

統合型防災情報提供サイトの構築について

企画情報技術グループ 松浦秀起

1. はじめに

京都大学防災研究所は、創立以来、日本における自然災害研究の中心的役割を担うと共に、災害科学研究者による研究ネットワーク形成を主導してきた。そして、研究の一環として年報を始め、論文やハザードマップなどの印刷物と実験・観測・調査など記録（記録紙、映像を含む）等が、防災研究基礎資料として蓄積されてきている。このような膨大な資料を一般に公開し、分かり易い形で提供することは、防災学研究を推進させ、防災資料の利用者にとって生きた防災情報源として付加価値を高めることになると考えられる。平成14年度に21世紀COE研究分担課題「防災研究所で蓄積された印刷物や映像情報の電子ファイル化とホームページで高速検索可能なシステムの構築」を技術室が担当することになった。その成果として、防災研究所が有している防災資料を利用者にとって、よりわかりやすい形で提供できるデータベース検索システム（以下、防災情報検索システム）を構築した。その後、防災情報検索システムに改良を加え、防災研究所ホームページにて試験公開している。

本報告では、防災情報システムの現在の機能と、システム改善すべき点を示し、今後目指すべき統合型防災情報提供サイトの構築案について報告する。

2. 現在の防災情報検索システムについて

図1のように、現在の防災情報検索システムは「タイトル」、「著作者」、「フリーワード」、「年代」、データタイプ（ダブリンコアに準じた素材の種類）、データカテゴリ（独自で定めた防災情報のグループ分け）を指定して、登録されている多数の防災情報から絞り込むことができる。例えば、著作者欄に河田、フリーワード欄に津波を入力して、検索ボタンを押すと、図2のように検索結果が下に一覧表示される。そして、その検索結果の中の詳細を閲覧したい場合は、詳細表示ボタン（別ウィンドウか右ページかを指定できる）を押すと図2のように、そのデータの詳細が閲覧できるようになっている。

システム構成は、PHP+MYSQL（RDBMS）であるが、実際にDBに登録されているのは、防災情報データそのものではなく、防災情報のメタデータであり、そのメタデータを通じて、防災情報にアクセスできる仕組みになっている。様々な防災情報を検索できるように設計しているため、格納できるデータ項目は80フィールド以上であるが、登録するために必要な必須入力データ項目は10フィールドであり、汎用性を持たせるようにしている。また、年報など既にホームページ上で公開されている防災情報については、このシステム内でダウンロードまで行うことができるようになっている。

