# 報告書

## 1 How to Convert Any Text Into a Graph of Concepts

テキストからのトリプル抽出の再現実験をした.
以下に “Ollama” への入力文とその和訳を示す.

### 1.1 LLM “Ollama” への入力文

Listing 1: “Ollama” への入力文

```
SYS_PROMPT = (
        "You are a network graph maker who extracts terms and their relations from a given context. "
        "You are provided with a context chunk (delimited by ‘‘‘) Your task is to extract the ontology "
        "of terms mentioned in the given context. These terms should represent the key concepts as per the context. \n"
        "Thought 1: While traversing through each sentence, Think about the key terms mentioned in it.\n"
            "\tTerms may include object, entity, location, organization, person, \n"
            "\tcondition, acronym, documents, service, concept, etc.\n"
            "\tTerms should be as atomistic as possible\n\n"
        "Thought 2: Think about how these terms can have one on one relation with other terms.\n"
            "\tTerms that are mentioned in the same sentence or the same paragraph are typically related to each other.\n"
            "\tTerms can be related to many other terms\n\n"
        "Thought 3: Find out the relation between each such related pair of terms. \n\n"
        "Format your output as a list of json. Each element of the list contains a pair of terms "
        "and the relation between them, like the follwing: \n"
        "[\n"
        " {\n"
        ’ "node_1": "A concept from extracted ontology",\n’
        ’ "node_2": "A related concept from extracted ontology",\n’
        ’ "edge": "relationship between the two concepts, node_1 and node_2 in one or two sentences"\n’
        " }, {...}\n"
        "]"
    )
```

あなたは, 指定された文脈から用語とその関係を抽出するネットワークグラフメーカーです. 以下のように提供された文脈チャンク (“で囲まれた部分) から, 言及されている用語のオントロジーを抽出するのがあなたの任務です. これらの用語は, 文脈に応じた主要な概念を表している必要があります.

考え 1：各文を読み進めながら, その中で言及されている重要な用語について考えてください.

用語には, オブジェクト, エンティティ, 場所, 組織, 人, 状態, 頭字語, 文書, サービス, 概念などが含まれることがあります. 用語はできるだけ原子レベルで細かくしてください.

考え 2：これらの用語が他の用語と一対一の関係を持つ可能性について考えてください. 同じ文や同じ段落で言及されている用語は, 通常お互いに関連しています. 用語は多くの他の用語と関連することができます.

考え 3：そのような関連する用語のペア間の関係を見つけてください.

出力を以下のようなリスト形式の JSON としてフォーマットしてください. リストの各要素は用語のペアとそれらの間の関係を含みます. 次のようになります：

node_1: 抽出されたオントロジーからの概念

node_2: 抽出されたオントロジーからの関連する概念

edge: node_1 と node_2 の関係、一文または二文で説明

## 1.2 対象となる文章の一部

以下にトリプルを取り出す文章の一例を示す.

Listing 2: 対象

```
Abstract'
Indias health indicators have improved in recent times but continue to lag behind those of its
    peer nations. The country with a population of 1.3 billion, has an estimated active health
    workers density of doctors and nurses/midwives of 5.0 and 6.0 respectively, for 10,000
    persons, which is much lower than the WHO threshold of 44.5 doctors, nurses, and midwives
    per 10,000 population. The issue is compounded by the skewed inter-state, urban-rural, and
    public-private sector divide. Calls to urgently augment the skilled health workforce
    reinforce the central role human resources have in healthcare, which has evolved into a
    complex multifactorial issue. The paucity of skilled personnel must be addressed if India is
     to accelerate its progress toward achieving universal health coverage and its sustainable
    development goals (SDGs).
The recent increase in the federal health budget offers an unprecedented opportunity to do this.
     This article utilizes the ready materials, extract and analyze data, distill findings (READ
    ) approach to adding to the authors' experiential learning to analyze the health system in
    India. The growing divide between the public and the burgeoning private health sector
    systems, with the latter's booming medical tourism industry and medical schools, are
    analyzed along with the newly minted National Medical Council, to recommend policies that
    would help India achieve its SDGs.

Categories: Public Health, Epidemiology/Public Health, Health Policy Keywords: working conditions
    , indian public health standards, auxiliary nurse midwives, human resource, health sector
    reform, india, health policy

Introduction And Background'
Indias health indicators have improved in recent times but continue to lag behind those of its
    peer nations. The country has an estimated active health '
    workers density much lower than the WHO recommended thresholds [1]. The issue is compounded
    by the skewed inter-state, urban-rural, and public-private sector divide. The paucity of
    skilled personnel is a multi-factorial issue and needs to be addressed if India is to
    accelerate its progress toward achieving universal health coverage and its sustainable
    development goals (SDGs).
```

Listing 3: triples

```
node_1 node_2 edge chunk_id count
0 india's health indicators peer nations continue to lag behind ae0fd26675d645e787964255667e90f4
    4
2 health workers density doctors and nurses/midwives for 10,00 persons
    ae0fd26675d645e787964255667e90f4 4
4 skilled health workforce india reinforces the central role human resources ha...
    ae0fd26675d645e787964255667e90f4 4
5 skewed inter-state urban-rural and public-private sector divide
    ae0fd26675d645e787964255667e90f4 4
7 health budget federal offers an unprecedented opportunity to do this
    ae0fd26675d645e787964255667e90f4 4
```

## 2 Text to KG の評価手法の調査

- WebNLG データセット
- TEKGEN データセット

いずれも文とその文から取れるトリプルの組. F1 値, Precision, Recall を用いて評価.

ホームズデータセットも文とトリプルの組のため, 同じ評価手法が可能である. しかし, 研究の目的が “物語の内容理解のためのナレッジグラフ生成” のため, データセット内のトリプルとの誤差をみると目的との相違が生じる. 他の評価手法を調査, 検討する.

## 3 今後

- Text to KG の評価手法の調査
- ホームズデータセットへの適用

## 参考文献

[1] Rahul Nayak, How to Convert Any Text Into a Graph of Concepts, https://medium.com/towards-data-science/how-to-convert-any-text-into-a-graph-of-concepts-110844f22a1a