Understanding Mineral Depositional Conditions from Geological Reports: A Knowledge Mining Approach

Eun-Jung Holden¹, Majiguren Enkhsaikhan², Wei Liu², Paul Duuring³, Tom Horrocks¹, Daniel Wedge¹

¹Centre for Exploration Targeting, The University of Western Australia
²Department of Computer Science & Software Engineering, The University of Western Australia
³Geological Survey and Resource Strategy Division, Department of Mines, Industry Regulation and Safety, Western Australia
Geological knowledge/Concept Discovery

“Nickel is deposited from an immiscible sulphide liquid in rapidly crystallising ultramafic extrusive rocks. ... The most important deposits of this type occur in komatiites, high-temperature magnesium-rich ultramafic rocks aged 3800 – 2500 Ma from the Archaean Eon.”

- KNOWLEDGE – collective information (e.g. concepts) that are acquired in the field of expertise
  - INFORMATION - refined form of data
    - DATA – structured (e.g. drill hole petrophysical measurements), unstructured (e.g. documents), semi-structured (e.g. field observations)
- Knowledge of interest for mineral exploration - a conceptual model
- Knowledge discovery from text is an active area of research in ML & AI
- Research Question: Can automated analysis of reports assist us in ‘evidence based’ conceptual modelling about mineral depositional conditions?
- Approach: Transform geological text into a structured database where knowledge can be stored, integrated and retrieved (re-usable) by harnessing the advances in natural language processing; and syntactic (structure of sentences) and semantic (meaning) analysis of text
Recent Advances in Text Analysis

- **Natural language processing (NLP)** allows ‘machine reading’ of text using linguistic analysis
- **Text data mining** (uses NLP) reveals meaning of text and/or recognise patterns across documents
  - Text categorisation, clustering, concept/entity extraction, document summarisation
  - Social media monitoring – user sentiment (mood, emotions and awareness)
- Interlinking information between different data types
  - A powerful knowledge discovery system for decision support

---

**Q:** Who is to the left of Barack Obama?
**A:** Richard Cordray

**Q:** Do all the people in the image have a common occupation?
**A:** Yes

**Q:** Who among the people in the image is called by the nickname Barry?
**A:** Person in the center

---

Advances in Geological Document Analysis

- Machine analysis of geological documents exists
  - Paleo DeepDive (Peters et al., 2014)
    - Locates and extracts text, tables and figures in publications – compilations of paleontological data i.e. geological history of taxonomic diversity and genus-level rates of origination and extinction (structured data)
  - Time scale analysis of stromatolites (Peters et al., 2017)
    - Machine reading of stratigraphic database and published literature - the occurrence of stromatolites in North America over geological timeline
  - Geological knowledge graph construction from Chinese geoscience literature (Wang et al., 2018, 2019)
  - Mineral deposit related text analysis for Lala copper deposit (Shi et al., 2018)
CET’s GeoDocA
(Geological Document Analysis) System

Performs machine reading of 25,419 WAMEX reports from GSWA

• Extracts and interactively visualises ‘geological contents’ within a report – **report summary graph**

• Extracts and visualises figures and tables within a report – **thumbnails**

• Identifies **similar reports** based on geological contents

• Assists the search of reports using **auto-completion of search terms** based on ‘learnt’ key word associations

https://www.youtube.com/watch?v=PBbDhRslBNY

Geological Knowledge Discovery

- **Document Retrieval** to **Knowledge Retrieval**

- PhD research by *Majigsuren Enkhsaikhan* (CSSE, UWA, Supervisors - Wei Liu, Eun-Jung Holden, Paul Duuring)

- **Challenge**: Building a robust conceptual model for mineral exploration framework is challenging

- **Approach**: Extract, store and retrieve the relations between locations, geological era, stratigraphy, rocks, minerals and ore/deposits using **Knowledge Graph** (Graph based database)

![Diagram of geological knowledge graph]

- **Locations** (mine, town, project, tenement, lease)
- **Geological Era** (Eon, Era, Period, Epoch, Age, Chron)
- **Stratigraphy** (Supergroup; Group; Subgroup; Formation; Member; Bed(s)) (Supersuite; Suite)
- **Rocks**
- **Minerals**
- **Ore/deposits**
Knowledge Graph (KG)

- How do we answer questions like:
  - “what are hosting rock types that have high probabilities of nickel deposits?”
  - “what is the stratigraphy relationship for the iron ore deposit X?”

- KG stores entities (locations, geological era, stratigraphy, rocks, minerals, ore/deposits) and their relationships

- Google, Facebook, Microsoft operate their own knowledge graph to store and retrieve information
  - Enhances AI systems (Alexa, Siri, Google assistant)
“The Knowledge Graph is a knowledge base used by Google and its services to enhance its search engine's results with information gathered from a variety of sources. The information is presented to users in an infobox next to the search results.”

(https://en.wikipedia.org/wiki/Knowledge_Graph)
A. Archaean sedimentary rocks occurred within the Kalgoorlie Group. Most sedimentary rocks contain either quartz or calcite.

B. Archaean sedimentary rocks occurred within Kalgoorlie Group. Sedimentary rocks contain quartz. Sedimentary rocks contain calcite.

C. Knowledge Graph

- STRAT: Kalgoorlie Group
  - occurs_in
  - ROCK: sedimentary rock
  - contains
  - MINERAL: quartz
  - contains
  - MINERAL: calcite

- TIMESCALE: Archaean

Knowledge Graph
Entity Extraction

- A sequence labeling model with bidirectional long short term memory (BiLSTM) neural network is used to recognise word sequences and label their entity types.
- 500 reports are labelled using the dictionary and validated for the training dataset.
- The 30,000 more reports are labelled with the sequence labelling model.

