An Innovative Approach to Represent Tacit Knowledge of Fishing with Knowledge Graphs

Tanjila Kanij[1][a], Shafia Husna [1], Afzal Azeem Chowdhary [1][b], Misita Anwar[1][c], Md. Khalid Hossain[2][d] and John Grundy[2][e]

[1]*Swinburne University of Technology, Melbourne, Australia*

[2]*Monash University, Melbourne, Australia*

103807822@student.swin.edu.au, {misitiaanwar, achowdhary, tkanij}@swin.edu.au, {md.khalid.hossain, jgrundy}@monash.edu.au

Keywords: Tacit Knowledge, Knowledge Graph, Fisherfolk, Large language Models

Abstract: Fisherfolk communities in developing nations face marginalization due to low literacy levels and socio-economic challenges, leading to reduced interest in fishing and the loss of vital, undocumented tacit knowledge. To address this, we conducted focus group discussions and interviews in Bangladesh and Indonesia, extracting knowledge components from the conversational data to develop knowledge graphs. These graphs visually represent facts, attributes, and relationships, facilitating knowledge extraction. We compared manual and automated graph development using Large Language Models (LLMs), demonstrating their potential to systematically identify, preserve, and share critical fishing knowledge. Future work involves testing this framework with fisherfolk to preserve and disseminate this essential knowledge.

1 INTRODUCTION

From our experience with a large-scale ICT4D (Information and Communication for Development) project, we worked closely with fishing communities in two developing nations to empower marginalized fisherfolk through digital technologies. During the user-centred design of prototype solutions, a critical challenge emerged: the loss of vital fishing knowledge. Due to low literacy and socio-economic barriers, fisherfolk acquire tacit knowledge through observation and mentorship of experienced "seniors". Being unexpressed and often semi-conscious, Tacit knowledge is rarely documented (Polanyi, 2009). However, declining interest in the profession among younger generations threatens this knowledge, leading to a scarcity of skilled fisherfolk. To address this, identifying, systematically preserving, and sharing this invaluable tacit knowledge within the community are essential.

While previous research proposed user-focused applications to facilitate knowledge sharing (Kanij et al., 2023), these do not address the systematic preservation of tacit knowledge. We systematically explore methods to identify and preserve tacit knowledge to address this gap.

After reviewing various approaches, we adopted "Knowledge Graphs" for two key reasons: (1) their ability to graphically represent heterogeneous facts, attributes, and relationships, and (2) the ease of extracting knowledge from the graphs. Developing knowledge graphs for tacit knowledge posed computational challenges. Initially, we manually created knowledge graphs to evaluate their feasibility. Then, we leveraged large language models (LLMs) to automate the extraction of text-based tacit knowledge components and develop these graphs. This article details both approaches and discusses insights from a pilot study, highlighting the potential for large-scale deployment to systematically identify, preserve, and share vital fishing knowledge.

The rest of the paper is organised as follows: Section 2 presents theories of tacit knowledge and knowledge graphs and information on fisherfolk communities in Bangladesh and Indonesia, Section 3 presents the details of user studies, Section 4 describes the idea of developing knowledge graphs based on the findings, Section 5 and Section 6 illustrate the manual and automated process of knowledge graphs generation,

[a] https://orcid.org/0000-0002-5293-1718
[b] https://orcid.org/0000-0003-2722-6424
[c] https://orcid.org/0000-0003-3979-1746
[d] https://orcid.org/0000-0001-5040-8619
[e] https://orcid.org/0000-0003-4928-7076

respectively, Section 7 discusses the implications and challenges and finally Section 8 concludes the article.

2 BACKGROUND

2.1 Tacit Knowledge

The concept of tacit knowledge, introduced by Michael Polanyi in The Tacit Dimension, highlights that "we can know more than we can tell" (Polanyi, 2009). Polanyi illustrated this with examples like face recognition and bike riding, showing that tacit knowledge resides in unconscious understanding and is challenging to externalize. Philosophers and cognitive psychologists differentiate tacit (unconscious) from explicit (conscious) knowledge. Pathirage et al. describe tacit knowledge as rooted in personal experiences, shaped by factors like attitudes, emotions, and perspectives (Pathirage et al., 2008). Sanderson emphasizes physical skills gained through practice and implicit learning (Sanderson, 2001), while Clarke highlights its evolving nature through interaction and experience (Clarke, 2010).

