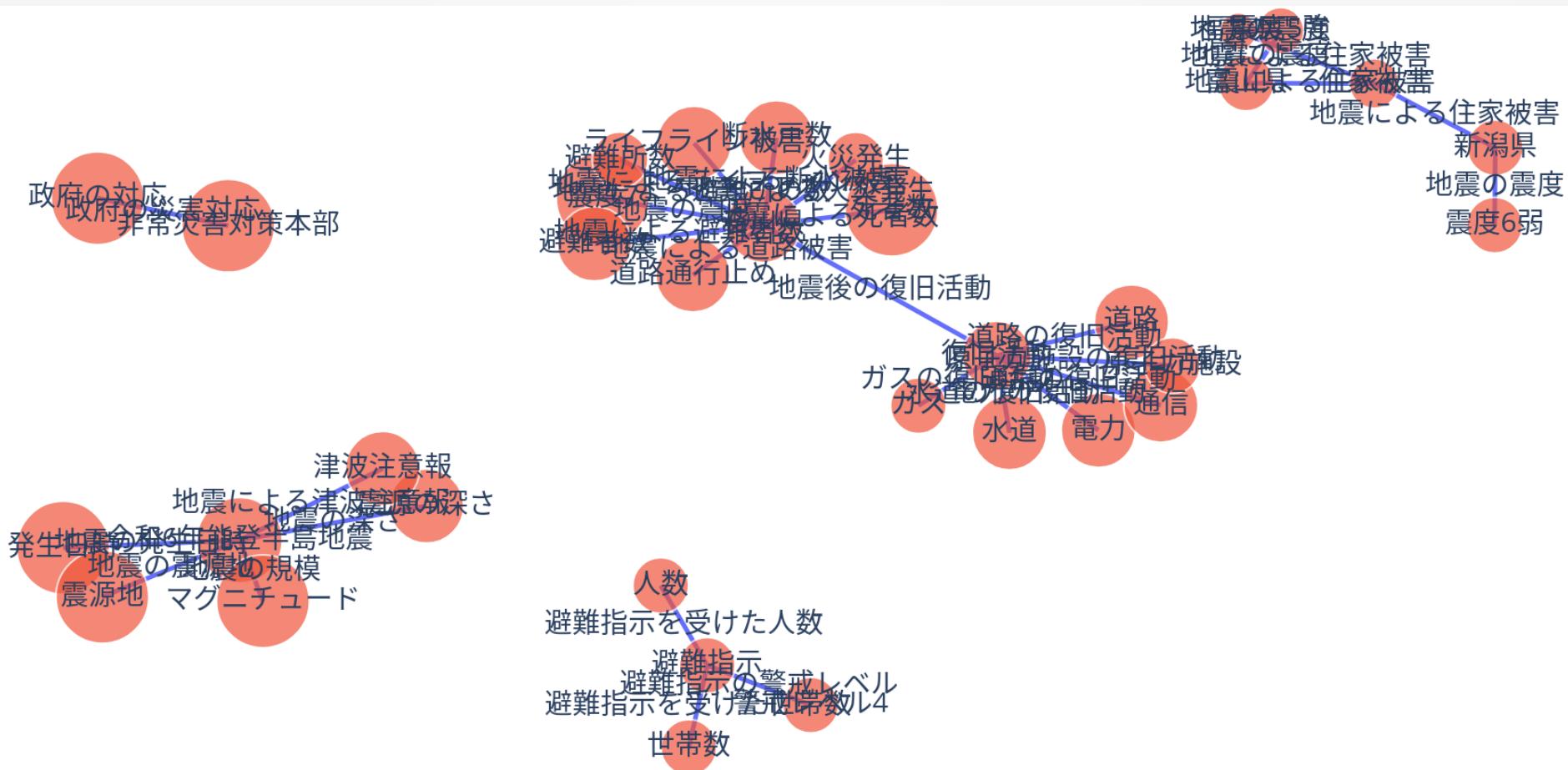
►各災害データの Knowledge Graph の可視化

2月2日



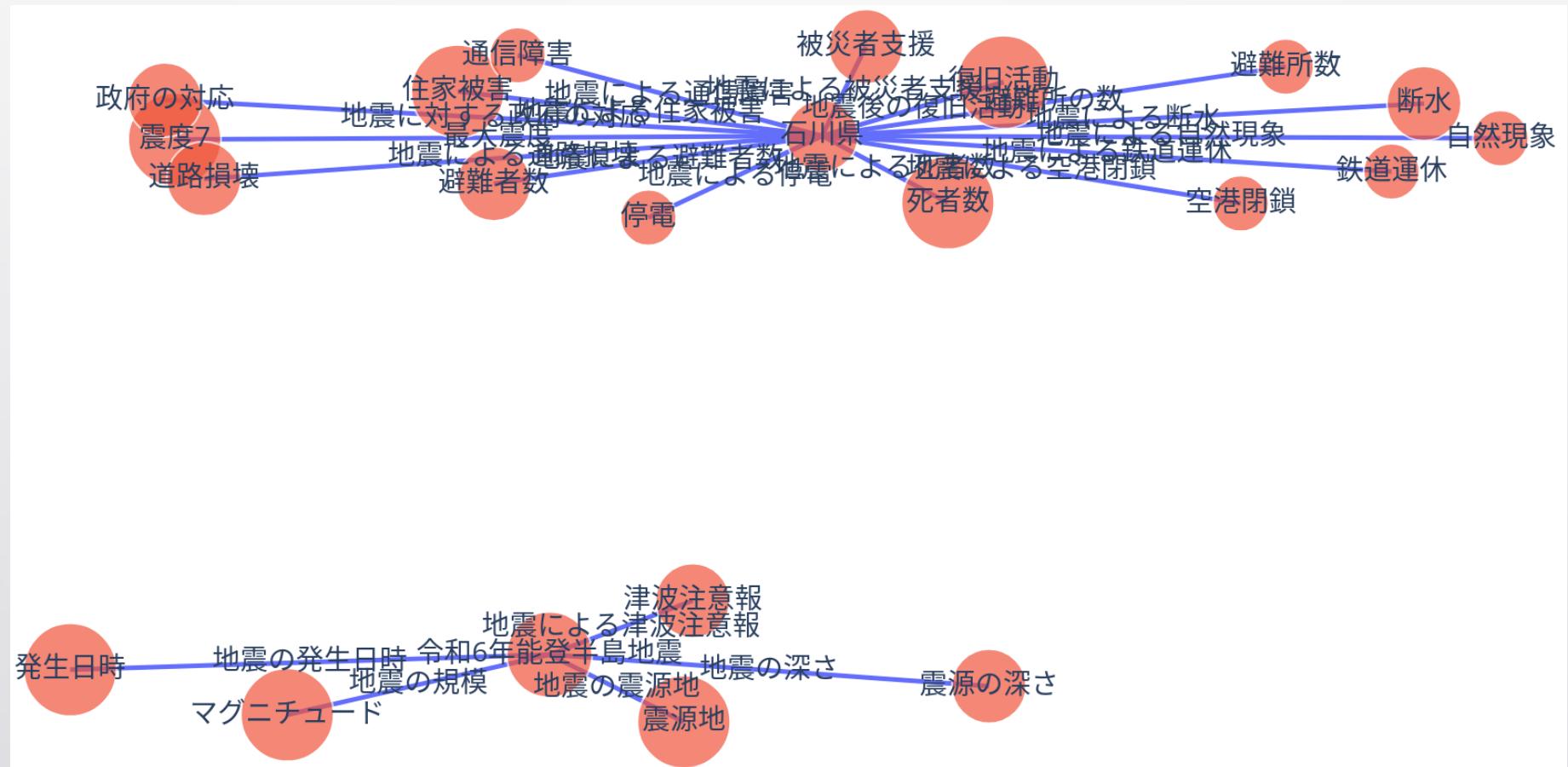
▶各災害データの Knowledge Graph の可視化

4月2日



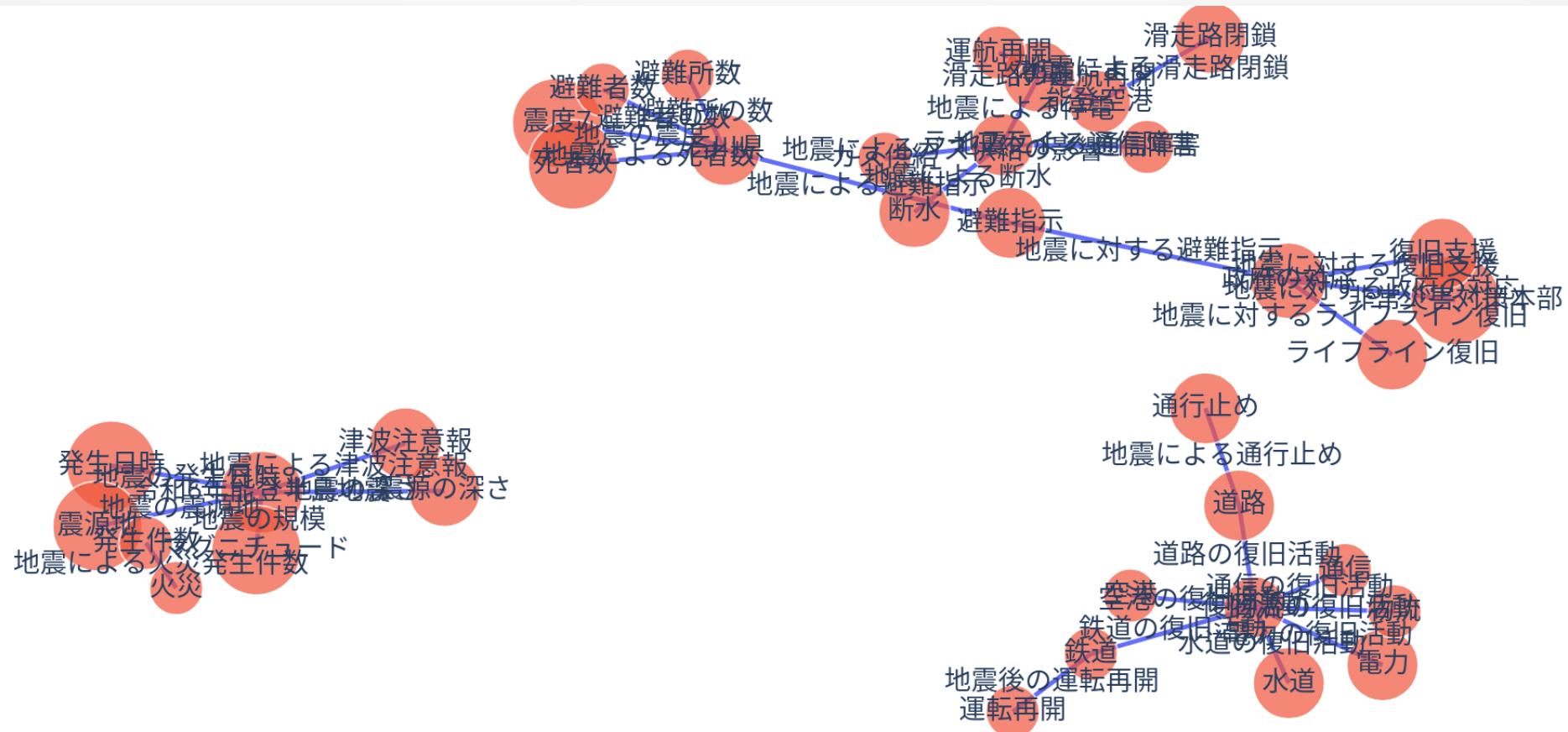
