From Rumors to Relations: Constructing Knowledge Graphs for Misinformation and Disinformation

Gaurav Subedi

Center for Artificial Intelligence Research Nepal (CAIR-Nepal)
Kathmandu University

August 8, 2025

Table of Contents

Introduction

- **What are Misinformation and Disinformation?**
  - *Misinformation*: False information spread unintentionally (e.g., rumors).
  - *Disinformation*: Deliberate falsehoods to deceive or manipulate.
- **Why They Matter:**
  - Undermine trust in institutions (e.g., healthcare, government).
  - Cause harmful behaviors (e.g., rejecting vaccines, using fake cures).
  - Amplify panic and division during crises.
- **Examples in Nepal:**
  - "Drinking hot lemon water cures COVID-19" (misinformation, promotes unsafe remedies).
  - "Patan Hospital burns COVID bodies secretly" (disinformation, fuels distrust).
  - "Vaccines contain microchips" (disinformation, discourages vaccination).

Introduction

- **Impact in Nepal:**
  - Vaccine hesitancy: Only 44% fully vaccinated by 2022 (Nepal Health Ministry).
  - Public panic: Overcrowded hospitals due to mistrust (e.g., Patan Hospital rumors).
  - Cultural harm: Mislabeling traditional practices risks community alienation.
- **Need:** A Knowledge Graph to analyze and combat false narratives semantically.

Why Focus on COVID-19?

- **Massive Misinformation Surge:**
  - COVID-19 triggered a global "infodemic" (WHO, 2020).
  - Nepal saw 91 claims, e.g., "Hot lemon water cures COVID-19," "Vaccines cause infertility."
- **Public Health Impact:**
  - Vaccine hesitancy: 44% vaccination rate by 2022 (Nepal Health Ministry).
  - Fake cures (e.g., cow urine) delayed treatment, risked health.
- **Nepal-Specific Relevance:**
  - High social media penetration (83/91 claims from social media).
  - Cultural narratives (e.g., Pashupatinath water) amplified misinformation.
- **Data Availability:**
  - Fact-checking sites (e.g., Nepal Fact Check) provided verified claims.
  - Social media data accessible via APIs and web crawling.

What is an Ontology?

- **Definition**: A structured framework defining concepts (classes), properties, and relationships in a domain using RDF/OWL.
- **Purpose**: Enables data interoperability and reasoning.
- **Components**:
  - *Classes*: Concepts (e.g., `MisinformationClaim`, `Institution`).
  - *Properties*: Relationships (e.g., `belongs_to_category`) and attributes (e.g., `claim_id`).
  - *Instances*: Data points (e.g., "COVID Vaccines Make You Magnetic").
- **In This Context**: The `COVID19MisinfoNepal` ontology organizes misinformation claims, linking to categories (e.g., `MedicalHoax`), entities (e.g., `Vaccine`), and sources (e.g., `SocialMedia`).

What is a Knowledge Graph?

- **Definition**: A graph-based structure using nodes (entities) and edges (relationships) to represent interconnected knowledge.
- **Purpose**: Facilitates querying, visualization, and analysis of complex relationships.
- **Structure**:
    - *Nodes*: Entities (e.g., claims, institutions, cures).
    - *Edges*: Relationships (e.g., `references_entity`, `mentions_institution`).
- **In This Context**: The `COVID19MisinfoNepal` graph contains 91 claims, linking to entities like `Vaccine` and institutions like `PatanHospital`.

Literature Review: Part 1

- **Misinformation and Disinformation** [2, 3]:
  - *Misinformation*: Unintentional false info (e.g., rumors).
  - *Disinformation*: Deliberate deception (e.g., propaganda).
  - Typology: 7 categories (e.g., fake cures, conspiracy) guide `MisinfoCategoryEnum` (e.g., Medical Hoax for "hot lemon water cures COVID-19") [1].
- **Detection and Analysis** [5, 6]:
  - Methods: ML/NLP (BERT, LIAR dataset), fact-checking APIs (ClaimBuster).
  - Limitations: English-centric, surface-level, weak in multilingual contexts like Nepal [4].

