東京大学地震研究所技術研究報告，No. 23，10–13 頁，2017 年.
Technical Research Report, Earthquake Research Institute, the University of Tokyo, No. 23, pp. 10–13, 2017.

報　告

全文検索に対応した日本歴史地震データベース検索システムの紹介

パンタ　ボーラ*†・鶴岡　弘*・佐竹健治*

Introduction to the Full-Text Search System for Historical Earthquake Database

Bhola PANTA*†, Hiroshi TSURUOKA* and Kenji SATAKE*

1. はじめに

地震研究所では，地震・火山に関する理学・工学的な観測・研究を行うだけでなく，東京大学史料編纂所（以後史料編纂所）と連携し，近代以前に発生した地震・火山噴火等の自然災害に関する史料のデータベース化が始められている．地震については史料による歴史地震の研究が 1 世紀以上にわたって実施されており（西山，2015），その膨大なデータが蓄積されつつある．このような文理融合研究をさらに拡張し，奈良文化財研究所（以後奈文研）と連携し，「発掘調査現場で見つかった災害の痕跡等を含め，近代的な観測データが整う以前の地震や火山活動にともなう情報」（村田，2014）をも横断的に検索できるシステムの構築が要求されている．具体的には，史料の全文と写真等の画像データを有機的に結合できる基礎システムの開発が必要となっていた．システム開発にあたり，歴史時代の災害データのコンパイル作業を進めると同時に，データ公開作業の一環として，全文検索に対応した検索システムのプロトタイプを作成したので報告する．

2. システム全体構成

本検索システムの設計は，どのデータベースエンジンを採用するのかの検討から始めた．今回は，オープンソースで無料のデータベースエンジンを採用することにした．検討したのは PostgreSQL と MySQL である．両方とも高い性能と豊富な機能を持っているオブジェクト関係データベース管理システム（ORDBMS）で，無料で使えるデファクトスタンダードである．ただし，MySQL は Oracle に買収されて，現時点ではオープンソースでなくなった．MySQL 派生として開発されているオープンソースの ORDBMS もあるが，本検索システムには両条件が揃っている PostgreSQL を採用することにした．

本システムの特徴の一つは日本語の「全文検索」機能である．しかし，PostgreSQL の標準機能の全文検索はアルファベットと数値だけに対応していて，日本語や中国語などのマルチバイト文字はサポートしていない．そこで，オープン系の日本国産の PGroonga というソフトを採用することにした．PGroonga は村川（2017）による災害記事の全文検索にも使われており実績があるためである．PGroonga を PostgreSQL にインストールすると全言語対応の超高速全文検索機能を使えるようになる．PGRoonga は，PostgreSQL の機能を拡張するエクステンションであり，別途のコンパイルまたはバイナリファイルでのインストールが必要であるが，PostgreSQL との親和性が高いので操作性やシステムのメンテナンス性に優れている．図 1 にシステム全体構成の概要を示した．テキストデータである史料編纂データをもとに，PGroonga により高速な全文検索を可能とするインデックス作成を行い，災害情報および画像とともにデータベースに格納した．ユーザは，これらのデータが有機的に結合されたデータベースに全文検索エンジンを用いてアクセスする．なお，ユーザは Web ブラウザーを用いたインタフェースにより，任意のキーワードや地震・火山のイベント発生期間などを指定した高度な検索をネットワーク経由で行うことができる．それぞれの部分的な説明は，3 章のシステム機能において詳細に記述する．

3. 開発したシステムの機能と特徴

本検索システムの検索対象は，XML 形式の構造化テキストデータである．近代以前に発生した地震・火山噴火に

2017 年 9 月 29 日受付，2017 年 11 月 24 日受理.
† panta@eri.u-tokyo.ac.jp
* 東京大学地震研究所地震火山情報センター
* Earthquake and Volcano Information Center, Earthquake Research Institute, the University of Tokyo.