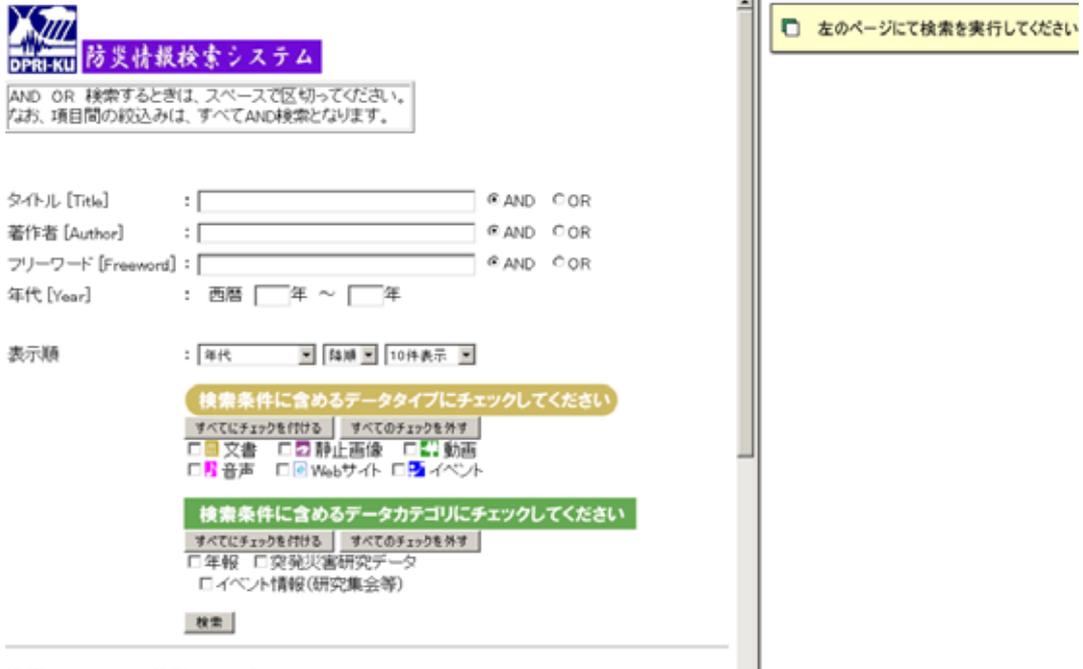


図 1 防災情報検索システムのトップ画面



図 2 防災情報検索システムにおけるデータ検索の様子

3. 防災情報検索システムの特徴

(1) NDC (日本十進分類法) の適用

NDC は図書館で図書分類に使用される日本十進分類法であるが、目的は主題別に図書を区分するために用いられる。通常の論文に NDC を使用することは、あまり例がなかったが、最近になって、NII (国立情報学研究所) が進めている IRP (学術機関リポジトリ構築連携支援事業) における各大学の学術リポジトリには NDC が用いられている例がある。IRP とは、平成 17 年度から NII が中心となって、各大学における学術的な論文などの研究成果を無料で閲覧できるように学術情報の収集、組織化、保存及び発信のための仕組みを構築する事業であり、現在は 70 大学の図書館において、大学内の学術情報の収集、および検索システムへの登録を積極的に行っている。

このように、元々は図書の分類をするために使用されてきた NDC だが、論文の主題を分類するために用いることも可能であることが実証されている。防災情報検索システムに登録されているデータ中で、最も多いカテゴリは年報である。年報とは、1957 年に創刊以来、毎年発行されている防災研究所の研究成果を報告する冊子のことである。内容や形式も学会誌に近く、非常に学術的な論文が投稿されている研究成果報告集的な位置づけである。防災研究所は、IRP に先駆けて、この年報の中の 1 つ 1 つの論文に NDC を付与している。これにより、防災情報検索システムにおいて NDC が付与された防災情報は、主題により分類することが可能であり、また防災情報の主題をキーとした新たな検索手法が可能となっている。ただし、現在はまだ主題を使用した検索機能は実装されていない。

(2) 他の検索システムとの連携

防災情報検索システムは、他の独立した検索システムにて、関連する情報を追加取得できるように、自動リンクする機能を持っている。現在、自動リンクを行っている検索システムは、防災研究所職員総覧、京都大学学術情報リポジトリ KURENAI、京都大学図書館蔵書検索 KULINE である。例えば、検索結果リストの著作者名の中に、検索した時点で着任している教員の名前があった場合、クリックするだけで、職員総覧の教員情報を閲覧できたり (図 3)、検索した防災情報に関する関連書籍や論文の一覧表を閲覧できたりする。

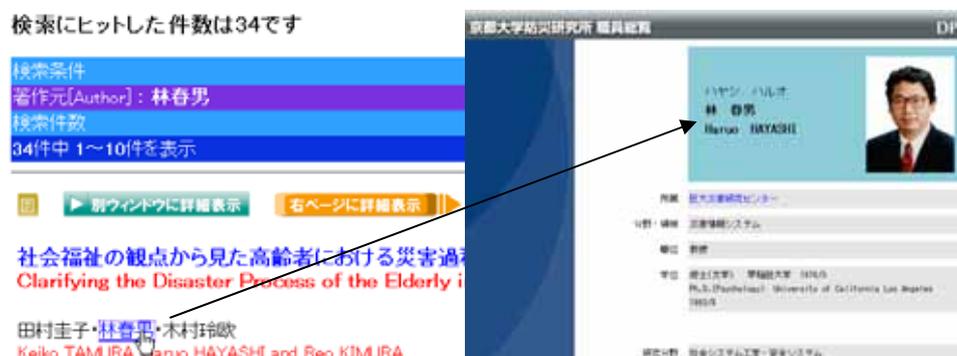


図 3 防災情報検索システムと職員総覧との連携

4. 防災情報検索システムの改善すべき点

平成 17 年度に防災情報検索システムの基盤となるプログラムが完成以降、少しずつ改良を重ねていったが、プログラムの煩雑化に伴い、データベース構造やデータ登録などの運営面で大幅な見直しを行う必要があることが分かっている。そのため、当初予定していた機能追加もその見直し後に行うようにする予定である。以下に改善すべき点をまとめる。