Nonaka and Takeuchi define tacit knowledge as personal, context-specific, and hard to formalize, contrasting it with explicit knowledge, which is codified and easily communicated (Nonaka and Takeuchi, 1995). Leonard and Sensiper position these as opposite ends of the knowledge spectrum (Leonard and Sensiper, 1998). Knowledge management literature debates two views: (1) knowledge exists in distinct structures (unconscious and conscious) that are not mutually convertible, and (2) tacit and explicit knowledge are two states that can be converted (Grandinetti, 2014). Nonaka et al. support the externalization of tacit knowledge into explicit forms but note that this process is complex due to its integration with dynamic human processes (Nonaka, 2009).

To address these challenges, an Entity-Relationship Management (ERM) system has been proposed to codify tacit knowledge into explicit forms and manage it in a digital repository (Chen and Nunes, 2019). Unlike traditional databases reliant on Boolean logic, the ERM model is envisioned as a flexible, heterogeneous repository capable of accommodating the complexities of tacit knowledge.

2.2 Knowledge Graph (KG)

Knowledge Graphs (KGs) organize and link information in a graph-like structure, where nodes represent entities (e.g., people, places, or concepts) and edges define their relationships; as semantic networks, they integrate diverse information sources to represent knowledge on specific topics (Fensel et al., 2020). Knowledge Management Systems structure and connect information, simplifying navigation and discovery within organizations (Dalkir, 2013).

KGs capture complex relationships and semantic information, enabling the querying of interconnected data, supporting knowledge discovery, decision-making, and efficient information retrieval. Consequently, they are widely adopted in AI, data integration, NLP, recommendation systems, and general knowledge management (Ladeinde et al., 2023).

2.3 Fisherfolk

2.3.1 Fisherfolk in Bangladesh

Due to its geographical location, Bangladesh is prosperous in rivers and wetlands, making fishing a crucial livelihood for a significant portion of its population. Mahmud et al. (Mahmud et al., 2015) estimate that 10% of the population depends on fishing, while Tran et al. (Tran et al., 2023) report fisheries support 12% of the 170 million residents in full-time and part-time roles. Livelihood experiences and challenges vary by region and water source (rivers, seas, or wetlands) (Kabir et al., 2012), with common challenges including poverty, low living standards, limited credit, lack of knowledge, disasters, and diseases (Kabir et al., 2012; Das et al., 2015; Mahmud et al., 2015; Hossain et al., 2009).

Recent studies emphasize the importance of preserving and sharing tacit knowledge among fisherfolk. Kanij et al. (Kanij et al., 2023) designed a mobile application to facilitate knowledge exchange among boat captains, aligning with Shuva et al.'s findings that fisherfolk primarily rely on friends and family for information (Shuva, 2017). Both studies stress the importance of transforming the "information service culture" to promote equitable information access, addressing barriers such as poor mobile networks, radio signal limitations, illiteracy, and poverty. Furthermore, Miah and Islam (Miah and Islam, 2020) highlighted that inequitable benefits, weak regulations, and power imbalances also shape these practices.

2.3.2 Fisherfolk in Indonesia

Like Bangladesh, Indonesian fisherfolk communities face significant socio-economic and environmental challenges, including poverty, income instability, limited access to essential services, and environmental degradation (Prasetyo et al., 2023). Despite efforts by government and NGOs, these initiatives often fail to address entrenched structural issues (Supriati and Umar, 2020; Prasetyo et al., 2023).

Local wisdom and traditional knowledge are vital in sustaining livelihoods and social cohesion, as seen in Bunaken Island communities (Supriati and Umar, 2020). Practices such as sustainable fishing methods, cultural rituals, and communal activities promote environmental protection and unity, embodying values of cooperation and respect for nature.

3 DATA COLLECTION

To explore knowledge management practices and identify critical tacit fishing knowledge, we conducted eight focus group discussions each in Bangladesh (Barguna and Chandpur districts) and Indonesia (South Galesong, Mangarabombang, and North Galesong districts). The participants, involved in both marine and inland fisheries, were recruited through trusted local partner organizations.

Separate focus groups were held for men and women, employing a storytelling approach that began with, “Please describe a usual day in your life.” Discussions covered various topics, including demographics, information practices, decision-making, responsibilities, career aspirations, training, challenges, and environmental issues. Although not directly involved in fishing, women fisherfolk also shared their contributions. Focus group had 8–12 participants.

In addition, we interviewed 20 key stakeholders (referred to as Key informants (KI)) working closely with fisherfolk, including local government officials, fish business representatives, NGO workers, and fisherfolk’s association leaders to explore stakeholder engagement with fisherfolk and their perceptions of livelihood challenges, expectations, and tacit knowledge practices.

All focus groups were in person, and the interviews were both face-to-face and online, in local languages, and audio-recorded with participants’ consent. Using Thematic Analysis (Nowell et al., 2017), we developed over 200 codes grouped into 12 themes to find insight from the data.