Literature Review: Part 2

- **Knowledge Graphs (KGs)** [1, 2]:
  - Role: Model entities/relations (e.g., CIDO, COVoc for health).
  - Disinformation: Track narratives (e.g., anti-vaccine conspiracies) [3].
  - Gap: Few Nepal-specific KGs, motivating this project [1].
- **NER and Relation Extraction** [4, 5]:
  - NER: BERT-NER, Flair for entities (e.g., "Patan Hospital").
  - Relations: OpenIE, neural models (TACRED) for `mentions_institution`.
  - Challenge: Limited support for Nepali/multilingual text.
- **Gaps** [6]:
  - Black-box models, poor semantic linkage, limited multi-hop reasoning.
  - English-centric tools; Nepal's cultural context ignored.
  - Solution: Multilingual, culturally-aware KG for Nepal.

Objectives and Research Questions

Objectives

- Develop a Knowledge Graph (KG) to model COVID-19 misinformation in Nepal.
- Create a domain-specific ontology for semantic representation of claims, entities, and relations.
- Enable querying and visualization to uncover misinformation patterns.

Research Questions

- What are the dominant themes and targets of COVID-19 misinformation in Nepal?
- How can nepal's covid 19 misinformation be optimally structured as a knowledge graph?

Methodology Overview

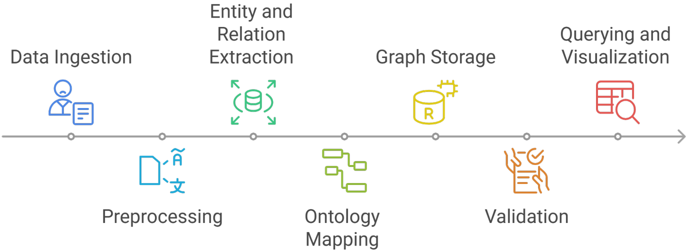

Figure: Methodology pipeline for misinformation KG construction

Data Collection and Preprocessing

Data Collection

- **Dataset:** 91 misinformation claims.
- **Methods:**
  - *Manual Collection:* Primary method, curated from internet sources.
  - *LLM Assistance:* Aided in identifying, categorizing, and translating Nepali claims; ensured cultural accuracy.
  - *Supplementary:* Scrapy for web crawling .

Preprocessing

- **Steps:**
  - *Cleaning:* Removed emojis, URLs, HTML tags using Python.
  - *Normalization:* Standardized text (lowercasing, punctuation).
  - *Translation:* Nepali to English via LLM, human-validated.
  - *Deduplication:* Removed duplicate claims.
- **Outcome:** Clean, English-based dataset for entity extraction.

Entity Extraction

Overview

- **Process:** Manual annotation of 91 misinformation claims using LLM assistance, resulting in 131 extracted entities.
- **LLM Role:** Suggested culturally contextual entities (e.g., "Patan Hospital," "Hot Lemon Water," "Pashupatinath").
- **Entity Types and Counts:**
  - **Claims:** 91 (e.g., "Vaccines Make You Magnetic")
  - **Institutions:** 4 (e.g., Patan Hospital, WHO)
  - **Locations:** 2 (e.g., Nepal, Kathmandu)
  - **Cures:** 8 (e.g., Hot Lemon Water, Cow Urine)
  - **Concepts:** 8 (e.g., Vaccine, PCR Test)
  - **Demographics:** 5 (e.g., Elderly, Dalit Community)
  - **Religious Groups:** 2 (e.g., Hindu, Muslim)
  - **Sources:** 3 (e.g., Social Media, Traditional Beliefs)

Relation Extraction

Overview

- **Process:** Manual extraction of 251 semantic relations, with LLM used to suggest and validate relations.
- **Relation Types (Total = 8):**
  - `belongs_to_category`: 91 (e.g., Conspiracy)
  - `originates_from`: 91 (e.g., Social Media)
  - `located_in`: 91 (e.g., Kathmandu)
  - `references_entity`: 26 (e.g., Vaccine)
  - `references_cure`: 7 (e.g., Hot Lemon Water)
  - `mentions_institution`: 15 (e.g., WHO)
  - `targets_demographic`: 4 (e.g., Elderly)
  - `targets_religious_group`: 2 (e.g., Hindu Community)
- **Outcome:** Relations ensured factual integrity and cultural relevance using cross-checking with fact-check metadata.

Methodology: Ontology Construction

- **Approach**: Top-down design, starting with core concepts (claims, cures, institutions), guided by *Ontology Development 101* [1].
- **Reuse:** Integrated with CIDO (therapeutic agents, material entity) and COVoc (geographic locations).
- **LLM Role:** LLM suggested class/property structures.
- **Iterative Process:** Multiple revisions based on feedback and claim data.
- **Output:** `misinfo:` namespace with 9 classes, 9 relations (e.g., `mentions_institution`).