▶各災害データの Knowledge Graph の可視化

6月4日



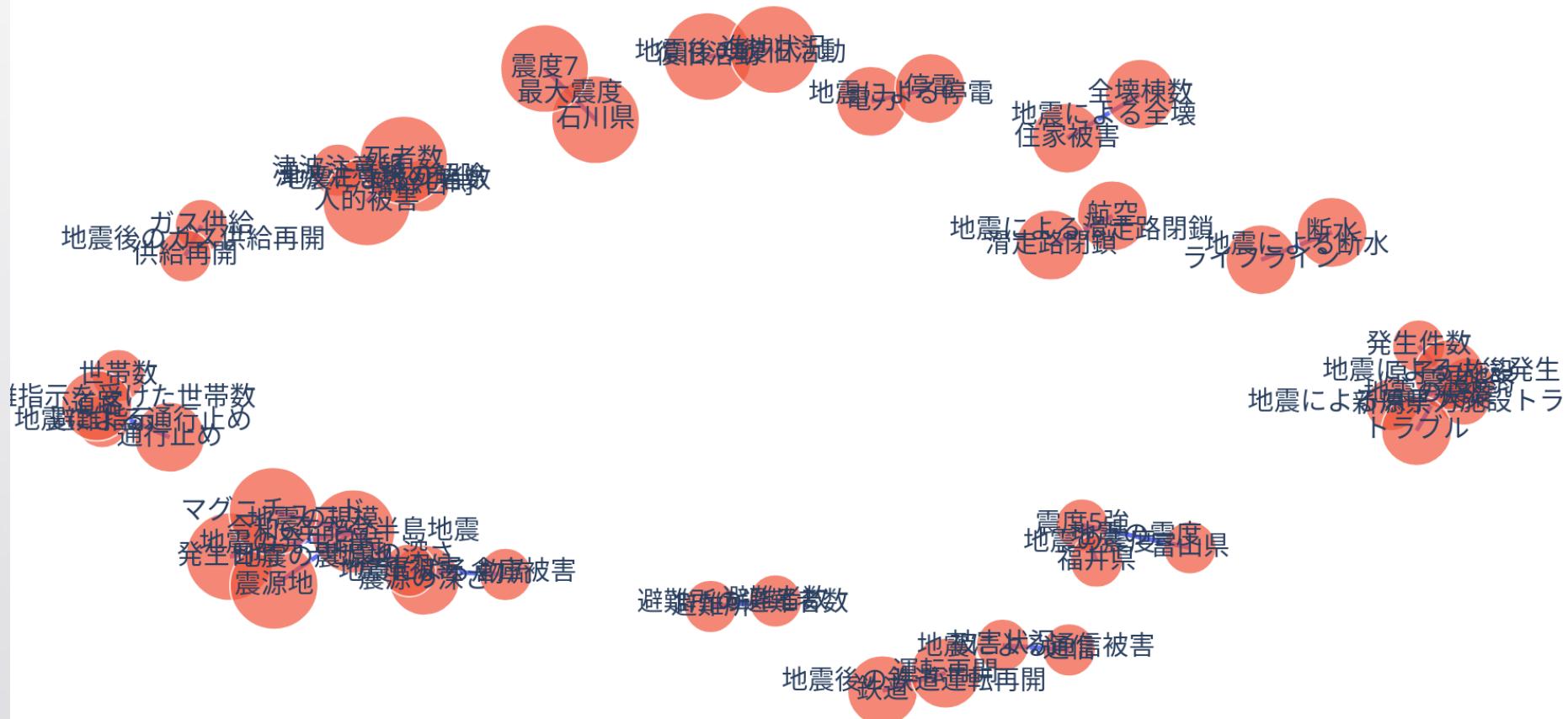
▶各災害データの Knowledge Graph の可視化

8月21日



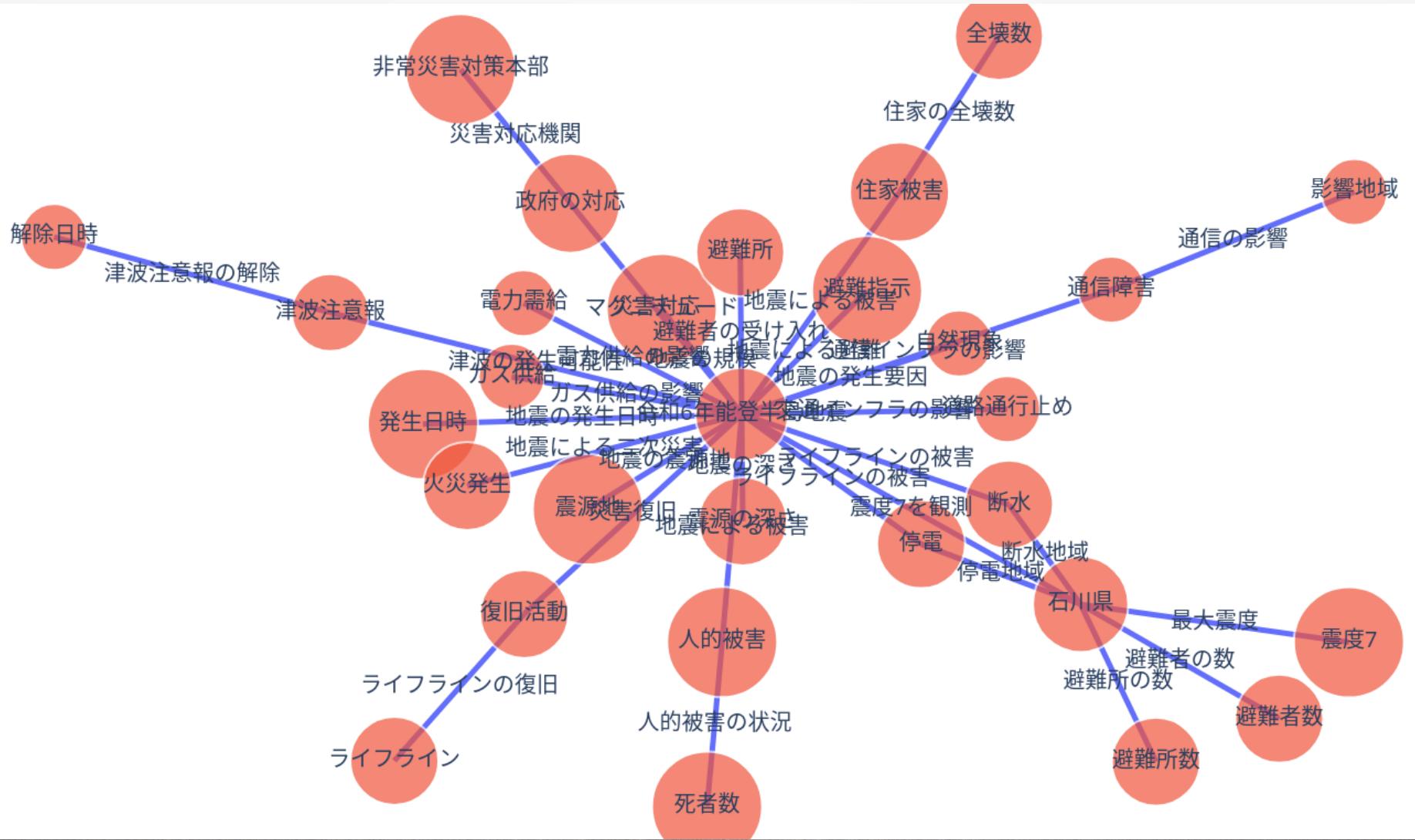
▶各災害データの Knowledge Graph の可視化

10月1日



