

地質学研究における研究データ共有のための 地理情報データベースサービスの構築

高橋伸弥^{†,*} 奥村勝^{‡,*} 鶴田直之^{†,*}

福岡大学工学部[†] 福岡大学総合情報処理センター[‡] 福岡大学国際火山噴火史情報研究所*

1 はじめに

著者らが所属する国際火山噴火史情報研究所は、火山・地質学を専門とする自然科学の研究者らと、情報工学を専門とする工学の研究者らによる学際的な研究グループであり、フィールドワーク等で得られる火山地質情報の収集や蓄積に情報技術を活用し、多目的に活用可能な新たな火山地質データベースを構築することを目的としている^[1]。火山地質情報の中でも特に、露頭（野外において地層・岩石が露出している場所）の情報をデータベース化することは、研究者個人だけでなく防災やアウトリーチの面からも非常に重要であることから、我々はこれまでに、

(1) 特定多数の専門家により編集された一般向けの情報を公開するための噴火史情報 wiki ページ^a及び(2) 不特定多数の一般ユーザからの情報提供・共有を目的とした、地図情報サイト「じおログ」^bの2種のサイトを試験的に公開し、データベース構築の方針について検討を進めてきた。

露頭情報は本質的に非定型かつ自由度の高いデータであるため、編集者毎に露頭情報の表記や表現が異なることも多々ある。また、それらを整理、構築するには専門家による手作業を必要とし、膨大な人的・時間的コストを要することとなる。これまでも露頭データベース構築の試みはなされてきた^{[2][3]}が、これらの課題の解決には至っていない。この問題に対し、利用者がデータに付与した任意のタグを用いて自動分類することによりデータベース作成を省力化することを考える。このタグによるデータベース作成手法の有効性を検討するため、本研究では、地質学研究室における卒業研究等の関連データを対象としたデータ管理支援システムを構築し、実際の運用を通して課題を明らかにすることを目指す。

2 メタ情報によるデータ管理

従来のデータベース構築においては、専門的な知識を持った編集者が人手により事前にデータを分類・整理し、さらに必要とされるデータ項目を洗い出した後、それらのフォーマットに従って手作業で入力をするといった膨大な労力が必要とされていた。これに対し、タグと呼ばれるメタ情報によるデータ管理では、自由に任意のキーワードを付加することができるので、データ項目の内容や項目数、フォーマット等に頭を悩ませることがなく、さらにはタグに基づくデータ間の類似度計算により自動分類が可能となるため、事前の分類・整理が不要といったメリットがある。この手法の特徴として、データを従来型の階層構造ではなく、フラットな構造として管理する点があげられる。逆に言えば、物理的なデータ管理の構造を考慮する必要がないことから、分散管理にも向いている。

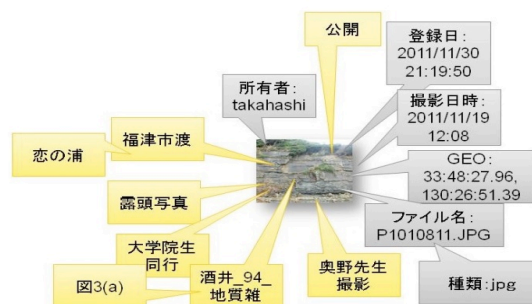


図1 露頭画像へのタグ付けの例

ファイルの種類や位置情報など、機械的に処理できる情報は自動的にタグ付けすることとし、所有者やアクセスレベル、登録日時、更新日時は隠しタグとして保持する(図1)。さらには、入力されたコメント等の文章からも自動的に適切なキーワードを抽出することも可能である。

3 地質学研究室向けデータ管理支援システム

はじめに述べたように、上述のタグ付けによる

データ管理・データベース作成手法の試行のため、地質学研究における研究データを共有管理するシステムを試作した。地質学研究室において、卒業研究等でまとめられる研究資料は膨大なものとなることから、それらのデータの散逸を防ぎ、かつ再利用を円滑にするようなデータ管理手法の確立は重要な課題となる。またデータの種類としても、写真や画像だけではなく各種ドキュメントファイルや測定結果の生データなど多岐にわたる。本稿では、これらの点を考慮したデータ管理支援システムを提案し、実際の運用を通して新たな課題や必要な機能を検討する。