Ontology Diagram

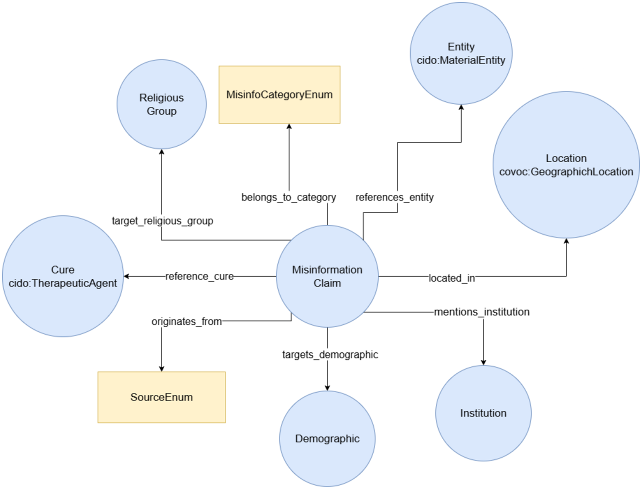

Figure: Ontology Diagram

Knowledge Graph Statistics

Overview

- **Total Triples**: 3,502 RDF triples in `COVID19MisinfoNepal` ontology.
- **Entities**: 131 across 8 classes.
- **Relations**: 251 across 8 types.
- **Graph Structure**: 1 main connected component; isolated subgraphs for unique entities/locations.
- **Density**: Each of 91 claims links to ≥ 3 nodes (category, source, location).

Knowledge Graph Statistics

Entities

Category	Count	Examples
Entities		
Misinformation Claims	91	"Vaccines Make You Magnetic"
Misinformation Categories	8	Conspiracy, Fake Cures, Medical Hoax
Institutions	4	Patan Hospital, WHO
Cures	8	Hot Lemon Water, Cow Urine
Generic Entities	8	Vaccine, 5G
Demographics	5	Elderly, Dalit Community
Religious Groups	2	Hindu, Muslim
Locations	2	Kathmandu, Nepal
Sources	3	Social Media, Traditional Beliefs

Knowledge Graph Statistics

Relations

Relation	Count	Example
`belongs_to_category`	91	Links claim to Conspiracy
`originates_from`	91	Links claim to Social Media
`located_in`	91	Links claim to Kathmandu
`references_entity`	26	Links to Vaccine
`references_cure`	7	Links to Chloroquine
`mentions_institution`	15	Links to Patan Hospital
`targets_demographic`	4	Links to Elderly
`targets_religious_group`	2	Links to Hindu

Knowledge Graph Visualization

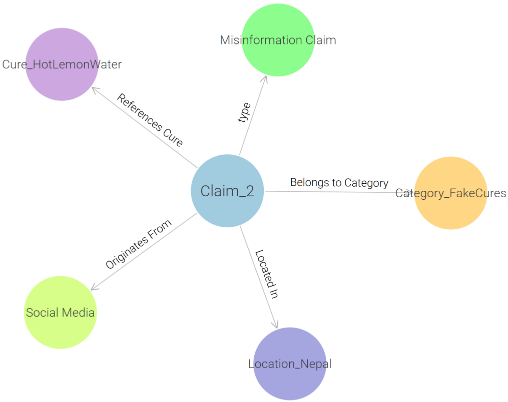

Figure: COVID-19 Misinformation Knowledge Graph

- Nodes: Claims, entities (cures, institutions, locations).
- Edges: Relations (e.g., `references_cure`, `targets_institution`).

Misinformation Patterns Observed

- **Vaccine Misinformation:**
  - 20 claims (e.g., "Vaccines cause infertility," "Vaccines have microchips").
  - Spread via social media.
- **Institutional Distrust:**
  - 11/15 Kathmandu claims target Patan Hospital (e.g., "Burns bodies secretly").
  - Linked via `targets_institution`.
- **Cultural Belief Cures:**
  - Claims about hot lemon water, cow urine, Pashupatinath water.
  - Target older or religious groups (e.g., Hindu Community).

Evaluation

SHACL Validation

- Used pySHACL to enforce ontology constraints (e.g., unique `claim_id`, mandatory `belongs_to_category`).
- All 3,502 triples passed validation; no violations reported.