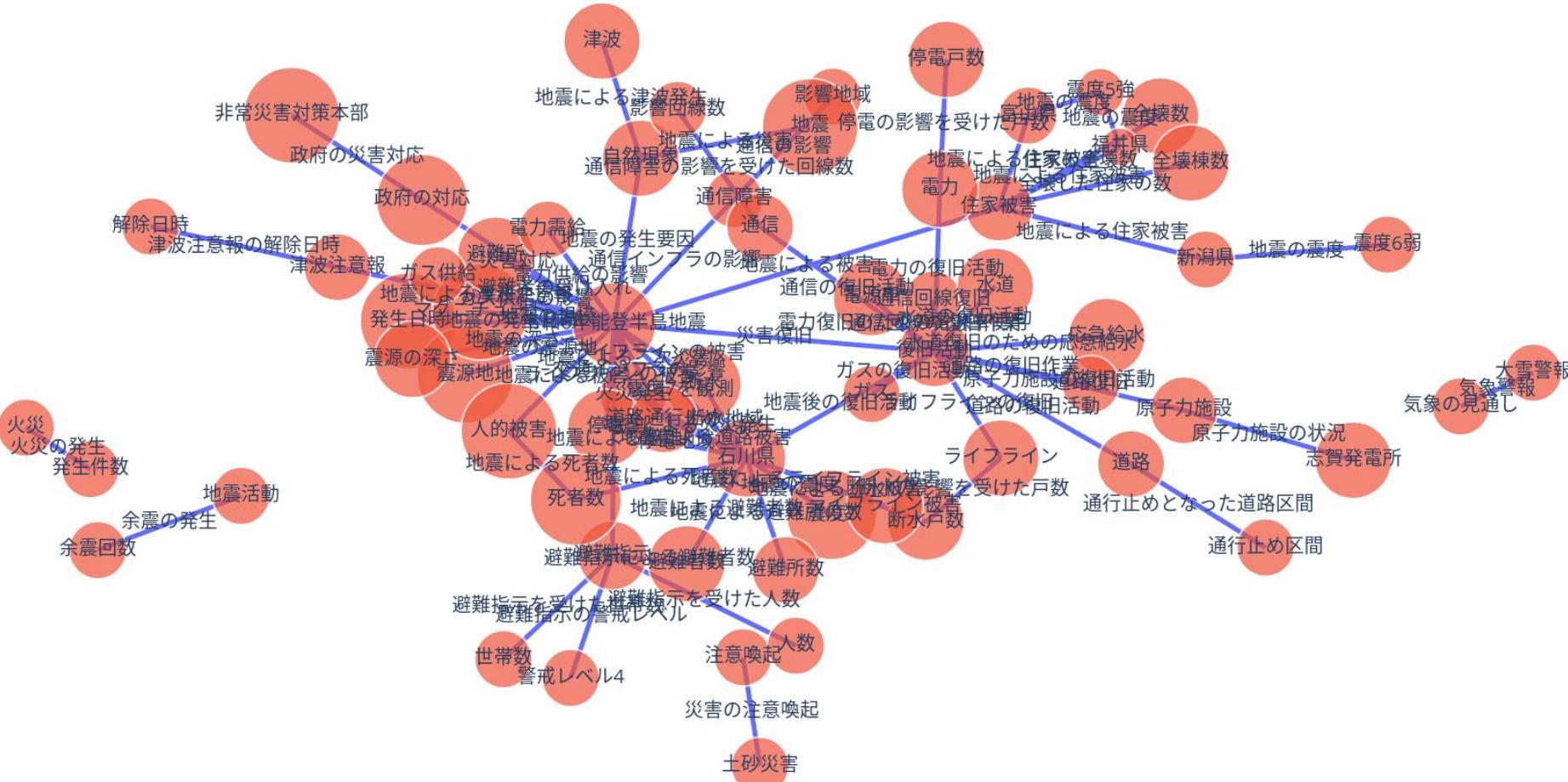
▶各災害データの Knowledge Graph の可視化

1月2日 - 1月8日

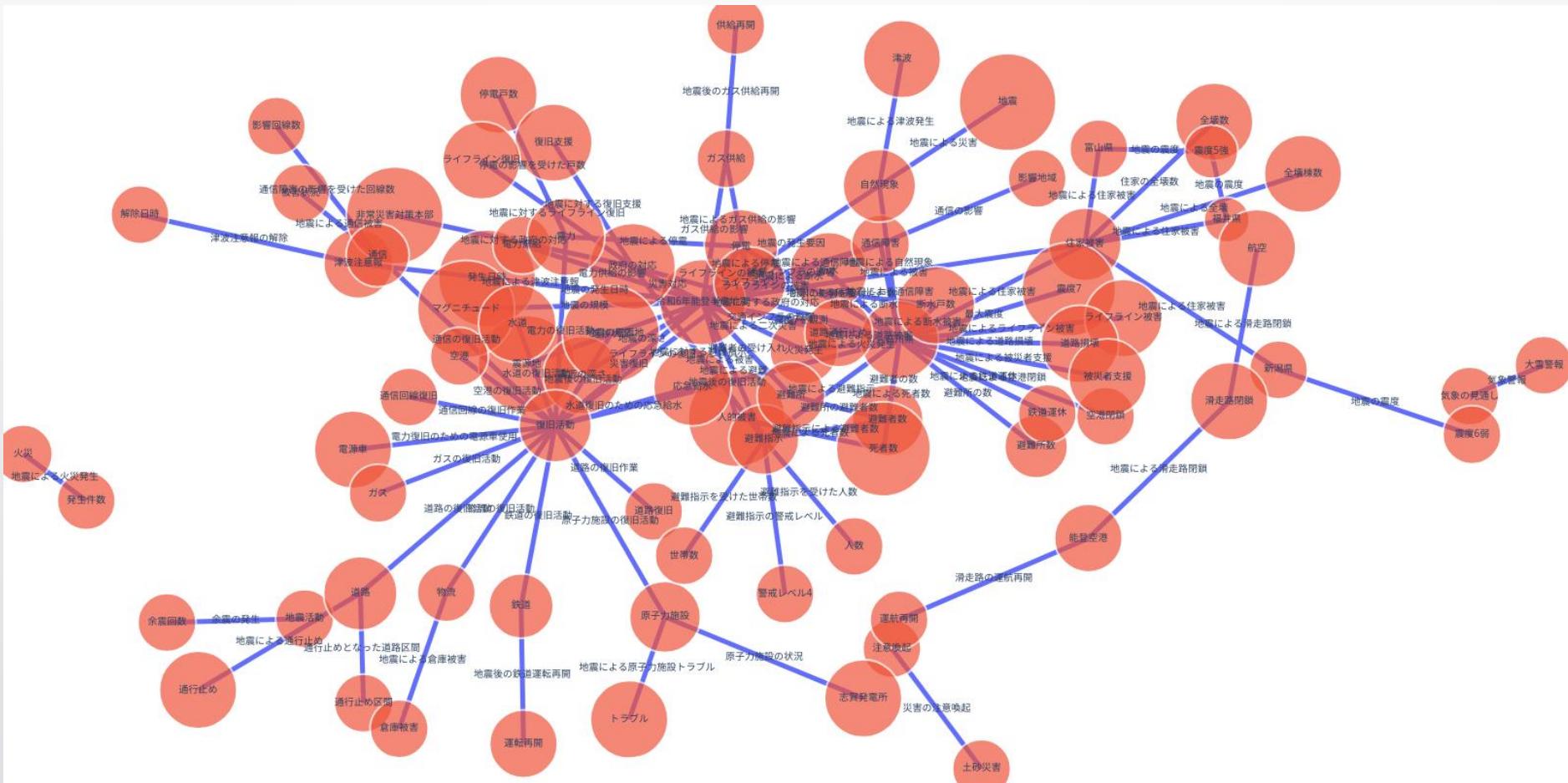


►各災害データの Knowledge Graph の可視化

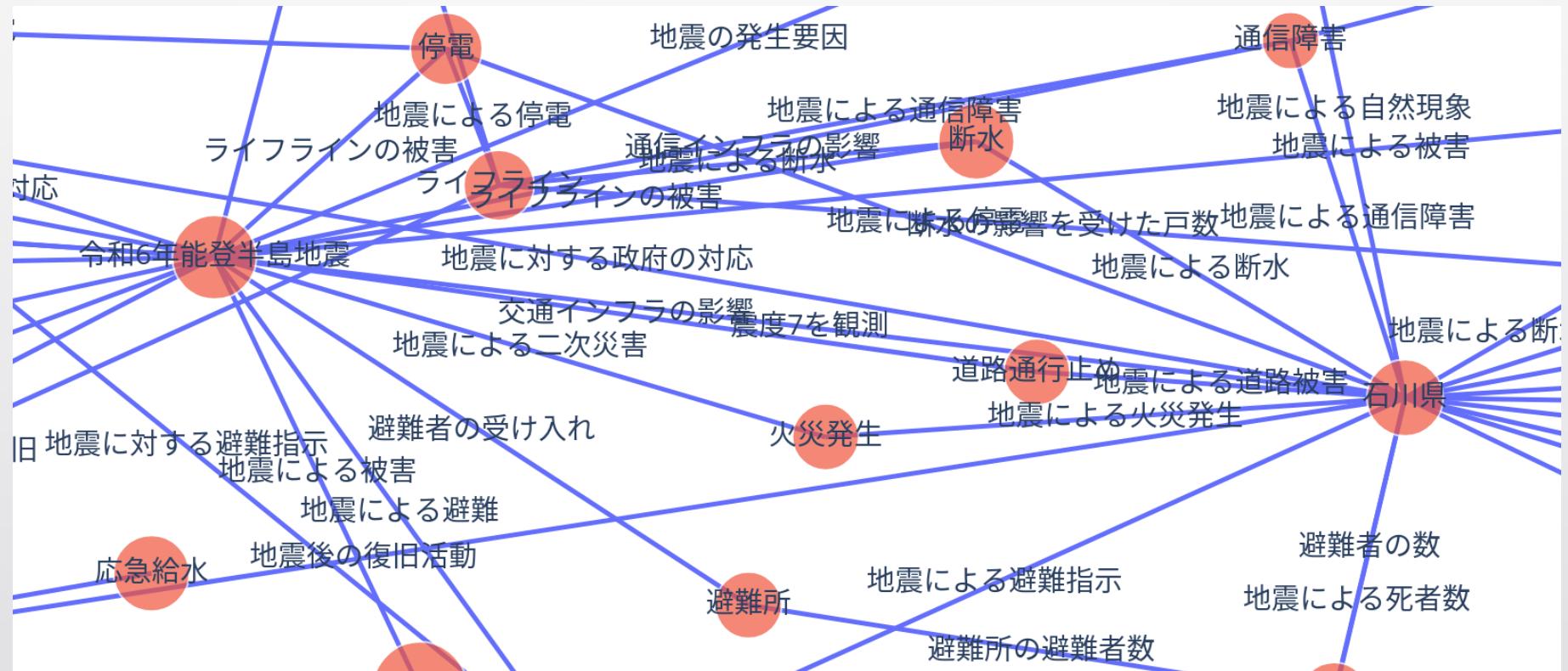