まずは、基本機能を組み込んだプロトタイプシステムを実装し、β版として2014年から試用を開始した。組み込んだ機能のうち主なものとしては以下の4つがある。

- ① タグによるファイル管理機能 (図2)
- ② 各種ファイルを一括管理する機能 (図3)
- ③ 直感的な操作によるタグ検索機能 (図2)
- ④ 位置情報との紐付け機能 (図3, 4)

システムは Web アプリとして提供することとし、PHP で開発した。フレームワークには CakePHP2 を用いている。DBMS には MySQL を使用しているが、次期バージョンでは、MongoDB 上を用いて構築した既存のシステム^[4]と連携する予定である。

地質学研究室の関係者ら6名を対象に2014年9月からの4ヶ月間を試用期間として収集したデータ総数は892個となった。うち画像が630個であり、その他 docx/xlsx 等のファイルが262個であった。また作成されたページ数は95となり1ページあたりの平均データ数は9.4となった。一方タグに関しては、総数が138種類となり、かなりの重複が見られた。使用頻度の最大値は90回で、10回以上使用されたタグは18種であった。逆に3回以下の使用頻度のタグは13種であった。

使用者からの要望としては、タグの表記の揺れに対処してほしい、タグ入力の際に補完機能があるとよい、経緯度から地名タグを自動でつけられないか、といったタグ入力に関するものが多く挙げられた。今後の改善項目として検討したい。

4 おわりに

本稿では、タグによるデータ管理手法の有効性を検討することを目的として、地質学研究室における卒業研究等のデータを対象としたデータ管理支援システムを提案した。今後は、実際の運用を通して問題点を洗い出し、機能の追加・改善を図る予定である。

参考文献

- [1] 鶴田他, “情報科学から活火山データベースを考える”, 月刊地球, Vol. 34, pp. 277-280, 2012.
- [2] 横田, “露頭データベース作成は何故困難か?”, 情報地質, Vol. 7, No. 4, pp. 297-301, 1996.
- [3] 勝野 *et. al.*, “参加型 WebGIS を活用した露頭情報システムの構築例,” 情報地質, Vol. 13, No. 2, pp. 74-75, 2002.
- [4] 奥村 *et. al.*, “データ活用のためのメタ情報を考慮した地理情報システム向けデータベースの提案,” 地理情報システム学会第22回学術研究発表大会, 2013.

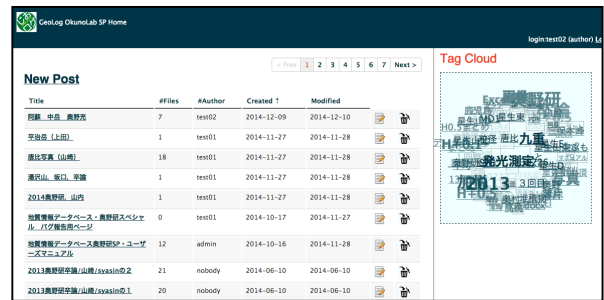


図2 タグクラウドによるファイル管理機能

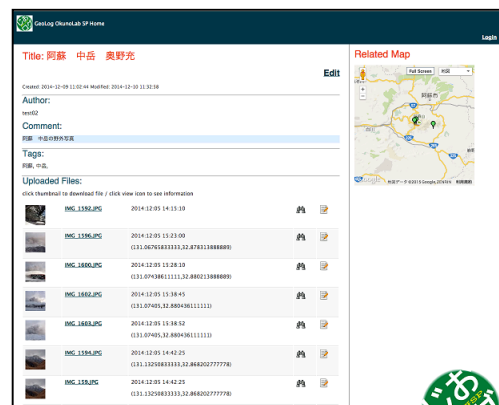


図3 各種ファイルの一括管理機能

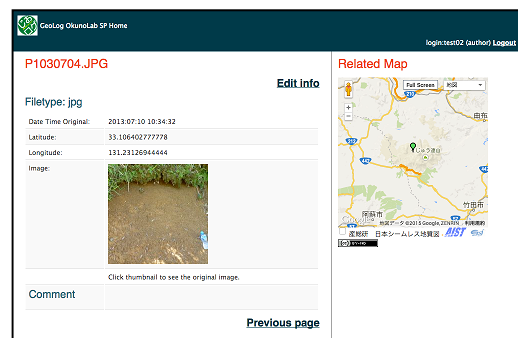


図4 位置情報との紐付け機能

謝辞

本研究は JSPS 科研費 26350410 の助成を受けたものである。

- a www.acrifis-ehai.fukuoka-u.ac.jp/mediawiki/
- b www.acrifis-ehai.fukuoka-u.ac.jp/geolog/