関する史料データは史料編纂所と奈文研が独自に管理・所有しているため，それぞれの機関から XML で記述されたテキストファイルおよび PNG などの画像ファイルをコピーし，地震研内におかれたサーバーに保存し，それらを検索対象とした．つまり，データの複製を作成するため，データ保存としては冗長となってしまうが，バックアップとなることと，集中化されたデータにアクセスするので，高速な検索が可能になるというメリットがある．まずは，全文テキストや画像のファイルを適切に配置した任意のフォルダーに置き，Python で書かれたプログラムを実行することによって，PostgreSQL データベースにデータを格納した．史料編纂所データ（XML 形式）の解釈（パーサーと呼ばれる）には Python の標準ライブラリを使用した．さらに，歴史災害データの PostgreSQL 用テーブルを設計し，インデックス作成用フィールド定義を行い，パーサー処理後，そのフィールドに対するインデックスが作成され

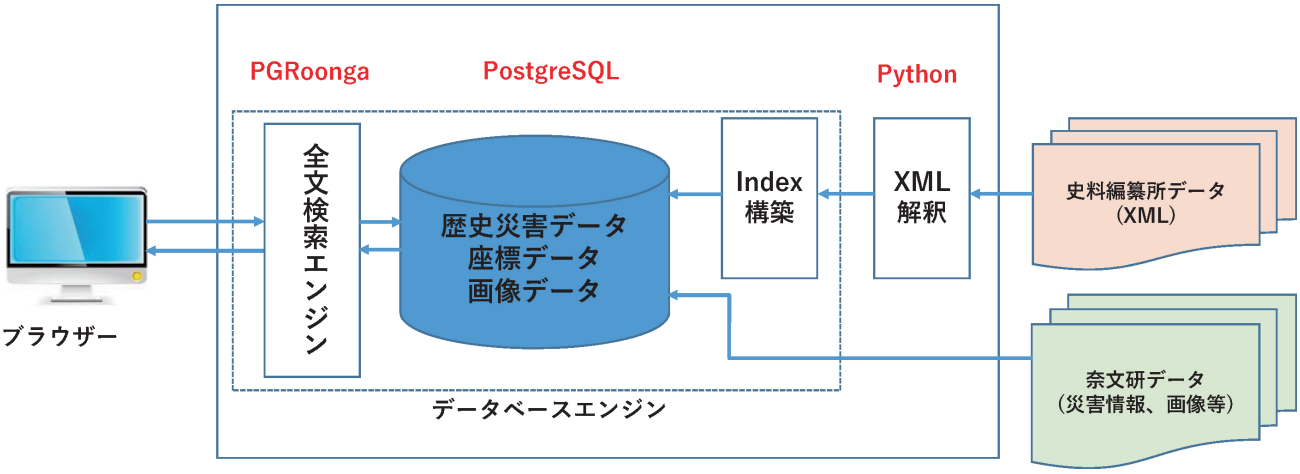

図 1.　システム構成の概要

表 1.　本システムで使用したソフトウェアの仕様（2017 年 9 月 27 日現在）

項目	ソフトウェア	バージョン，URL 等
ブラウザー	Chrome	バージョン：61.0.3163.100（64 bit）
	Firefox	バージョン：55.0.3（32 bit）
	IE	バージョン：11.608（32 bit）
OS	CentOS	リリース 6.9 バージョン：2.6.32-696.6.3.el6.x86_64 https://www.centos.org
検索エンジン全体	JavaScript	各ブラウザーの最新プラグイン
	PHP	PHP　バージョン：5.3 http://php.net/downloads.php
	データベース	PostgreSQL バージョン：9.6 https://www.postgresql.jp/
	全文検索 エンジン	PGRoonga 1.20 https://github.com/pgroonga/pgroonga
マップ API	Leaflet	バージョン：1.2.0 http://leafletjs.com
ベースマップ	地理院地図	https://maps.gsi.go.jp /development/ichiran.html
DB 管理ツール	PhpPgAdmin	バージョン：5.1 http:// phppgadmin.sourceforge.net/