1月2日 - 4月2日



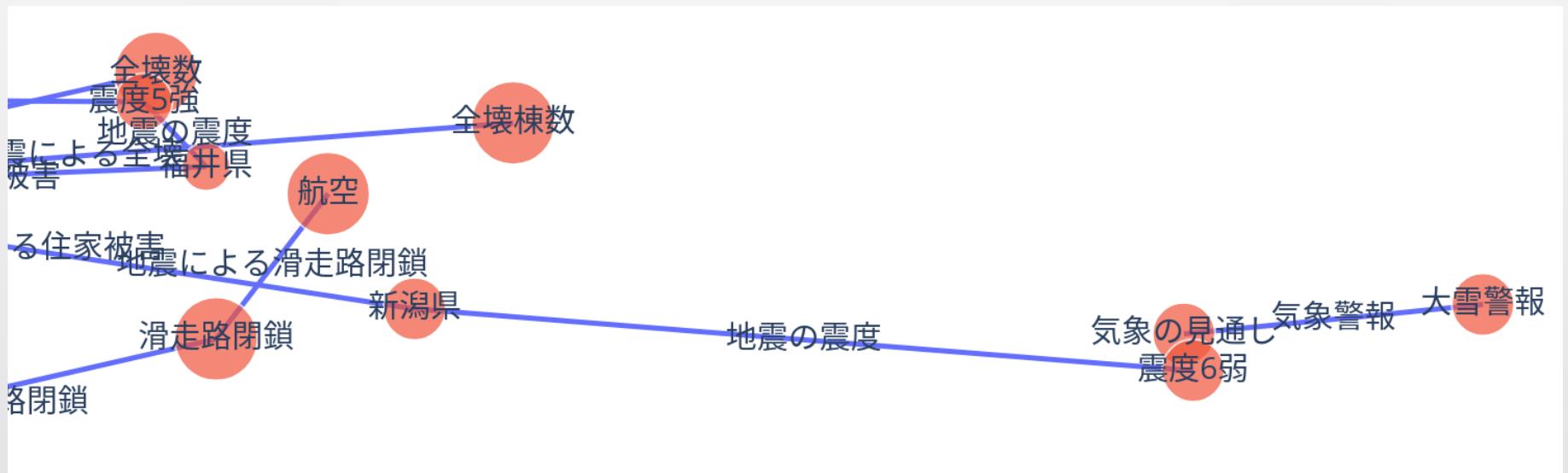
►統合した Knowledge Graph の可視化



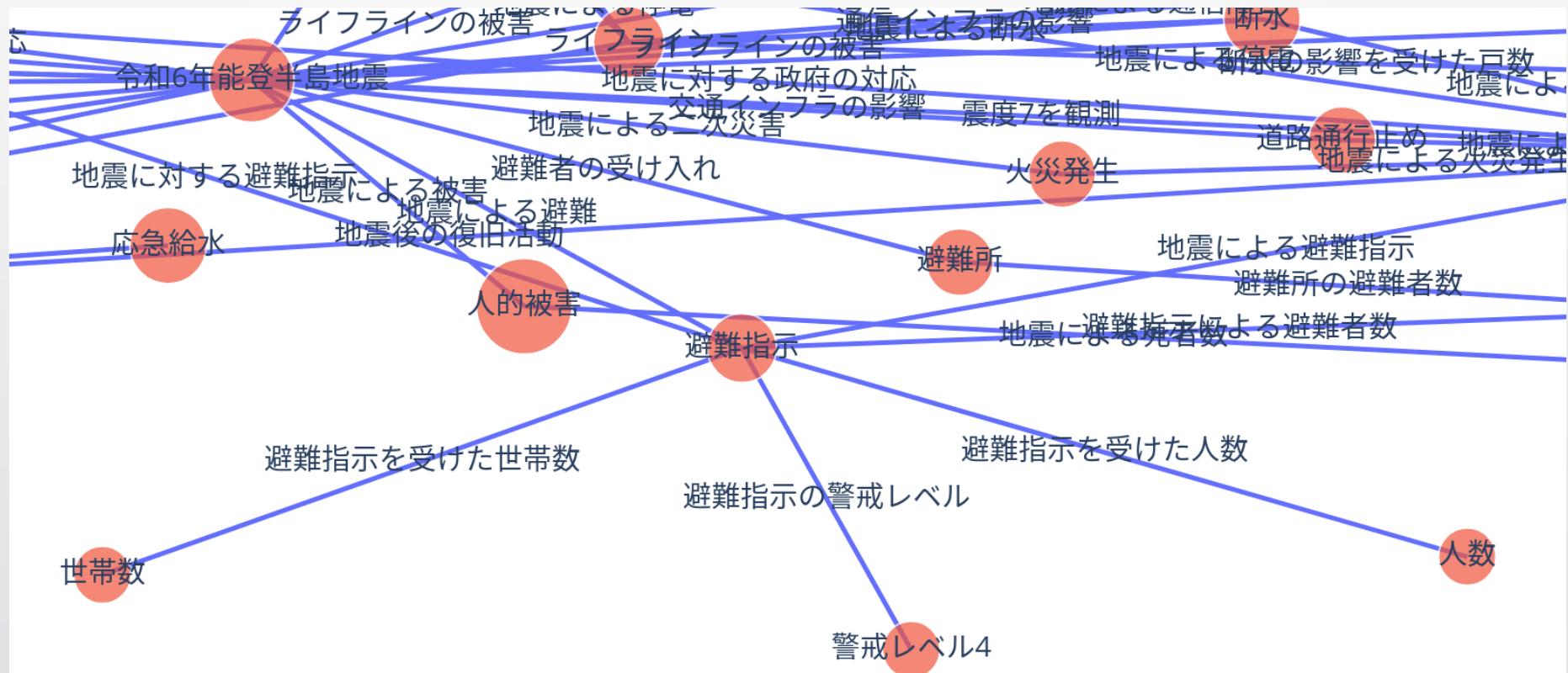
▶統合した Knowledge Graph の拡大図



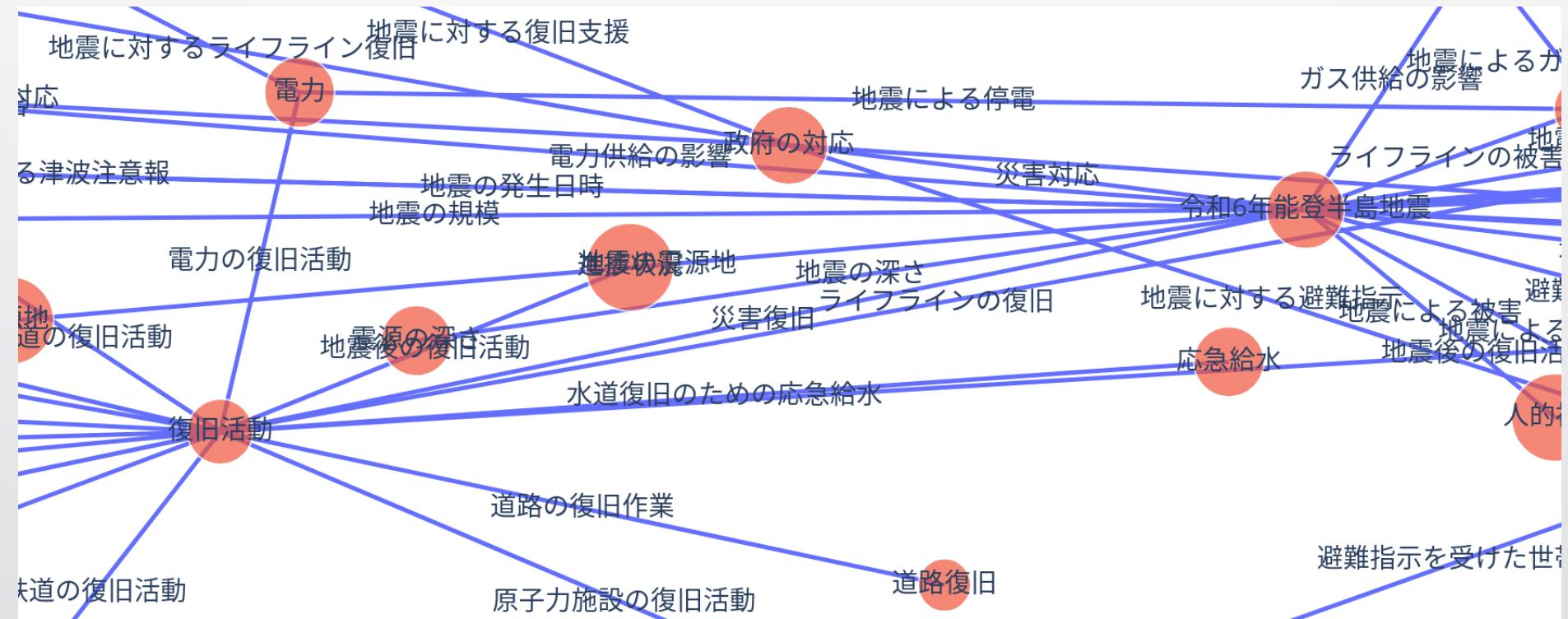