Our findings highlight that fisherfolk’s tacit knowledge management involves practical and cultural elements deeply embedded in generational wisdom and integral to daily life. It encompasses ecological understanding, economic considerations, and socio-cultural practices; the key features are:

- The information/knowledge and their respective sources are mainly heterogeneous.
- The relationships between knowledge components are complex and do not always conform to a specific pattern (semantic ambiguity).

4 REPRESENTATION OF TACIT KNOWLEDGE OF FISHING WITH KNOWLEDGE GRAPHS

Our research highlights the reliance of fisherfolk in Bangladesh and Indonesia on tacit knowledge for fishing, emphasizing the urgent need to document and systematically preserve this critical resource. To address this, we developed Knowledge Graphs (KGs) that offer structured visual representations of diverse fishing information, showcasing complex interconnections to aid in preserving and visualising tacit fishing knowledge.

KGs enhance accessibility and streamline the sharing and preservation of fishing knowledge, promoting their adoption among fisherfolk in both countries. Their selection followed an evaluation of paradigms like knowledge bases, ontologies, and semantic networks. Unlike knowledge bases, which focus on structured data storage, KGs utilize graph structures to capture and represent experiential knowledge effectively.

While ontologies enable semantic modeling and link concepts, they lack the flexibility and granularity of KGs. By integrating varied data types and representing complex, domain-specific relationships, KGs proved to be the most suitable approach for preserving and sharing the nuanced knowledge of fisherfolk.

5 DEVELOPMENT OF KNOWLEDGE GRAPHS

A manual approach was initially adopted to ensure the effectiveness and accuracy of the knowledge graphs. Focus group discussions and interview transcripts were used as inputs to extract knowledge components. These responses typically contained long sentences, so the information was first organized by separating sentences based on their themes.

The subjects of these sentences were identified to create graph nodes, and their relationships to themes were structured and encoded in Cypher. The queries were executed in Neo4j (NEuler) to construct and display the knowledge graph, leveraging Neo4j’s efficiency in extracting knowledge components with Cypher. The process is shown in Figure 1.

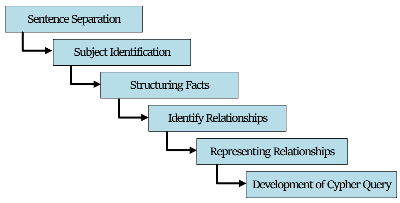

Figure 1: KG Development Process (Manual)

5.1 Demonstration with An Example

The following example demonstrates the manual process of developing knowledge graphs using a focus group discussion and interview transcript.

The transcript from an Indonesian focus group reads as follows: ”We all use Jabba (fish trap) to catch fish. Jabba we set at the bottom of the lake. So those of us who use Jabba as our main means of catching fish must be clever and endure diving to the bottom of the lake. We don’t have a specific time when we should install Jabba. It just depends on the feeling and the proper time slot available.”

Table 1: Sentence Separation & Subject Identification

Sentence Separation	Subject Identification
S1: We all use Jabba (fish trap) to catch fish.	We [Sub: Fisherfolk]
S2: Jabba we set at the bottom of the lake.	Jabba [Sub: Jabba]
S3: Those of us who use Jabba as our main means of catching fish must be clever and endure diving to the bottom of the lake.	Those of us [Sub: Fisherfolk]
S4: We don't have a specific time when we should install Jabba.	We [Sub: Fisherfolk]
S5: It depends on the feeling and proper time slot available.	It [Sub: Jabba installation]

Table 2: Fact & Relationship Identification and Representation

Fact Identification	Relationship Identification	Relationship Representation
Jabba is a fish trap.	Jabba ≡ Fishtrap	Jabba ≡ Fishtrap
Fishermen use Jabba to catch fish.	Relationship: Fishermen use Jabba nets to catch fish from the lake.	Fishermen (identity) [:use] Jabba (tool) to catch fish.
Jabba is set at the bottom of the lake.	Relationship: Deepest part of the lake location where Jabba is positioned for use.	Jabba(tool) [:is_set_at] the bottom (position) of the lake(location).
Fishermen require wit and diving endurance to set up the Jabba.	Relationship: Deploying Jabba requires fishermen to demonstrate wit and diving endurance to reach and install it at the lake bottom.	Fishermen (identity) [:require] wit (attribute) and diving endurance (attribute) to set up Jabba (tool).
Jabba installation depends on the feeling and availability of proper time slot.	Relationship: The decision to install Jabba underwater is influenced by subjective factors, intuition, and practical considerations, such as the availability of a suitable time slot.	Jabba (tool) [:installation_depends_on] the intuition (attribute) and availability of proper time slot (attribute).