(1) コンテンツ管理手法

防災情報コンテンツの中でもドキュメントはタイトルや著者がはっきりと判別がつき、ファイル名もある程度の規則に従った命名が可能であるが、災害調査報告で撮影された写真や動画などは、同じ時期、同じ場所、同じ著作者で作成されたコンテンツが多数存在し、タイトルや著作者、整理されたファイル名などメタデータ作成において、非常に時間と手間がかかることが分かっている。現時点での対処法は、関連する日時や場所や災害などの名前を管理者側が設定することで処理しているが、今後も増大する資料の整理方法、システムへの効率的な登録手法を確立し、データベース構造などもそれに合わせて変更する必要があると考えている。

また現在のデータベース内のテキストデータの文字コードは ujis である。プログラムもそれに合わせてコーディングされているため、ごく一部ではあるが漢字が文字化けを起こしてしまう。その対策として、近年になってようやく主流になってきた文字コード utf-8 を採用し、すべてのデータを utf-8 で統一し、プログラムも一から見直す必要がある。

(2) 検索結果のランキング付け

現在の検索結果は、年代、タイトル、著作者の 3 つをキーにして、ソートできるようになっている。しかしながら、そのソートは数字の大小、もしくは、あいうえお順であり、Google や Yahoo にみられる検索語の重要度や頻度などを重視したソートではない。本当に使いやすい検索システムを目指すのであれば、tf-idf などを用いたキーワードの重み順にソートした結果で一覧表示するほうが望ましいと考える。ただし、tf-idf を使用する場合は問題点もある。それは、新規コンテンツが追加・削除されるたびに tf-idf の値は変動するため、そのたびに形態素解析や再計算を行わなければならない点である。

(3) 運営コスト

現在の防災情報検索システムのデータベースへの入力、現在すべて手動である。データ入力はデータベース構造に合わせたデータセットを作成し、ローカルから直接追加しているため、データベースへの入力作業が非常に煩雑になり、入力が少量でも時間と手間がかかるのが現状である。開発当初からこの問題はあったが、データベース構造の正規化を最小限にとどめ、データベース構造を単純化することで、入力作業の負担をある程度軽減することしか対応できていない。今後の本格運用するにあたっては、効率的なデータ入力に加え、データ更新と削除が容易に行うことができるような高性能なデータ管理システムが必要と考える。

5. 次世代の統合型防災情報検索システムの開発に向けて

2009年度に防災研究所 Bulletin (英語版年報の位置づけ、査読有、現在は廃刊)の完全電子ファイル化を京都大学で行っている学術リポジトリ事業に委託することになり、2010年度には Bulletin の正式公開が可能になる予定である。しかしながら、4で示したように、現在の防災情報検索システムには改善すべき点が多い。そして、本格的に運用を始めると、防災資料の管理には人手と時間にかかるコストが大きく、データベース構造の変更などが困難になることが予想される。そこで、現在は手動で行っている防災情報検索システムへの登録を半自動にするための管理システム案、および、より効率的に防災情報を管理できるようなデータベース構造案の検討を行っている。その際、機能面で改善すべき点や現在のデータをより防災研究に活用するため、新機能の考案を行ったので、以下に示す。

(1) データベース登録に関する改善案

データベース構造に合わせてデータセットを作成し、登録作業を行うのは、技術者ではなく研究者であることが望ましいが、現実問題として現在のデータベース構造では研究者に輸入してもらうことは難しい。なぜなら、登録作業自体が煩雑で面倒だからである。裏を返せば、手軽にいつでも簡単に登録できるような仕組みを用意すれば解決することになるが、言葉では簡単でも、実際に仕組みを用意することは難しい。現時点ではっきりした答えが出たわけではないが、現在試行中の案をいくつか示す。