▶統合した Knowledge Graph の可視化



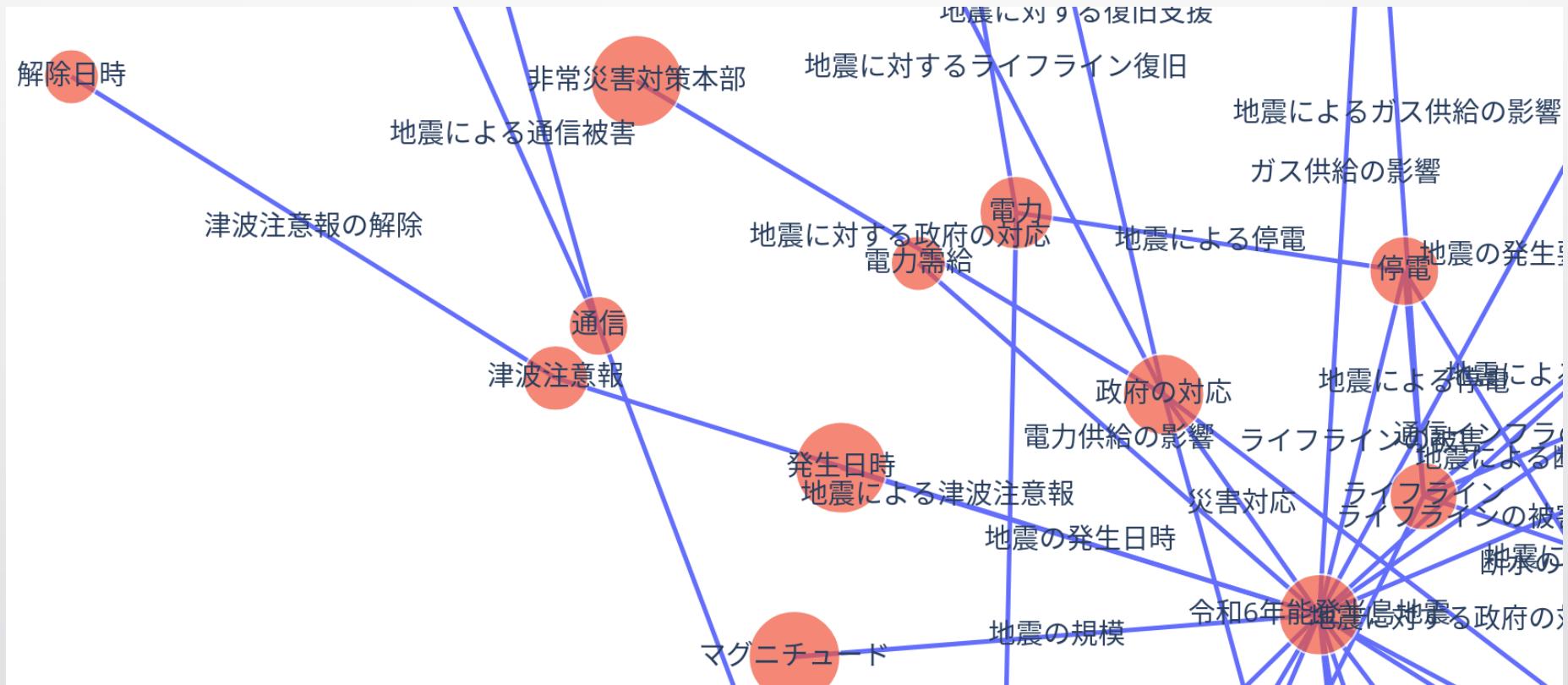
▶統合した Knowledge Graph の可視化



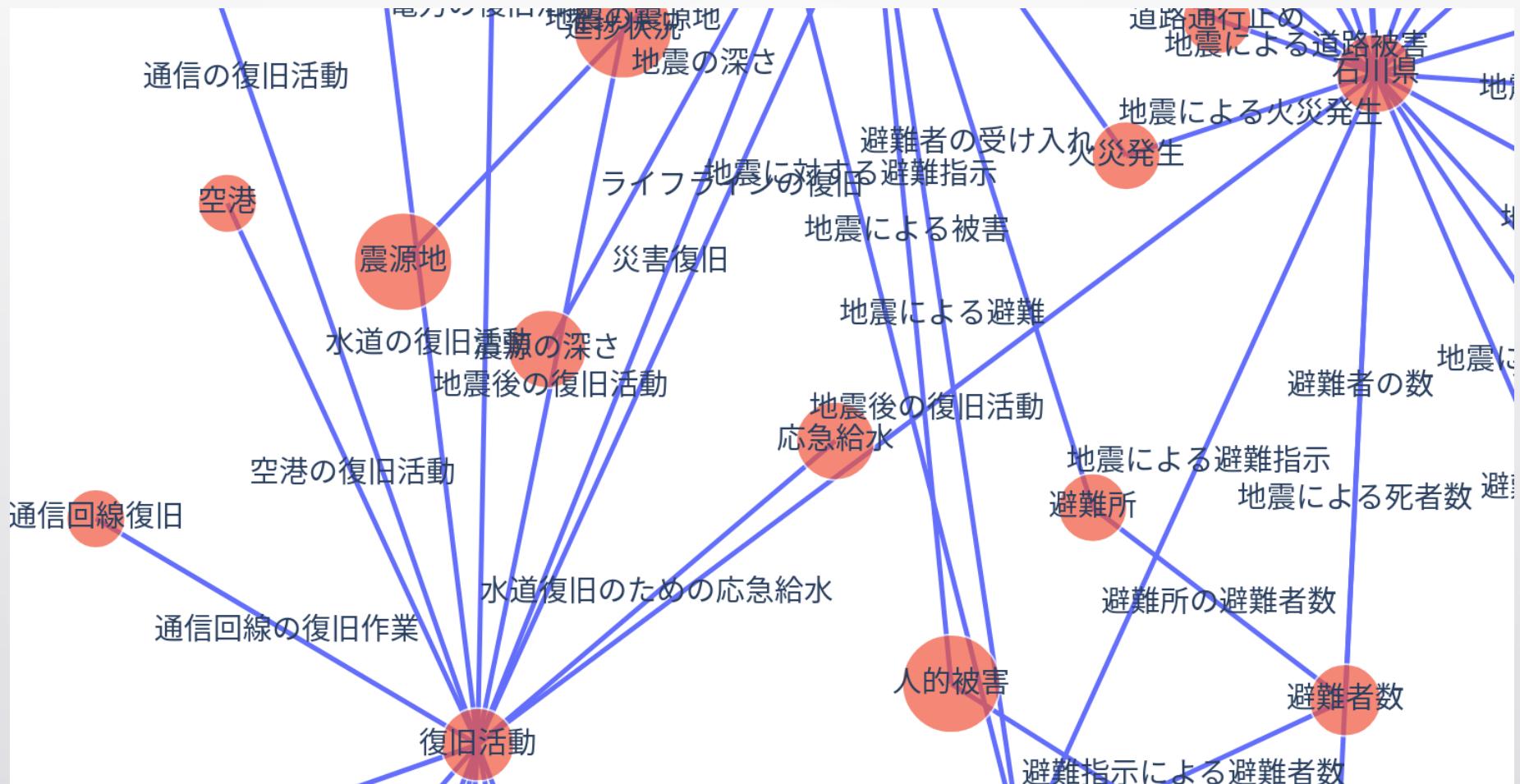
►統合した Knowledge Graph の可視化



▶統合した Knowledge Graph の可視化