Based on the method outlined in Section 5 and shown in Figure 1, the following steps were followed:

1. **Sentence Separation** The interview response was split into five sentences.
2. **Subject Identification** The subjects of each sentence were identified by analyzing both grammatical structure and context to ensure accurate understanding. This step is shown in Table 1.
3. **Structuring Facts** We extracted five key facts from the sentences and subjects. To simplify, we assumed that all fisherfolk in the example used "Jabba" for fishing.
4. **Identifying Relationships** Next, Five relationships were identified from the facts extracted.
5. **Representing Relationships** The relationships were formatted for easy conversion into Cypher queries. Facts were enclosed in curly braces, with relationships following the same format, starting with a colon (:), as illustrated in Table 2.
6. **Development of Cypher Query** Nine Cypher queries were created for node generation and six for establishing relationships, executed on the Neo4j platform to produce the knowledge graph shown in Figure 2.

Creating nodes:

```
CREATE (fishermen:Identity {name:'Fishermen',
description:'Catches fish using Jabba' })
CREATE (fish:Identity {name:'Fish'})
CREATE (jabba:Tool {name:'Jabba', description:'Fish trap' })
CREATE (depths:Position {name:'Depths', description:'Bottom
of the lake' })
CREATE (lake:Location {name:'Lake'})
CREATE (wit:Attribute {name:'Wit skills'})
CREATE (diving:Attribute {name:'Diving skills'})
CREATE (intuitiveTiming:Attribute {name:'Intuition'})
CREATE (availability:Attribute {name:'Timeslot
availability'})
```

Creating relationships:

```
CREATE (fishermen)-[: USE]→(jabba)-[: TO_CATCH]→(fish)
CREATE (jabba)-[: IS_SET_AT]→(depths)-[: OF]→(lake)
CREATE (fishermen)-[: REQUIRE]→(wit)-[: TO_SET_UP]→(jabba)
CREATE (fishermen)-[: REQUIRE]→(diving)-[: TO_SET_UP]→(jabba)
CREATE (jabba)-[: SET_UP_DEPENDS_ON]→(intuitiveTiming)
CREATE (jabba)-[: SET_UP_DEPENDS_ON]→(availability)
```

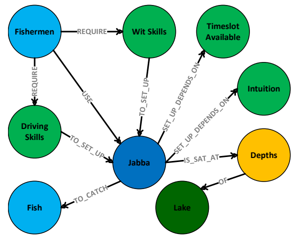

Figure 2: Knowledge Graph Snippet

5.2 Feedback on KG

We developed separate KGs for each country, with the Indonesian knowledge graph presented to five experts academics from Indonesian universities, with experience working with fisherfolk in the region. All experts welcomed the idea of visualizing knowledge in this format but expressed concerns about the complexity of the graphs as more knowledge components are added.

This manual exercise was a feasibility study to assess whether knowledge graphs could effectively preserve tacit knowledge within fisherfolk communities in developing countries. The successful creation of two comprehensive knowledge graphs and positive expert feedback suggests that this approach is viable for preserving and sharing tacit knowledge.

6 AUTOMATED DEVELOPMENT OF KNOWLEDGE GRAPHS

Building on the successful creation of knowledge graphs and positive feedback, we streamlined the process by employing computational methods to extract facts and relationships from text, automating the generation of Cypher queries.

This section explains the automated extraction of tacit knowledge from translated focus group discussions and interview transcripts (in English) to create knowledge graphs. Using a six-step framework (see Figure 3), the process mirrors the manual approach, with the best algorithms validated against manual results. Each step is detailed in the following subsections.

6.1 Demonstration with An Example

In this section, with an example, we demonstrate the automated process of developing knowledge graphs from the focus group discussion and interview transcripts. We will consider the previous example transcript from the manual KG part Section 5.1.

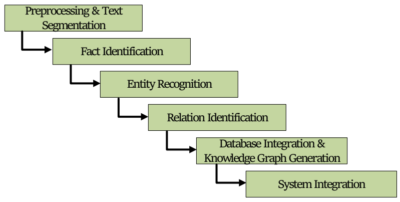

Figure 3: Automated Knowledge Graph Methodology

6.1.1 Preprocessing and Text Segmentation

The preprocessing and segmentation stage was vital in preparing transcription data for knowledge graph creation. It automatically identifies and extracts question-answer (Q&A) pairs and filters sentences containing key knowledge components while removing irrelevant information.

Initial Approach: SpaCy (Honnibal and Montani, 2017) retained conversational parts but lacked precise filtering and segmentation, often missing relevant tacit fishing knowledge. NLTK (Bird et al., 2009) improved segmentation using its Text-Filling Algorithm but struggled with accuracy and filtering irrelevant content. Both identified five and four Q&A pairs in the example transcript but failed to filter sensitive information consistently.