Relation Extraction

- Two nodes (geological entities) can be connected through a variety of expressions
- Extract relation types (generalised relations)
Relation Extraction using Triples

**head_entity** *relation_words* **tail_entity**

“Iron ore is confined to Robinson Range Formation”

**Table 1.** The semantic relations: *Name* is the simplified name for the relation, *Num* refers to a number of mentions in the text, % refers to a percentage of the mentions.

<table>
<thead>
<tr>
<th>Relation</th>
<th>Name</th>
<th>Num</th>
<th>%</th>
<th>Example clause</th>
</tr>
</thead>
<tbody>
<tr>
<td>Component-Whole</td>
<td>partOf</td>
<td>11,050</td>
<td>42.8%</td>
<td>[iron ore]_E1 is confined to robinson range formation]_E2</td>
</tr>
<tr>
<td>Cause-Effect</td>
<td>causes</td>
<td>329</td>
<td>1.3%</td>
<td>[iron]_E1 occur in banded iron formation]_E2</td>
</tr>
<tr>
<td>Content-Container</td>
<td>contains</td>
<td>2,332</td>
<td>9%</td>
<td>shale]_E1 found on [peak hill]_E2</td>
</tr>
<tr>
<td>Entity-Origin</td>
<td>originatesFrom</td>
<td>5,742</td>
<td>22.2%</td>
<td>[schist]_E1 derived from [andesite]_E2</td>
</tr>
<tr>
<td>Member-Collection</td>
<td>memberOf</td>
<td>6,027</td>
<td>23.3%</td>
<td>[chlorite]_E1 may represent [sediment]_E2</td>
</tr>
<tr>
<td>Message-Topic</td>
<td>isAbout</td>
<td>356</td>
<td>1.3%</td>
<td>[soil]_E1 covering [greenstone]_E2</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td></td>
<td>25,836</td>
<td>100%</td>
<td></td>
</tr>
</tbody>
</table>

SemEval2010 Task8 dataset (10 semantic relations) is used for training

Predicting relations of the triples using BiLSTM
Gold deposits in the Coolgardie Gold Project, Yilgarn Craton, Western Australia

The **Coolgardie** district is located on the western side of the **Archaean Menzies-Norseman Greenstone Belt**.

- **Coolgardie** is connected to **Menzies Norseman Greenstone Belt** through **Kalgoorlie Terrane**.
ROCKS: Coolgardie domain is described as a belt of complexly folded and faulted mafic to ultramafic volcanic rocks, with minor interbedded black shales and volcanioclastic rocks overlain by a thick succession of felsic volcanic and volcanioclastic rocks interbedded with clastic sedimentary rocks. The sequence is intruded by felsic to intermediate porphyries, dolerites, and gabbros. The Bali and Calooli granodiorites cause doming of the greenstone belt, and the wrapping of these rocks around the granitoid margins.
**Stratigraphy:** The relations include *Kalgoorlie Group* contains *Coolgardie Subgroup*, which contains *Lindsays Basalt*, *Gleesons Basalt* and *Brilliant Formation*; *Black Flag Group* overlain by *Gleesons Basalt*; *Brilliant Formation* current name of *Brilliant Ultramafics*.

**INTERLINKING OF Australian Stratigraphic Unit Database (ASUD) for stratigraphy hierarchy**
Entities
• The main ore deposits were lode gold deposit and supergene, while minerals included gold, quartz, sulphide, carbonate, chlorite, talc, oxide, nickel, lead, pyrite, pyrrhotite and biotite.

• The locations included Coolgardie, Big Blow, Denmark, Kalgoorlie, Burbanks, Black Flag, Kunanalling, and Emu Hill.

• Many rock types were connected with gold such as basalt, quartz vein, dolerite, amphibolite facies, ultramafic rock, ultramafic mafic rocks, breccia, granite greenstone, porphyry, calcrete, saprolite and granitoid.

Relationships
• Gold occurs in quartz, gold hosted in sulphide.
• Coolgardie contains gold, and gold deposit dominated by gold.
On-going & Future Development

• This study demonstrated:
  – Feasibility of transforming unstructured text data into a structured data in KG that is enriched with information such as labelled entities and their relation types
  – Integration of external databases such as ASUD further enriched KG

• Future use of KG for intelligent spatial search system
  – Improving visualisation, incorporating new/conflicting information
  – Extracting & spatially search for key relationships between entities (e.g. mineralisation & hosting rocks)
  – Extracting anomalous relationships between a mineral deposit and key elements associated with the deposit
  – Searching and spatially locating near-miss exploration cases by identifying non-geological keywords (e.g. event detection) & KG

• Future use of KG for updating databases
ACKNOWLEDGEMENTS:

- Geological Survey and Resource Strategy Division, Department of Mines, Industry Regulation and Safety, Western Australia
- Lynda Burnette, Chris Wijns, Jon Hronsky

PUBLICATIONS:

Enkhsaikhan et al., (In revision), “Understanding Mineral Depositional Conditions using Machine Reading of Text”, Ore Geology Reviews


Enkhsaikhan et al. (2018), “Towards geological knowledge discovery using vector-based semantic similarity”, Proceedings of the 14th Int. Conf. on Advanced Data Mining and Applications