

対話システムによる地質・施工情報の要約技術

鹿島建設(株) 正会員 ○戸邊勇人 升元一彦 松川剛一

1. はじめに

土木構造物を合理的かつ安全に施工するには、施工予定地周辺の地質とそれに起因する施工上の問題とを関連付け、その情報を迅速に現場に提供することが重要である。しかしながら、これを実行する地質技術者の人的資源が不足している。そのため筆者らは、この不足を補うための技術の一つとして、地質や施工に関する情報を文献から自動的に抽出するシステムのプロトタイプを作成してきた¹⁾。今回、その続きとして、このシステムに対話システムを追加することによって、文献から抽出されたキーワードの間に意味(アノテーション)を付与することを可能にした。この結果、地質・施工の文献から有効な情報を集約するシステムとすることができたため、その成果について報告する。

2. 開発したシステムの概要

地質技術者は、施工において問題が発生したときには、周辺の地質や施工記録などの文献から必要な情報を収集し、それに基づいて対策の検討を行っている。そして、文献から情報を収集する際には、重要なキーワードを抽出している。そのため、このキーワード抽出の作業を自動化することにより、地質技術者の負担が軽減できると考えられる。

文献からキーワードを自動抽出するには、近年のAI技術の発達によりテキストマイニングが使われてきている。本システムの開発においても、このテキストマイニング技術を取り入れ、下記の手順で文献からキーワードを抽出するとともに、キーワードから文献を検索(逆引き)できるシステムを構築した。

2.1 単語の抽出(形態素解析)

日本語は単語ごとに区切った書き方(分かち書き)をしないため、文章を解析するには、文を単語(形態素)に分解する必要がある。形態素解析を行うソフトウェアには多種のプログラムが存在しているが、本開発では無償かつ高速なMeCabを使用した。しかし、地質や施工の専門用語は、初期状態のMeCabでは正確に認識できないため、地学辞典²⁾と土木用語辞典³⁾から約25,000語を抽出し、辞書ファイルに追加した。

2.2 単語の出現頻度算出(文のベクトル化)

文献から抽出された単語を、出現頻度の表にまとめた。この表は、単語の種類を次元、出現頻度を要素とするベクトルと等価であるため、この表の作成は文のベクトル化と呼ばれている。表-1には3.7kmの本坑に複数切羽を採用し高速に施工したトンネルの施工文献⁴⁾をベクトル化した例を示す(上位10語を抜粋)。

2.3 共起ネットワーク解析とキーワード抽出

文献中の単語の重要性は、出現頻度、文中の位置、他の単語との関連性、および出現する文や単語の類似性などを基にして、多変量解析により算出した。文中で重要性の高い単語は、キーワードとして抽出した。この解析は共起解析と呼ばれており、解析結果は共起ネットワークと呼ばれる。表-1と同じ文献の共起ネットワーク解析例を図-2に示す。本システムではこの解析に無償かつ高速な解析が可能なKHCorder⁵⁾を用いた。