Final Approach: The Mistral 7B Model (Jiang et al., 2023), trained without tacit fishing knowledge, significantly improved text segmentation by accurately pairing relevant questions and answers. This ensured precise filtering, retaining only sentences with valuable tacit fishing knowledge.

For example, it identified the following Q&A pair:
"Question": *"Why do you use Jabba?"*,
"Answer": *"We use Jabba (fish trap) to catch fish. We set Jabba at the bottom of the lake."*
The automation reduced manual effort and enabled faster, more efficient processing of large datasets.

6.1.2 Summarisation and Fact Identification

This step involves identifying essential facts from Q&A pairs to create summarised sentences that do not contain pronouns.

Initial Approach: Various supervised and unsupervised machine learning techniques were explored for extracting meaningful facts from text, including libraries like spaCy and NLTK, LLMs such as Bard, GPT, Llama 2, Facebook/bart-large-cnn, and summarization models like Falconsai/text_summarization and IlyaGusev/mbart_ru_sum_gazeta.

Testing on 20 segmented paragraphs revealed Facebook/bart-large-cnn as the most accurate, though it introduced pronouns, ambiguous sentences, and omitted crucial details. For example: Facebook/bart-large-cnn: *"We all use Jabba (fish trap) to catch fish. Jabba we set at the bottom of the lake. We don't have a specific time when we should install Jabba."*.
Llama 2 initially produced accurate outputs without omitting key facts but struggled with larger inputs by generating random content.

Final Approach: The Mistral 7B model, chosen over Llama 2 for its efficiency and accuracy, addressed context retention and fact identification.

Prompt engineering preserved critical information by removing pronouns and refining context. Prompts also extracted fishing-related tacit knowledge while maintaining gender-inclusive language by referring to individuals as fisherfolk.

Using the demo example, the Mistral 7B model summarised the sentences as follows, **Sentence 1:** *"Fisherfolk use Jabba as a fishing tool."* **Sentence 2:** *"Jabba is set at the lake bottom for catching fish."*

6.1.3 Named Entity Recognition (NER)

This step focuses on categorisation accuracy and efficiency. It extracts key concepts from the input data and categorises them into predefined labels.

Initial Approach: The initial use of spaCy struggled with entity extraction, especially native terms from Bangladesh and Indonesia. BERT (Devlin et al., 2019) performed poorly on unstructured data and native words. Manual labelling via Doccano was inefficient, leading to the adoption of the OpenAI API for auto-labelling, which improved consistency. However, fine-tuning BERT with limited training samples still yielded low accuracy. Both approaches misidentified Jabba as PERSON.

Final Approach: The Mistral 7B model effectively captured key facts from summarised text and accurately identified entities, categorizing them into 16 predefined labels (Table 3). Context-based categorization structured the knowledge for graph creation. Automation of labelling expedited processing and significantly reduced manual effort for large datasets.
Using the demo example, the following entities were found in each sentence: **Sentence 1**: "Fisherfolk use Jabba as a fishing tool." **Entities:** [{"entity": "Jabba", "label": "TOOL" }, {"entity": "Fisherfolk", "label": "GROUP" }]; **Sentence 2:** "Jabba is set at the lake bottom for catching fish." **Entities:** [{"entity": "Jabba", "label": "ID" }, {"entity": "lake bottom", "label": "LOC" }, {"entity": "for catching fish", "label": "ACT" }, {"entity": "fish", "label": "FISH" }]

6.1.4 Relationship Identification

This component identifies relationships between key entities in a text, represented as edges in Cypher queries for the Neo4j knowledge graph.

Initial Approach: The BERT model, fine-tuned with OpenNRE (Han et al., 2019), struggled with low accuracy due to limited training data and predefined relations, resulting in generic labels. TinyLlama was

Table 3: Description of the 16 Entity Categories

Entity Name	Description
Identity (ID)	Names of individuals or entities involved in fishing practices (e.g., "Fisherman Ali")
Location (LOC)	Geographical locations relevant to fishing activities (e.g., "Bay of Bengal")
Tool (TOOL)	Equipment and tools (e.g., "Fishing Net")
Food (FOOD)	Types of food or bait (e.g., "Shrimp")
Season (SSN)	Seasonal information affecting fishing practices (e.g., "Monsoon season")
Attribute (ATT)	Characteristics or qualities related to fishing (e.g., "Strong currents")
Product Aspect (PROD)	Specific aspects of fishing products (e.g., "Fish size")
Infosource (INFO)	Sources of information such as local knowledge or expert advice (e.g., "Local guide")
Group (GRP)	Groups or communities involved in fishing (e.g., "Fishing cooperative")
Factor (FAC)	Factors influencing fishing practices (e.g., "Water temperature")
Date (DATE)	Specific dates relevant to fishing activities (e.g., "March 2023")
Time (TIME)	Time-related information (e.g., "Early morning")
Countable (CNT)	Quantitative data such as counts or measurements (e.g., "10 nets")
Concept (CON)	Abstract concepts or ideas related to fishing (e.g., "Sustainability")
Activity (ACT)	Specific activities or actions related to fishing (e.g., "Casting nets")
Cost (COST)	Financial aspects or costs associated with fishing (e.g., "Cost of the boat")

used to identify entity relationships but often produced hallucinated, unstructured outputs. Efforts to format results in JSON increased inaccuracies, highlighting the need for a more robust LLM.