る．全文検索にはこのフィールドを使用するので，高速な検索が可能となる仕組みとなっている．

本システムの特徴は，すべてオープンソースで無料のソフトウェアを使用し，簡単に日本語の「全文検索」に対応した検索システムを構築できたということである．本システムの運用は，Linux（CentOS サーバー）上で行うこととした．CentOS は，RedHat 互換の無料の OS であり，セキュリティパッチの提供も行われているため，データベースの開発・運用には問題ないと考えた．実際のプロトタイプシステムの開発及び運用は Dell のラップトップ PC（Precision M4600）で行った．

検索処理でのレコードの取得および表示には Web プログラミングでは広く使われている JavaScript と PHP 言語を用いた．地図表示には地理院地図と表示用ライブラリとして Leaflet を採用した．Leaflet は近年（2011 年以降），広く使われている Web 地図のための JavaScript ライブラリであり，スマートフォンなどのモバイル端末やデスクトップ端末のプラットフォームのほとんどに対応しており，HTML5 と CSS3 に対応している．Google Map のようなものであるが，異なる技術で地図表示ができるツールである．PostgreSQL や PGRoonga の設定は，オンラインチュートリアル等参考資料を参照し，設定を実施した．プロトタイプ用 OS，ソフトウェアやブラウザーの仕様を表 1 に記述し，図 2 ～ 5 にユーザ操作による検索結果を表示した例を示す．

4.　開発したシステムの使い方

本システムでは，Web 上の操作で指定した URL にアクセスすることにより検索処理を行うことができる．検索条件を入力するインタフェースは，簡易検索画面（図 2）と複数の条件を入力する複合検索画面（図 3）を用意した．簡易検索画面で二つの単語が入力された場合，OR 検索または AND 検索をできるようにしている．表示された一覧結果（図 2，3）において，リンクされたキー項目をクリックすることでそのレコードの詳細情報が表示される（図 4）．災害が発生した地名がある場合，その項目をクリックすることにより地図が表示される（図 5）．さらに，座標マーカーとマーカーへのクリック操作により絵図等の付属情報を表示することが簡単にできるようになっている．

5.　まとめ

本報告では，歴史地震資料（XML 化されたテキスト）および絵図等の画像データの登録機能，全文検索エンジン，検索閲覧インタフェースなどを持つ歴史地震データベース検索システムのプロトタイプの設計，構成，機能や操作について紹介した．本システムが取り扱うデータベースは史料編纂所および奈文研で開発中のため，本検索システムの本格運用に向けては，各研究所とさらに連携しながら取り扱うデータの仕様や関連する各種のデータやファイルなどを一元的に管理するリポジトリへの格納方法の自動化などを検討・調整する必要がある．