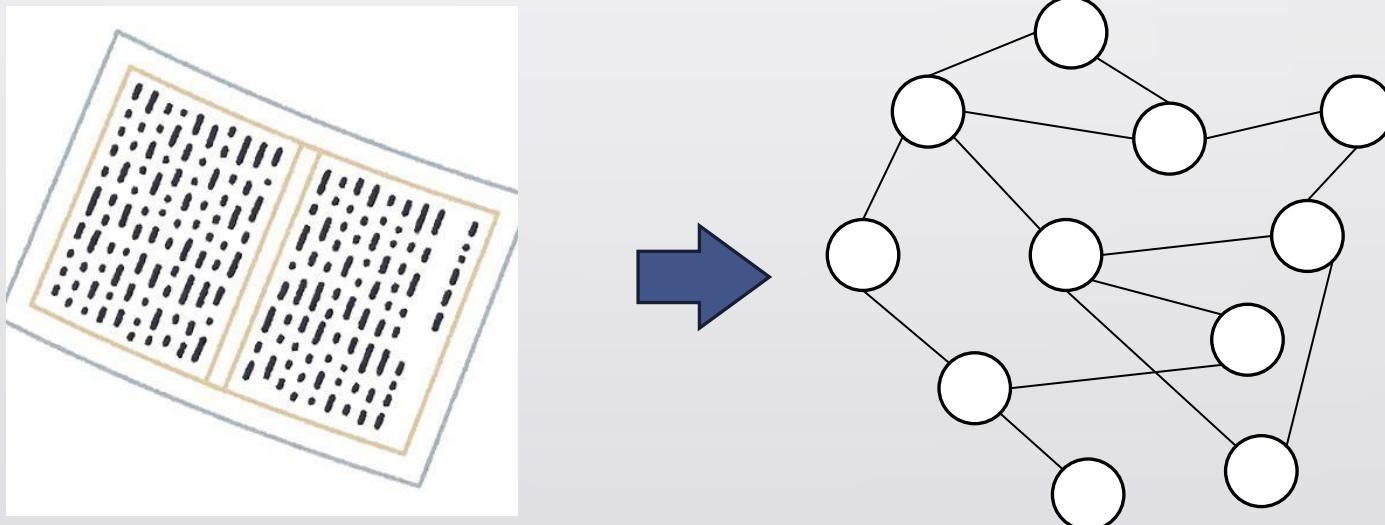
▶統合した Knowledge Graph の可視化



まとめと今後の課題

➤ 今後の課題

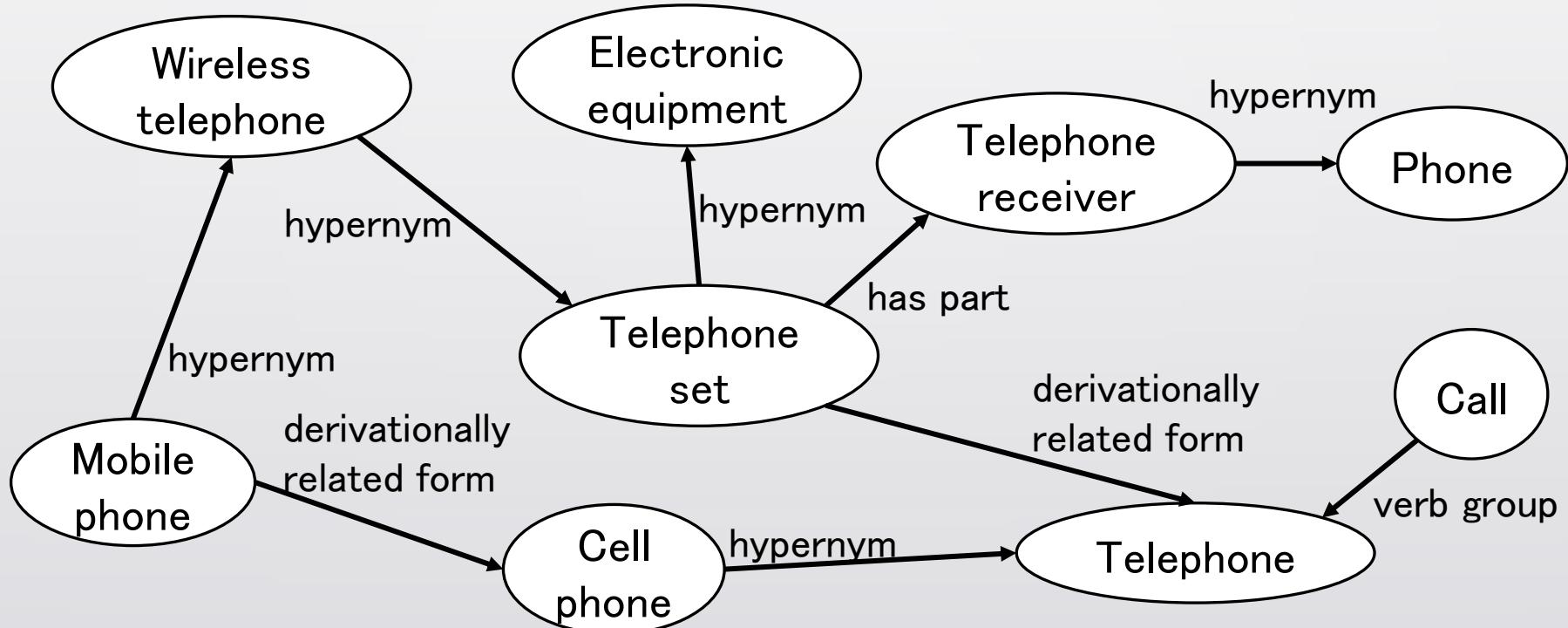
- ◆ 文章から triple の抽出
- ◆ Knowledge Graph の自動生成による視覚的な文章要約