XML と RDBMS との併用

RDBMS にて管理すべき項目を最小限に減らし、残りの要素は XML で管理することで運用にかかる手間を減らすのが目的である。XML を利用したデータベースは、開発当初は敷居が高かったが、2007年に XPath、XQuery が標準化され、最近では通常の RDBMS に実装されることが多くなっており、PHP と XML の連携を使ったアプリケーション作成が容易になってきている。よって、もし研究者が入力したデータを、ローカルで XML を利用して管理することができ、その XML を RDBMS で適切に管理できるようなアプリケーションを開発できれば、研究者は自分の研究成果を自分の PC で管理することができ、さらに Web でも簡単に公開できるようになると考えている。

画像、動画、音声などテキスト以外の防災情報コンテンツの仮想フォルダ + コメント管理
画像や動画、音声などのデータは前述したように、ローカルでも管理し難い。研究者の多くは、フォルダ + ファイル名などで管理していると考えられる。また著作権があいまいな場合などが理由で災害調査の写真の多くは共有できずに埋もれている場合が多い。

そこで、防災情報検索システムが管理母体となって、画像を管理できるような体制を作ることができれば、研究者は自身の PC で写真などを管理する負担が少なくなるため、いつでも好きな時に画像を効率よく探すことができる。ただし、実際の登録の煩雑さの問題については、発想の逆転が必要であると考えている。つまり、画像や動画、音声などを登録する際

に、タイトルなどを必須とせず、先に研究者からは素材のみを提供して頂くようにする。その際、提供して頂いた時と同じフォルダ構造を Web プログラムで仮想的に作成する。データベースへ登録するメタデータは、フォルダ構造とファイル名、素材作成日時（写真なら撮影日時）、登録日時、提供頂いた研究者情報に加え、研究者自身が後で確認しやすいように、提供時のコメントも登録できるようにする。また登録後も Web 上で様々情報を更新できるようにし、さらに外部公開してもよいものも選択できるようにする。

これによって、研究者は自分だけで閲覧できる素材、研究所内だけで閲覧できる素材、外部に公開する素材をまとめて管理できるようになり、高性能なアーカイブ的な役割をシステム自体にもたせることもできる。また、画像などのファイル名は、FlexRena のような一括リネームするようなソフトを使い、EXIF の情報を元に、登録作業やファイル名の自動化を試みる予定である。

防災研究に特化したランキング手法

tf-idf などを用いた一般的なランキング手法は、コンテンツ中にあるキーワードの重みを tf-idf で評価し、コンテンツ 1 つの中に重みが高いキーワードが何個あるかで、コンテンツをランキングする手法である。つまり、すべてのコンテンツが計算後には既にランキングされており、後は検索語に引っかかったものをそのランキング順に並べれば良い。しかしこのやり方だと、新しい記事を追加するごとに tf-idf の値が変わり、追加コンテンツの形態素解析を行う必要があるため、管理する負担が大きい。そこで、キーワードの重みづけやコンテンツ追加時に半自動でランキングするための tf-idf などの数値を計算するため、ランキングに関わるキーワードを防災研究に関係ある重要なキーワードに限定することを考えた。ランキングに作用するキーワードが固定されている場合、形態素解析の必要がなく、tf-idf の計算も自動更新できるようなプログラムも作成可能である。現在、2200 の防災関連キーワードをデータベース化しており、これに tf-idf の数値を付け加えるようなシステムを開発予定である。また tf-idf に頼ったランキング手法だけでなく、ユーザが特定のキーワードに重みを段階的に与えられるような仕組みも考えている。例えば、特定の災害に関連した語の重みを倍にする、もしくは最大値の半分を足した重みでリランキングするなどを考えている。

6. おわりに

現在の防災情報検索システムはある一定の役割を果たしているといえるが、研究者が積極的に使うようなシステムとは言い難い。前述したような改善すべき点がきちんと改善されて初めて、研究者が使いやすいシステムとなり、また研究者の研究成果が公開されることで、防災研究所外の研究者間との情報共有、地域社会への貢献ができると考えている。防災情報検索システムが公開されるようになって既に何年も経つが、最近になって、外部の方がそれら閲覧して詳細を知りたいといった問い合わせがきたり、学生が研究者を目指す道具として効果的に利用していたりすることが判った。構築したシステムが多少なりとも役に立っているがあることと受け止め、今後も出来る限りシステム改善に努めていきたいと思っている。