図 2.　キーワードを指定した簡易検索の画面例

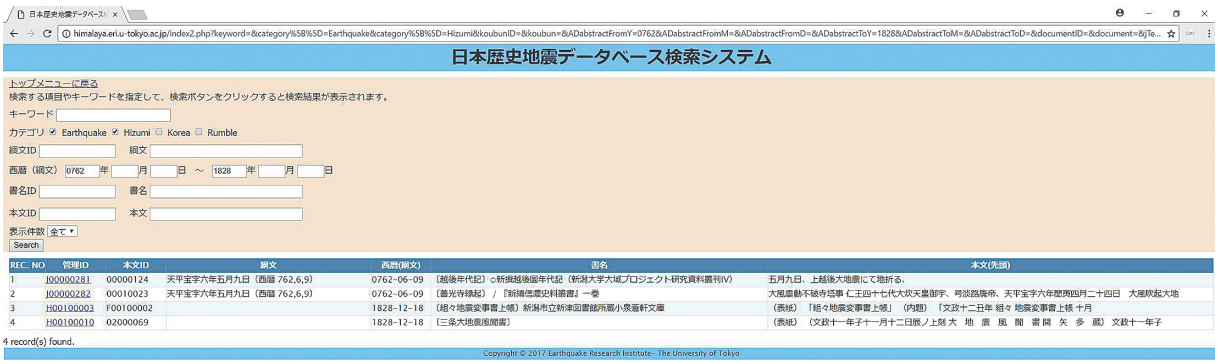

図 3.　複数の条件を指定して検索した画面例

図 4.　検索結果の詳細情報表示例

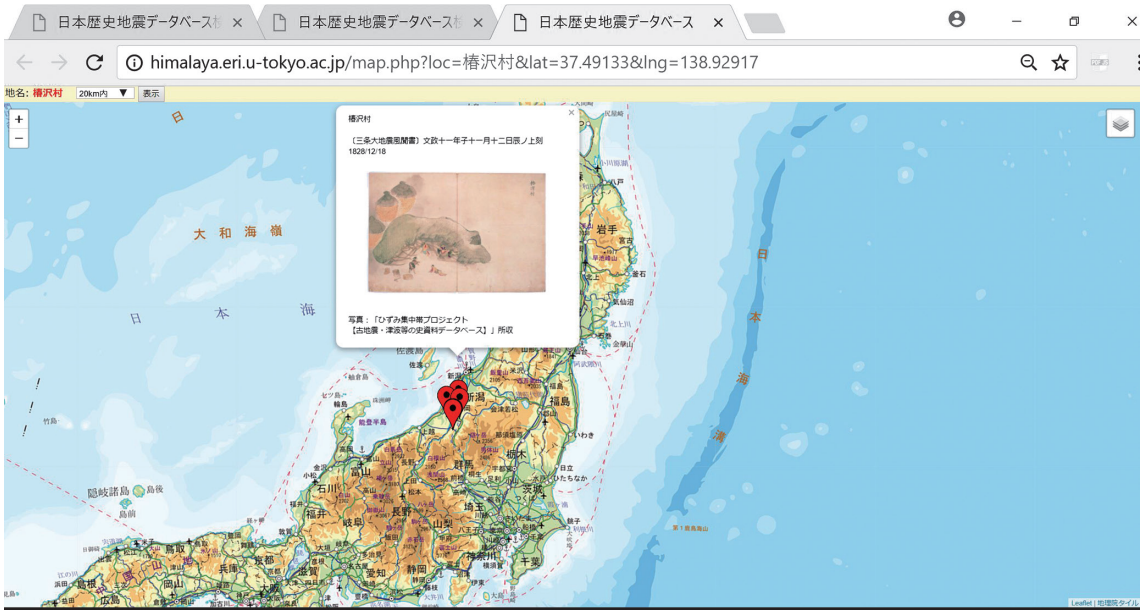

図 5.　位置情報により，地図および画像データ等を表示した画面例

謝　辞：奈良文化財研究所埋蔵文化財センター遺跡・調査技術研究室の村田泰輔アソシエイトフェロー，史料編纂所の榎原雅治教授，同所の山田大造助教および地震研究所地震予知研究センターの西山 昭仁助教に有益なご指摘を頂きました．また査読者の岩崎貴哉教授と飯高隆准教授には本稿を改善するうえで有益なご指摘を頂きました．ここに記して感謝申し上げます．

文　　献

西山昭仁，2015，史料を用いた歴史地震の研究，地震調査研究推進本部，コラム，http://www.jishin.go.jp/resource/column/column15win_p10/，（参照 2017 年 9 月 27 日）．

村田泰輔，2014，平城第 530 次発掘調査で発見された巨大地震の痕跡，奈文研ニュース，55，http://repository.nabunken.go.jp/dspace/bitstream/11177/2526/1/AA11581556-55-1t.pdf，（参照 2017 年 9 月 25 日）．

村川猛彦，2017，災害記事データベースの構築および応用―記事収集，全文検索，およびテキスト分析―，和歌山大学災害科学教育研究センター研究報告，**1**，1，http://www.wakayama-u.ac.jp/bousai/kiyou/number/2017033000172/files/C_020_murakawa_20170216.pdf，（参照 2017 年 9 月 25 日）．

PostgreSQL/PGRoonga チュートリアル，https://pgroonga.github.io/ja/tutorial/，（参照 2017 年 6 月 1 日）．