キーワード: テキストマイニング, 対話システム, 文献検索, 施工の合理化

連絡先 〒182-0036 東京都調布市飛田給2-19-1 鹿島建設(株)技術研究所 TEL042-485-1111

表-1 文のベクトル化

名詞	出現頻度
掘削	39
本坑	35
トンネル	33
切羽	30
採用	21
避難	21
施工	20
沈下	18
空洞	17
短縮	13

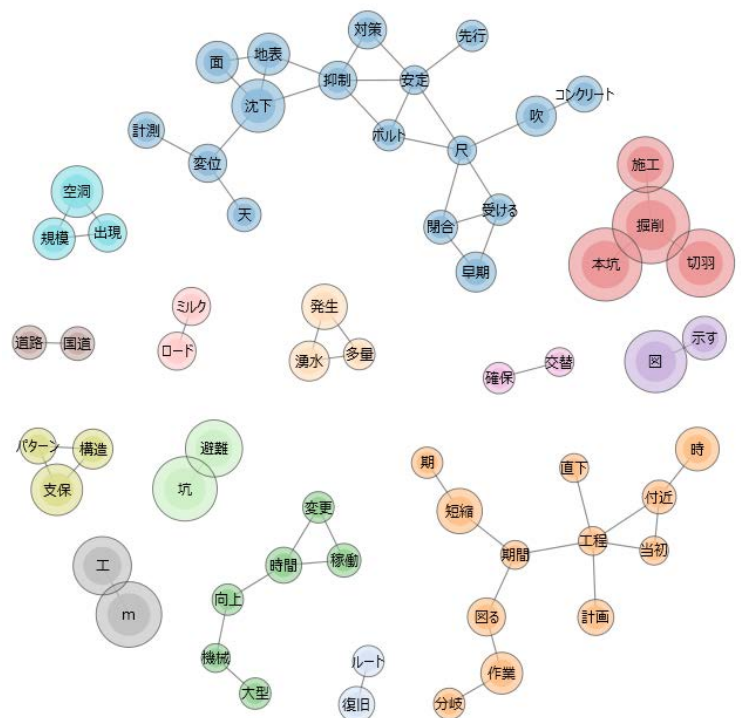


図-2 共起ネットワークの解析例

2.4 データベース化と文献検索

共起解析によって抽出されたキーワードを文献のタイトルとともにデータベース化することにより、キーワードを検索キーとして文献の検索（逆引き）ができる。また、検索の結果得られた文献から新たに別のキーワードを抽出し、それを検索キーとしてさらに別の文献を検索することができる。これにより、施工上の問題が発生した場合に関連性の高い多数の文献を即時に引用できるだけでなく、その文献に記述されている内容をキーワードに基づいて短時間で判断できるため、地質・施工情報の集約に必要な時間を削減できる。

3. 検索性能向上のための新機能の追加

3.1 新語登録とアノテーション付与

文献をデータベース化する場合には、抽出されたキーワードを羅列するだけでなく、新語登録とアノテーション付与を行うことにより検索の性能を向上することができる。

新語登録は、形態素解析エンジンに未登録の専門用語を追加する仕組みである。たとえば図-2では「避難」「坑」の各単語が関連性の高いものとして抽出されているが、これらは本来1単語の「避難坑」として扱うべきであるため、形態素解析にフィードバックし辞書ファイルに追加する。

アノテーション付与は、キーワード間の意味を付与する仕組みである。たとえば図-2では「空洞」「出現」（および「規模」）のキーワードが塊（クラスタ）で抽出されている。空洞は切羽に出現するものであるため、これらのキーワードの関係性（アノテーション）は「切羽に」とする（図-3）。これにより、データベースは「切羽に空洞が出現した文献」といった、より高度な検索が可能になる。

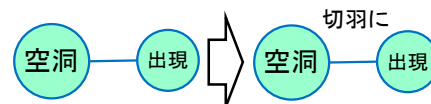


図-3 キーワード間のアノテーション付与

3.2 対話システムの導入による運用の効率化

アノテーションの付与をすべて手動で行うと1アノテーションあたり数分程度の時間を要する。また、アノテーションの自動付与技術は様々に研究されているものの、その多くはネットワークを用いたシステムを必要とする⁶⁾。施工情報には秘密保持を要するものが多く含まれているため、本システムはネットワークを介して外部に情報が流出しない仕様とした。そこで、対話システムを採用し、キーワード間のアノテーションをPCと地質技術者のチャットにより付与するシステムとした（図-4）。これにより、1アノテーションあたりの付与時間は10秒以下に短縮された。また、この対話システムでは前述の新語登録の機能も兼ねることにより、その登録時間も数分から10秒以下に短縮することができた（図-5）。

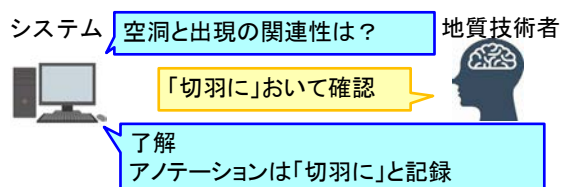


図-4 対話システムによるアノテーション付与の模式図

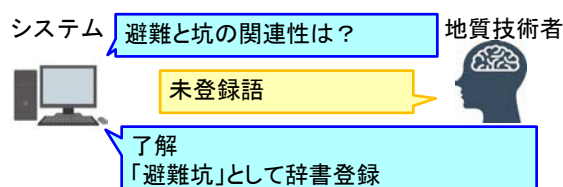


図-5 対話システムによる新語登録の模式図

4. おわりに

本稿では、地質・施工情報の自動集約システムの開発を進めるにあたり、その一部として対話システムによるアノテーションの付与機能の追加を行った。これにより、ネットワークを介することなく、文献から抽出したキーワードのアノテーションの付与を迅速に行うことが可能となった。しかしながら、アノテーションを完全に付与するには地質技術者との対話によるシステムの学習を進める必要がある。そのため今後は、システムの学習を進めることにより、より実用性の高い地質・施工情報の要約技術の開発を進めていく予定である。

参考文献

- 1) 松川ら：テキストマイニングによる地質・施工文献の定量解析とデータベース化,2021年度土木学会年次講演会講演論文集,2021.
- 2) 地学団体研究会：新版地学事典，平凡社，1443p，1996.
- 3) 土木用語辞典編集委員会：土木用語辞典，コロナ社，1421p，2000.
- 4) 廣瀬ら：3.7kmの本坑に複数切羽を採用し避難坑とともに20か月で掘削—国道57号二重峠トンネル，トンネルと地下，5月号，2020.
- 5) 樋口：社会調査のための計量テキスト分析 第2版，251p，2020.
- 6) 長野ら：人工知能による科学技術研究の加速，人工知能，34巻，6号，p783-789，2019.