要素技術

➤ Knowledge Graph

知識をグラフ構造として表した知識ネットワーク



結果（考察）

□ 今後の方針

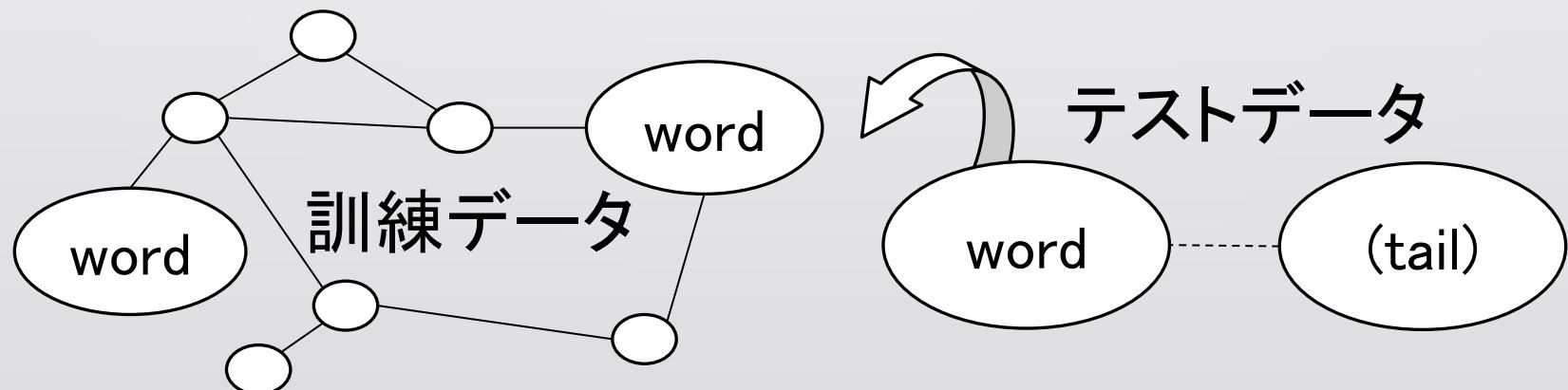
◆ Knowledge Graph 補完

… 存在しない entity に対して適切な triple を生成



しかし

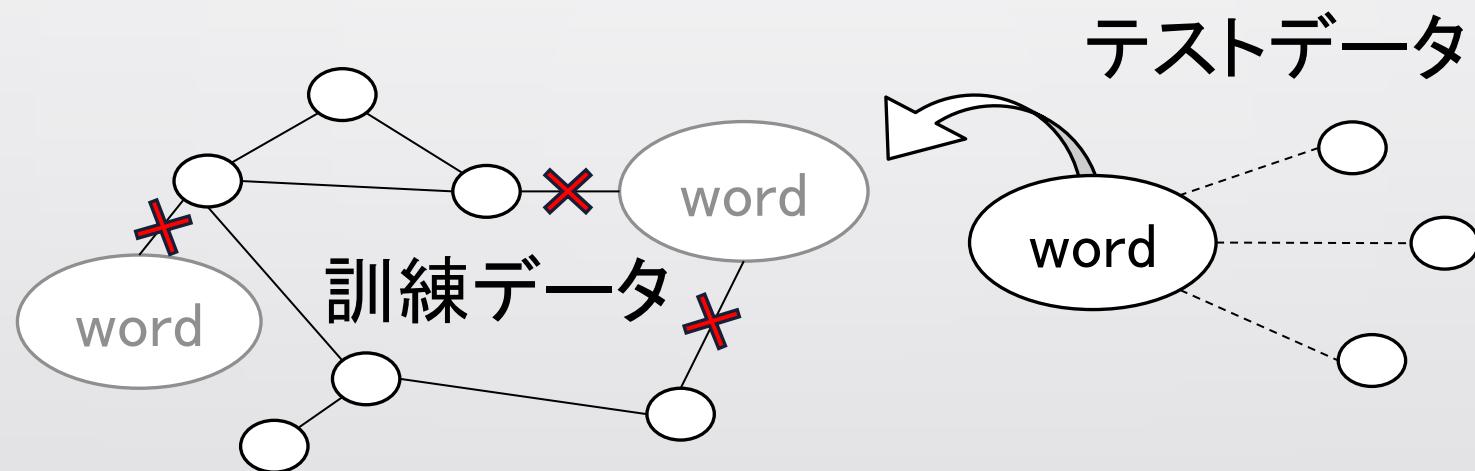
◆ テストデータ内の entity が訓練データ内に存在



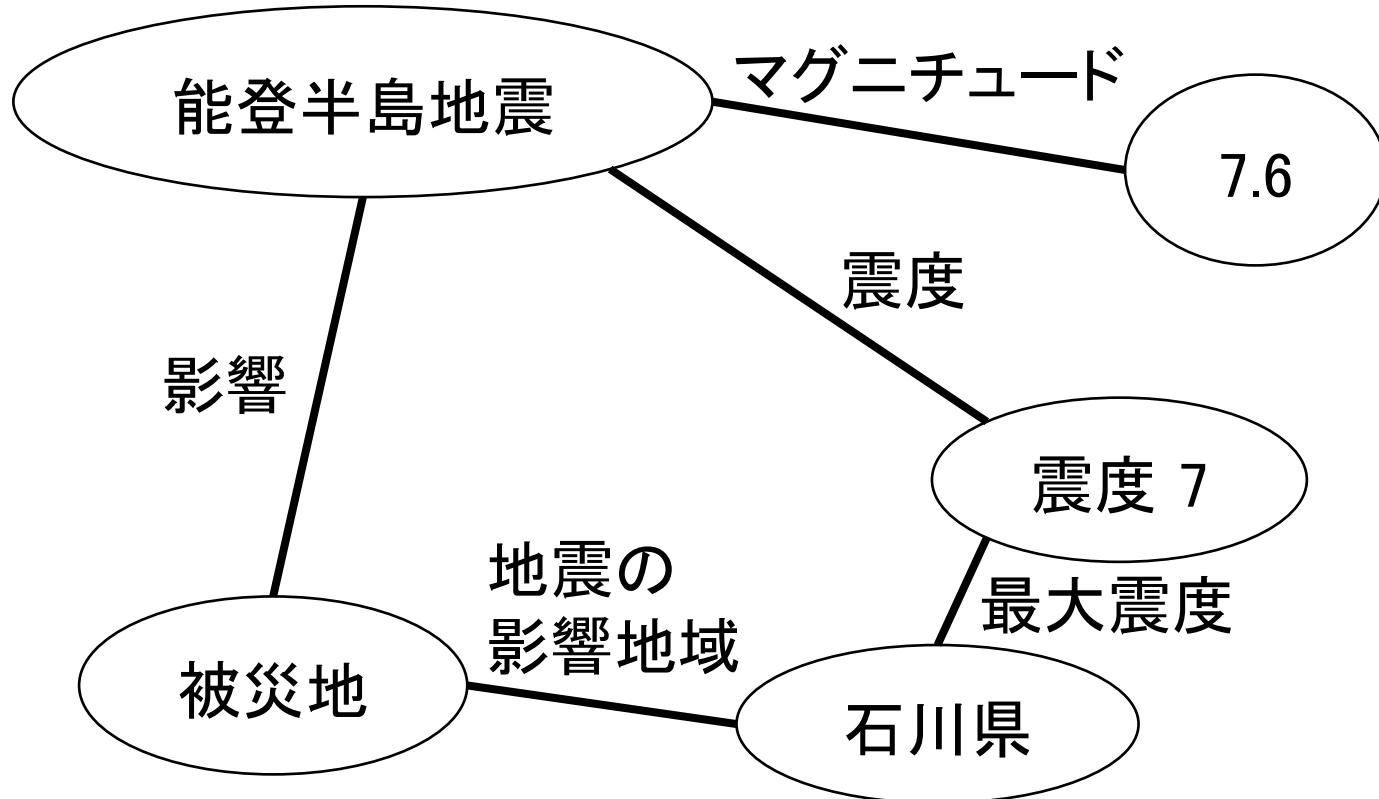
結果（考察）

□ 今後の方針

