2017年8月18日

奈良文化財研究所

都道府県別の発掘調査報告書特徴語ワードマップ等の公開

考古学ビッグデータの応用事例：地域ごとの発掘調査成果の特徴を１枚の画像で表現すると？

１、公開機能の概要

（1）機能公開

日　時：2017 年8月22日（火）より公開

ＵＲＬ（全国遺跡報告総覧）：http://sitereports.nabunken.go.jp/ja

（2）機能の概要

全国遺跡報告総覧において、考古学ビッグデータの特性を活かした４つの新機能を公開することで、発掘調査報告書（以下、「発掘報告書」という。）のアクセシビリティをさらに向上させました。

①　【都道府県別の発掘調査報告書特徴語ワードマップ】

全国遺跡報告総覧（以下、「総覧」という。）に登録されている約 20,000 件の発掘報告書（テキストデータ 15 億文字）に対し、都道府県ごとに考古学関係用語の特徴語を、図化しました。当該都道府県の発掘報告書内において頻出頻度が高くかつ他の都道府県では出現頻度が低い用語を加味した上でマップを生成しているため、<u>当該地域の強い特徴を示す用語を可視化</u>できました。自然言語処理技術のベクトル空間モデルの TF（索引語頻度）と IDF（逆文書頻度）を組み合わせた TF-IDF にて算出しました。

②　【類似報告書の自動表示機能】

発掘報告書毎の頻出用語を整理することで、ある報告書と内容が類似している報告書を自動表示します。

③　【検索キーワードの入力補助機能（簡易版）】

全文検索時のキーワード入力補助機能を実装しました。

○もしかして機能：入力したキーワードに間違いが含まれている可能性がある場合、自動的に正しいと思われるキーワードを表示します。最低２文字で機能します。

○サジェスト機能：入力したキーワードと類似の用語を自動表示します。総覧搭載の用語辞書をベースに部分一致したものを表示します。

④　【都道府県別発行機関の刊行物登録状況】

都道府県毎に地方公共団体（都道府県・市町村）・法人調査組織の登録状況に応じて、日本地図の都道府県部分を濃淡表示します。

２、経緯と期待される効果

【経緯（現状の課題）】

・日本では<u>埋蔵文化財行政の着実な推進によって、膨大な発掘報告書が発行され、長年にわたる日本考古学の研究蓄積</u>があります。総覧においても 2017 年 8 月 14 日時点で、20,437 冊の資料、文字数約 15 億字、総ページ数約 250 万ページのデータ量があります。考古学は調査事例の蓄積によって研究が進展する蓄積型の学問です。「考古学ビッグデータ」というべき新たな可能性がある一方、<u>情報過多によって処理不能となる「情報爆発」という一面</u>があります。既に推定十数万冊の報告書が発行されており、さらに年間約 1,700 冊程度の発掘報告書が発行されているため、もはや<u>人間がすべての発掘報告書を実際に閲覧し、内容を把</u>

握することは困難な状況となっています。

・発掘報告書は、「将来にわたって保存されるとともに、広く公開されて、国民が共有し、活用できる」（『発掘調査のてびき－整理・報告書編－』）ことが求められています。しかし、一般市民や初学者の方が発掘報告書を手軽に検索し、その成果を活用するにはハードルがありました。検索には、専門用語の入力が必要であり、まず専門用語を事前に学習し知っておく必要があります。そのため専門用語を知らなくとも情報検索できるようにできれば、ハードルを下げることができます。

【期待される効果】

① 【都道府県別の発掘調査報告書特徴語ワードマップ】

考古学的成果（考古学用語）について、地域ごとの特徴を可視化でき、地域性を理解する手がかりにできます。考古学用語は地域によって表記ゆれがあります。特徴語ワードマップにより、表記ゆれを効率的に把握できます。これを総覧の考古学関係用語シソーラスに反映させることで、漏れのない検索ができるようになります。