Final Approach: The Mistral 7B model, optimized using Unsloth techniques for efficiency, significantly improved accuracy in defining relationships between multiple entities within sentences. Prompt engineering, similar to the approach with TinyLlama, ensured the output was structured in JSON format, enabling easier post-processing and seamless system integration.

Mistral 7B, using the demo example, extracted three relations from sentences,

- { "node_1": "Fisherfolk", "node_2": "Jabba", "relation": "FISHING_TOOL" }
- { "node_1": "Jabba", "node_2": "lake bottom", "relation": "PLACED_AT" }
- { "node_1": "Jabba", "node_2": "fish", "relation": "CATCHES" }

and two unique entities as a part of the relations:

- { "entity": "fisherfolk", "label": "UNKNOWN" }
- { "entity": "jabba", "label": "UNKNOWN" }]

While Mistral 7B showcased substantial advancements over others, future iterations will require further refinement through prompt engineering and the reduction of hallucinations.

6.1.5 Database Integration & KG Generation

After extracting nodes, relationships, and labels, the next step was to generate knowledge graphs by integrating the Neo4j database, executing Cypher queries, and visualizing the results. A cloud-based solution was implemented to optimize data ingestion, query execution, and graph visualization. Below are the key steps:

Database Integration & Cypher Queries: Neo4j AuraDB was selected for its scalability, reliability, and ability to efficiently handle large datasets and complex queries. Cypher, a declarative language tailored for graph databases, was used to query and update the database. Two primary Cypher queries were developed: one created nodes with specified labels and entities, ensuring no duplication. In contrast, the other established relationships between nodes only if they did not already exist.

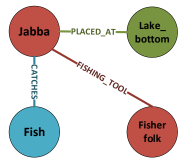

Figure 4: Automated Knowledge Graph for Demo Example

KG Generation: The knowledge graph data was retrieved and visualised once the database was populated with nodes and relationships. Python's Plotly and NetworkX (Hagberg et al., 2008) libraries were used for creating interactive and visually appealing graphs. Nodes, labels, and relationships from AuraDB were processed to generate an interactive graph with distinct colours for node labels, annotated relationships, and individual node visualizations for detailed insights. NetworkX managed the graph structure, while Plotly rendered the visuals. The demo example is presented in Figure 4.

6.1.6 System Integration

Google Colab was selected for its intuitive interface and seamless integration, facilitating cohesive functionality and efficient execution. The pipeline streamlined data preprocessing, entity extraction, relation identification, and query generation, creating a modular, maintainable, and scalable system. Validation using Mistral 7B ensured correct formatting, data integrity, and compliance with a valid JSON structure, which was then converted into Python dictionaries for efficient handling and manipulation.

7 DISCUSSION

This pilot study created knowledge graphs from conversational data with fisherfolk in Bangladesh and Indonesia. Feedback on the manually developed graphs underscored their value in preserving orally transmitted tacit knowledge. These KGs provide an innovative way to document such knowledge by capturing complex relationships and contextual nuances. Integrating automated approaches with LLMs like Mistral 7B shows promise for scalability and efficiency, enabling systematic preservation on a scale unattainable through traditional methods.

7.1 Technical Contributions

The successful implementation of this framework serves as the "back-end" module for a complete platform aimed at preserving and sharing tacit knowledge. Fisherfolk would communicate knowledge via a user interface in the proposed platform, while the back-end would extract knowledge components and construct knowledge graphs. Future plans include developing a front-end interface and integrating it with the back-end module to assess the feasibility of the entire process, as visualised in Figure 5. KGs effectively preserve tacit knowledge, enabling easy extraction via Cypher queries. Future plans include an interface for fisherfolk to ask questions, automatically converting them into Cypher queries for retrieving specific knowledge.