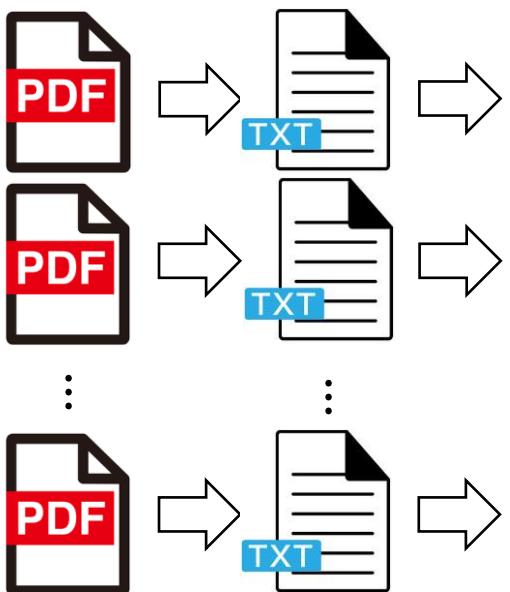
- ◆ ある entity を含む triple を訓練データより除外
- ◆ それらをすべてテストデータに追加



⇒ Knowledge Graph 補完の再現



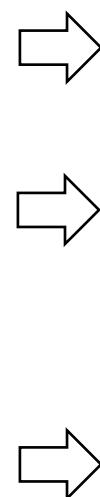
災害データ



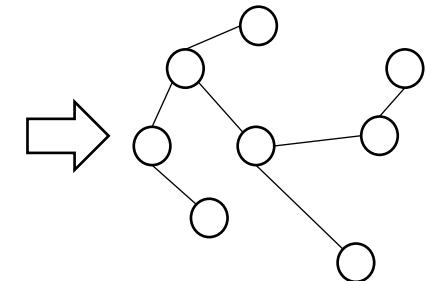
GPT-4o

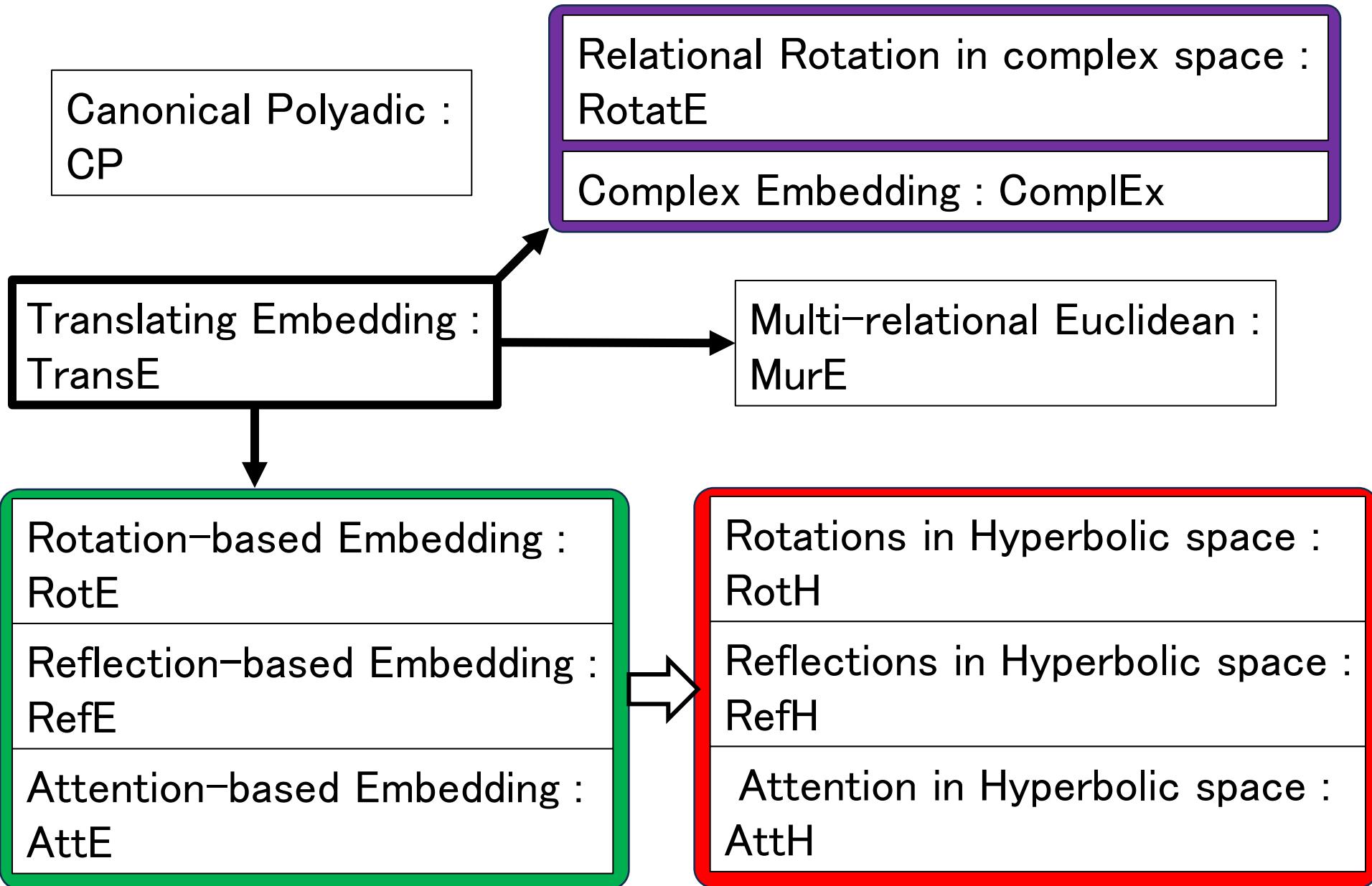
プロンプト

Knowledge Graph



triple





おわり