② 【類似報告書の自動表示機能】

考古学では類例調査が重要です。ある報告書を閲覧した時に、内容の近い報告書が自動表示されることで、効率的で精度の高い類例調査が可能となります。専門家にとっても自分が把握できていなかった報告書を知る機会の一つとなります。

③ 【検索キーワードの入力補助機能（簡易版）】

専門的な考古学用語の一部を知っているだけで、検索できるようになります。一部入力間違いの場合でも候補が表示されることで、検索のハードルが下がります。

④ 【都道府県別発行機関の刊行物登録状況】

都道府県毎の登録状況を視覚的に確認できます。

【参考情報】

・全国遺跡報告総覧とは

全国遺跡報告総覧は、埋蔵文化財の発掘調査報告書を全文電子化して、インターネット上で検索・閲覧できるようにした“報告書のインデックス”です。「総覧」は、全国遺跡資料リポジトリ・プロジェクトによって構築された遺跡資料リポジトリ・システムとコンテンツを国立文化財機構 奈良文化財研究所が引き継ぎ、運用しているものです。

貴重な学術資料でありながら、流通範囲が限られ一般に利用しづらい報告書をインターネット上で公開することで、必要とする人が誰でも手軽に調査・研究や教育に利用できる環境の構築を目指しています。

現在、全国の大学や自治体等の 381 機関の報告書類 20,437 件を収録しています（2017 年 8 月 14 日時点）。

お問い合わせ先　奈良文化財研究所企画調整部　高田祐一　℡ 0742-30-6711　Mail ytakata@nabunken.go.jp

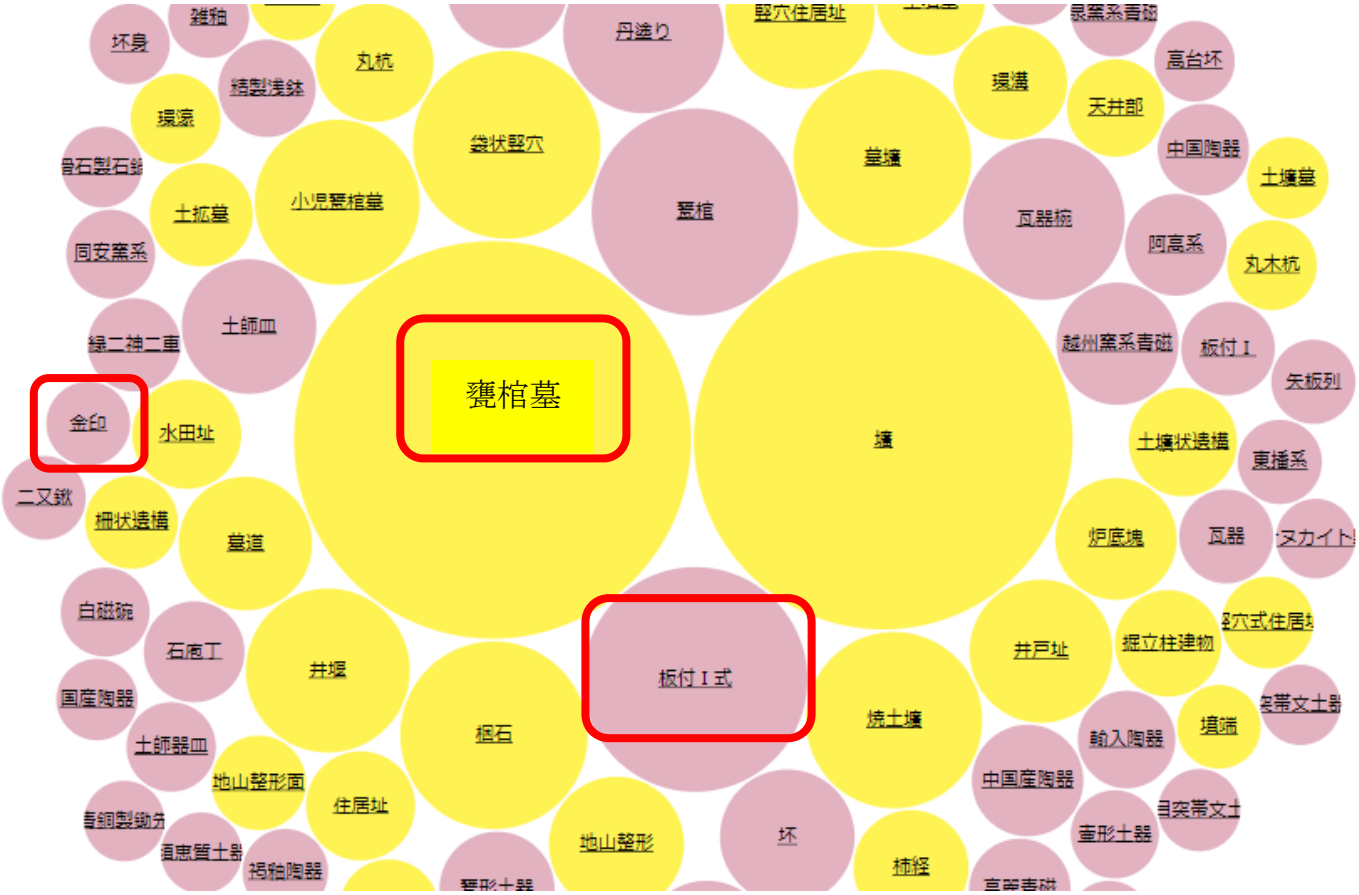
① 【都道府県別の発掘調査報告書特徴語ワードマップ】

僕の住んでいる地域では考古学的にどんな特徴があるんだろう？？
でも確認しようにも報告書がたくさんあって読めないよ！

ある地域には多くても他の地域には少ない考古学用語は自然言語処理で可視化できたよ。図は福岡県の例で、北九州に多い甕棺墓が特徴語になっているね。

【報告書特徴語ワードマップ-福岡県】（開発中画面）

http://sitereports.nabunken.go.jp/ja/visualization/term/pref/40



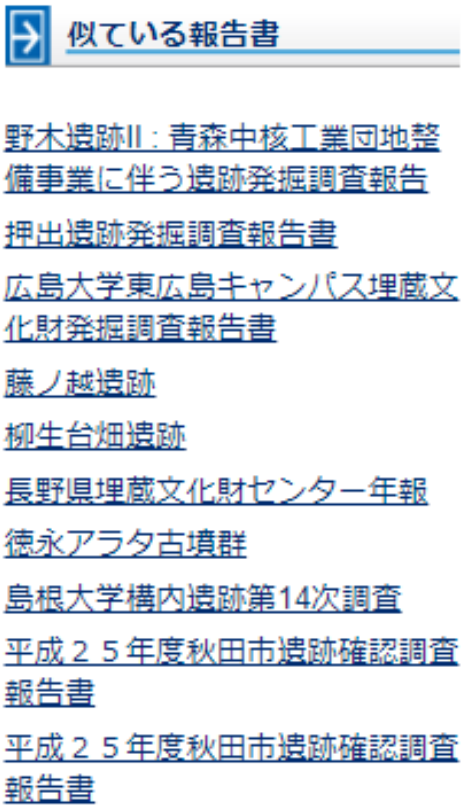
特徴をおよそつかめるね。なぜこの考古学用語が多いのか、背景を調べたくなったよ！〇〇についてもっと調べよう！