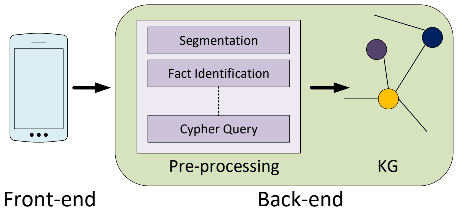

Figure 5: Proposed tacit knowledge management platform

Currently, no LLMs, including spaCy and Mistral 7B, address the specific needs of the fishing occupation. Developing specialized LLMs could help fisherfolk share knowledge efficiently and process queries to retrieve information from knowledge graphs via Cypher, improving usability and accessibility.

7.2 Societal & Broader Implications

This work's societal impact extends beyond preserving knowledge, addressing key socio-economic challenges fisherfolk face. Systematizing tacit knowledge enables intergenerational transfer, with knowledge graphs helping younger fisherfolk acquire traditional skills and bridging generational gaps amid declining interest in the profession. Moreover, the framework democratizes access to critical fishing knowledge, reducing dependence on hierarchical knowledge-sharing networks and advancing digital inclusion by integrating traditional practices.

The knowledge graphs provide actionable insights for policymakers and NGOs, aiding in designing targeted interventions. For example, sustainable fishing practices encoded in the graphs can inform environmental conservation efforts while supporting livelihood sustainability. The model's adaptability also makes it a blueprint for preserving and sharing tacit knowledge in other marginalized communities, such as those relying on indigenous agriculture or forestry.

This approach is crucial for preserving cultural heritage and promoting sustainable practices facing socio-economic and environmental changes. It also advances digital equity by demonstrating how technologies like LLMs can serve marginalized communities, fostering inclusive technological progress.

7.3 Challenges

Focus group discussions and interviews were conducted in the local language, transcribed, and translated into English, a process that can alter meanings. To mitigate this, multiple researchers collaborated to cross-check the translations.

The accuracy of the KGs generated using automated algorithms on the Google Colab platform was manually evaluated. However, the evaluation was limited due to the small dataset from focus groups and interviews, with plans to expand data collection. While the Mistral 7B model outperformed its peers, it occasionally produced hallucinated information, resulting in undocumented entities and relationships. Future efforts will focus on reducing such hallucinations through refined prompt engineering and enhanced model training with more labelled data.

8 CONCLUSION

This study introduces an innovative approach to preserving tacit fishing knowledge for fisherfolk in Bangladesh and Indonesia. Initially, knowledge was manually identified from collected data, structured into entities and relationships, and visualized as knowledge graphs using Cypher queries on the Neo4j platform. Encouraged by positive feedback on the manually developed graphs, we experimented with various LLMs to automate the process, following similar steps and verifying outcomes at each stage.

Despite challenges encountered during development, the automatically generated knowledge graphs demonstrate significant potential for leveraging LLMs to help fisherfolk preserve their tacit knowledge. This approach can also benefit other marginalized communities in similar contexts. Future plans include trialling the framework with fisherfolk in Bangladesh and Indonesia, addressing identified challenges, and refining the methodology. We also propose future research directions to explore the full potential of this work.

ACKNOWLEDGEMENTS

Kanij and Grundy are supported by ARC Laureate Fellowship FL190100035. Research funding is also provided by the Empowerment Charitable Trust and the Whyte Fund. We thank all the participants and local collaborators.

REFERENCES

Bird, S., Klein, E., and Loper, E. (2009). *Natural Language Processing with Python: Analyzing text with the natural language toolkit*. O'Reilly Media, Inc.

Chen, H. and Nunes, M. B. (2019). Retaining professional tacit knowledge and evidence of experience through electronic records management. *IADIS International Journal on Computer Science & Information Systems*, 14(2).

Clarke, T. (2010). *The Development of a Tacit Knowledge Spectrum based on the Interrelationships between Tacit and Explicit Knowledge*. PhD thesis, Cardiff Metropolitan University.

Dalkir, K. (2013). *Knowledge Management in Theory and Practice*. Routledge.

Das, M., Ray, S., Kumar, U., Begum, S., and Tarafdar, S. (2015). Livelihood assessment of the fishermen community in the South West region of Bangladesh. *Journal of Experimental Biology and Agricultural Sciences*, 3:353–361.

Devlin, J., Chang, M.-W., Lee, K., and Toutanova, K. (2019). BERT: Pre-training of Deep Bidirectional transformers for language understanding.

Fensel, D., Şimşek, U., Angele, K., Huaman, E., Kärle, E., Panasiuk, O., Toma, I., Umbrich, J., and Wahler, A. (2020). *Introduction: What Is a Knowledge Graph?*, chapter 1, pages 1–10. Springer International Publishing, Cham.

Grandinetti, R. (2014). The explicit dimension: What we could not learn from polanyi. *The Learning Organization*, 21(5):333–346.