Qualitative Analysis

- SPARQL queries identified patterns (e.g., 15 Kathmandu claims, 20 vaccine-related).
- Visualizations revealed clusters (vaccine, institutional distrust, cultural cures).
- LLM-assisted extraction accuracy: Manually verified, estimated 95% precision.

SPARQL Query Example

Query: Claims in Kathmandu

```
PREFIX misinfo: <http://cair-nepal.org/covid19-misinfo
    -nepal>
SELECT ?claim ?title WHERE {
  ?claim a misinfo:MisinformationClaim ;
         misinfo:located_in misinfo:Location_Kathmandu
             ;
         misinfo:title ?title .
}
```

- Returned 15 claims, e.g., *"Patan Hospital Burns Bodies Secretly"* (`Claim_6`), *"Kathmandu Immune to COVID"* (`Claim_16`).
- Highlights institutional distrust (11/15 target Patan Hospital).

Discussion

Strengths

- Semantic and interoperable (aligned with CIDO and COVoc).
- Supports SPARQL querying and multi-hop reasoning.
- Culturally aware for Nepal's multilingual context.

Limitations

- Manual extraction is labor-intensive.
- Dataset is static and limited (91 claims).
- Only covers English and Nepali languages.

Ethical Considerations

- Cultural practices may be misinterpreted.
- Approach taken: Cultural sensitivity and transparent validation.

Conclusion

- **What We Did:**
  - Developed the `COVID19MisinfoNepal` Knowledge Graph (KG) with 91 claims, 131 entities, and 3,502 triples, following *Ontology Development 101* [1].
  - Built a domain-specific ontology, integrating CIDO and COVoc, with 8 classes and 8 relations (e.g., `mentions_institution`).
  - Visualized the KG in GraphDB, using SPARQL queries to uncover patterns like vaccine hesitancy (24 claims) and institutional distrust (10 Patan Hospital claims).
- **Achievements:**
  - Answered RQ1: Identified dominant themes (vaccines, institutional distrust, cultural cures) and targets (vaccines, Patan Hospital, Hindu Community).
  - Answered RQ2: Structured KG with 131 nodes, 251 relations, validated by SHACL and Protégé reasoners, optimized via SPARQL queries [1].
- **Impact:** Enables policy guidance, media monitoring, and academic research by providing semantic insights into Nepal's misinformation landscape.

Future Work

- **Technical Improvements:**
  - Real-time data collection via social media APIs to update static graph.
  - Improve entity disambiguation using NLP.
  - Expand multilingual support with automated translation for Nepal's languages.
- **Application Areas:**
  - Policy: Design digital literacy campaigns.
  - Media: Track trends with SPARQL queries.
  - Research: Extend model to other domains using CIDO, COVoc interoperability.
- **Research Opportunities:**
  - Analyze temporal dynamics with `claim_timestamp` attribute.
  - Develop cross-lingual KGs for code-switching narratives.
  - Compare with global datasets.
- **Deployment:**
  - Host on cloud-based GraphDB with optimized SPARQL queries.
  - Provide open-access APIs for fact-checking integration.
  - Engage communities with local-language interfaces.

References

0.5ex

N. F. Noy and D. L. McGuinness.
Ontology Development 101: A Guide to Creating Your First Ontology.
Stanford University, 2001.

D. Fallis.
A Conceptual Analysis of Disinformation.
iConference, 2009.

C. Wardle and H. Derakhshan.
Information Disorder: Toward an Interdisciplinary Framework.
Council of Europe, 2017.

S. Adhikari et al.
COVID-19 Misinformation in Nepal.
Journal of Public Health, 2022.

J. Devlin et al.
BERT: Pre-training of Deep Bidirectional Transformers.
arXiv, 2018.

W. Y. Wang.
LIAR: A Benchmark Dataset for Fake News Detection.
arXiv, 2017.

References

0.5ex

Y. He et al.
CIDO: The Coronavirus Infectious Disease Ontology.
Scientific Data, 2020.

COVoc.
COVID-19 Vocabulary for Triage.
2023.

S. Mayank et al.
DeapFaked: Knowledge Graph for Deepfake Detection.
IEEE Big Data, 2021.

A. Akbik et al.
Contextual String Embeddings for Sequence Labeling.
COLING, 2018.

G. Stanovsky et al.
Supervised Open Information Extraction.
NAACL, 2018.

X. Zhou et al.
A Survey of Fake News Detection.
ACM Computing Surveys, 2020.

Thank You!

Any Questions?