② 【類似報告書の自動表示機能】

僕は〇〇について調べているんだ。
でも類例を確認しようにも報告書がたくさんあってどれを読めばいいのかわからないよ！

〇〇を掲載している報告書と内容が似ている報告書は、自然言語処理で、自動表示できるよ。まずは似ている報告書をあたれば効率的に類例収集できるね。

【似ている報告書例】（開発中画面）
報告書詳細ページの右サイドバーに表示されます。



僕が知らなかった報告書が表示されているぞ。類例調査の漏れが減るかもね！

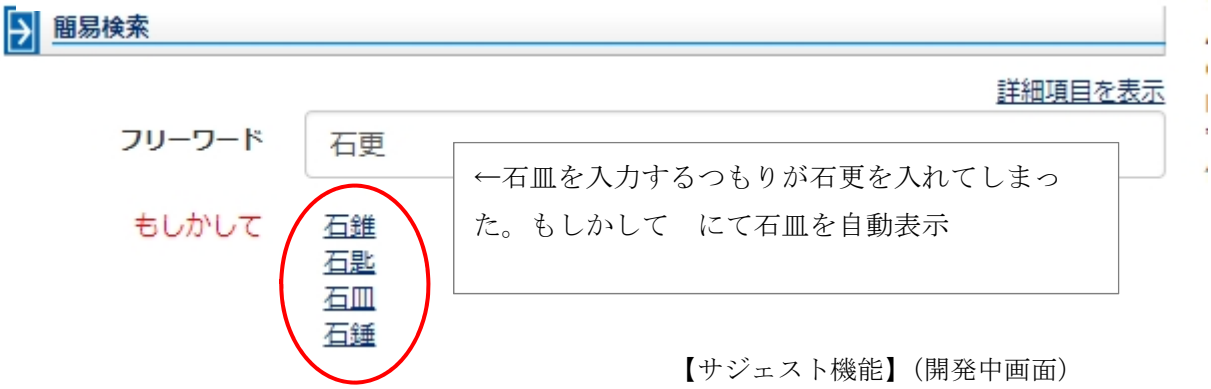
③　　【検索キーワードの入力補助機能（簡易版）】

○○について興味があるから、少しだけ用語はわかるよ。でも専門的な考古学用語を正しく覚えておかないと、検索できない。報告書がたくさんあってもたどり着けないと意味がないよ！

曖昧に用語を入力しても、自然言語処理で候補を自動表示できるよ。

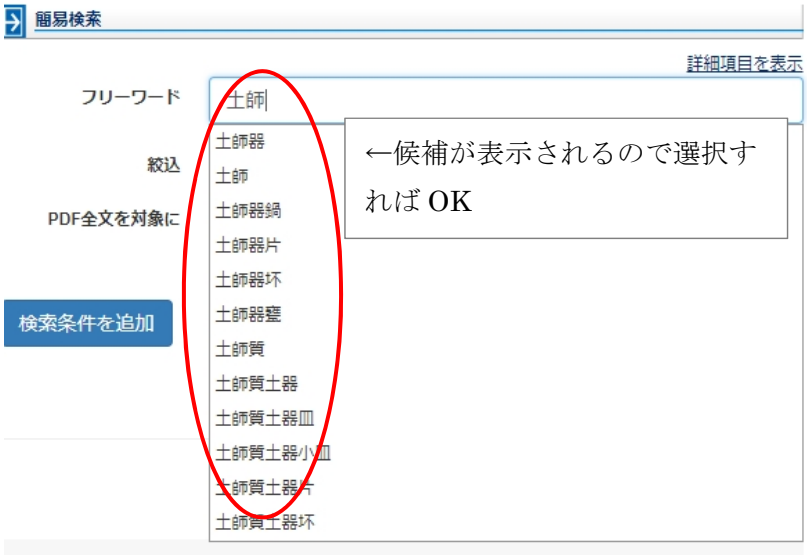
【もしかして機能】（開発中画面）

１字入力間違いをしても、類似の考古学用語を表示します。



【サジェスト機能】（開発中画面）

入力した用語と部分一致する用語を表示します。



用語の入力を一部間違っても、候補を表示してくれるから、確認したい報告書を探しやすくなったよ。また新しい□□という用語も知ったよ。○○に関連していそうな□□についても調べてみよう！