Hagberg, A., Swart, P., and S Chult, D. (2008). Exploring network structure, dynamics, and function using networkx. Technical report, Los Alamos National Lab.(LANL), Los Alamos, NM (United States).

Han, X., Gao, T., Yao, Y., Ye, D., Liu, Z., and Sun, M. (2019). OpenNRE: An open and extensible toolkit for neural relation extraction. *arXiv preprint arXiv:1909.13078*.

Honnibal, M. and Montani, I. (2017). spaCy 2: Natural language understanding with Bloom embeddings, convolutional neural networks and incremental parsing. *To appear*, 7(1):411–420.

Hossain, I., Chamhuri, S., Mokhtar, M., Madan, M., Dey, M., and Jaafar, A. (2009). Socio-economic condition of fishermen in seasonal floodplain beels in Rajshahi District, Bangladesh. *Research Journal of Social Sciences*, 4:74–81.

Jiang, A. Q., Sablayrolles, A., Mensch, A., Bamford, C., Chaplot, D. S., de las Casas, D., Bressand, F., Lengyel, G., Lample, G., Saulnier, L., Lavaud, L. R., Lachaux, M.-A., Stock, P., Scao, T. L., Lavril, T., Wang, T., Lacroix, T., and Sayed, W. E. (2023). Mistral 7B.

Kabir, K., Adhikary, R., Hossain, M. B., and Minar, M. H. (2012). Livelihood Status of Fishermen of the Old Brahmaputra River, Bangladesh. *World Applied Sciences Journal*, 16:869–873.

Kanij, T., Anwar, M., Oliver, G., and Hossain, M. K. (2023). Developing software for low socio-economic end users: Lessons learned from a case study of fisherfolk communities in bangladesh. In *2023 IEEE/ACM 45th International Conference on Software Engineering: Software Engineering in Society (ICSE-SEIS)*, pages 96–107. IEEE.

Ladeinde, A., Arora, C., Khalajzadeh, H., Kanij, T., and Grundy, J. (2023). Extracting queryable knowledge graphs from user stories: An empirical evaluation. In *International Conference on Evaluation of Novel Approaches to Software Engineering 2023*, pages 684–692. Scitepress.

Leonard, D. and Sensiper, S. (1998). The role of tacit knowledge in group innovation. *California management review*, 40(3):112–132.

Mahmud, S., Ali, M. L., and Ali, M. M. (2015). Present scenario on livelihood status of the fishermen in the Paira River, Southern Bangladesh: Constraints and Recommendation. *International Journal of Fisheries and Aquatic Studies*, 2(4):23–30.

Miah, M. and Islam, M. (2020). Defending fishing space: Conflicts and marginalization of small-scale fisheries in the Saint Martin's Island of Bangladesh. In *Blue Justice For Small-Scale Fisheries: A Global Scan*, pages 1–3. TBTI Global Publication Series, St. John's, NL, Canada.

Nonaka, I. (2009). The knowledge-creating company. In *The economic impact of knowledge*, pages 175–187. Routledge.

Nonaka, I. and Takeuchi, H. (1995). *The Knowledge-Creating Company: How Japanese Companies Create the Dynamics of Innovation*. Oxford University Press.

Nowell, L. S., Norris, J. M., White, D. E., and Moules, N. J. (2017). Thematic analysis: Striving to meet the trustworthiness criteria. *International Journal of Qualitative Methods*, 16(1):1–13.

Pathirage, C., Amaratunga, D., and Haigh, R. (2008). Tacit knowledge generation and utilisation in the construction industry: from process perspective. In *Proceedings of the RICS Construction and Building Research Conference (COBRA 2008)*.

Polanyi, M. (2009). *The tacit dimension*. Routledge.

Prasetyo, T., Widodo, P., Saragih, H., Suwarno, P., and Said, B. (2023). Poverty Reduction For Coastal Communities In Indonesia Through Community Empowerment Training. *International Journal Of Humanities Education and Social Sciences (IJHESS)*, 2.

Sanderson, M. (2001). Records management and the capture of tacit knowledge. *Records Management Journal*, 11:7–17.

Shuva, N. Z. (2017). The Information Practices of the Fishermen in the Bay of Bengal, Bangladesh. *Open Information Science*, 1(1):21–40.

Supriati, A. and Umar, M. (2020). The role of coastal communities to preserve local wisdom (a study in community of bunaken island, indonesia). In *2nd Annual Civic Education Conference (ACEC 2019)*.

Tran, N., Rodriguez, U.-P., Chan, C. Y., Aung, Y. M., Chu, L., Islam, A. H. M. S., Barman, B. K., and Phillips, M. J. (2023). Future scenarios of fish supply and demand for food and nutrition security in Bangladesh: An analysis with the AsiaFish model. *Aquaculture*, 568:739288.