

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 ホームページ上で高速検索可能なシステムの構築

平野憲雄・多河英雄・吉田義則・○松浦秀起・○辰己賢一
三浦勉・和田博夫・高山鉄朗

1. はじめに

京都大学防災研究所は昭和 26 年に災害の学理とその応用の研究を行うことを目的に設置された。今後、防災学研究をさらに推進させるためには、これまで防災研究所が蓄積してきた膨大な印刷物や映像情報を一般に公開し、分かり易い形で提供することが必要である。そして、それらの情報を簡単に検索できる形にすれば、防災研究所の基礎資料が社会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た。

今回技術室から申請した 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」(3 年) が採択された。平成 15 年度は、印刷物の中でも基本となる防災研究所年報に注目し、インターネット上で防災研究所年報を電子ファイルの形で一般公開し、高速検索によって誰でも必要な情報だけを取得可能なシステムを構築することを目的とした。

2. 年報電子ファイル化作業概要

防災研究所年報は、防災研究所が設立されてから 7 年経った 1958 年に第 1 号が出版され、2003 年度に出版された第 46 号まで続いている。第 46 号はすでに電子ファイル化されているため、第 45 号から過去に遡って、年報を電子ファイル化する作業を実施した。

年報電子ファイル化に必要な作業として一番労力がかかる部分は、大きく分けて以下の二点である。

- (1) イメージスキャナを使用して、年報を 1 ページずつデジタル画像ファイル（ビットマップ）の形で取り込み保存する。
- (2) (1) で保存したデジタル画像は細かい汚れが目立つため、画像編集ソフトによって 1 ページずつ手動で汚れを取り除く。

以上の作業は非常勤職員を雇用し、3 ヶ月間実施した。その結果、(1)、(2) の作業によって、約 2 万ページ (12 年分) が公開可能となった。

実際には、(1)、(2) の手順によって作成したデジ

タル画像ファイル（以下、ページ画像と表記する）を加工し、年報を論文単位で電子書類のデファクトスタンダードである PDF (Portable Document Format) ファイル形式に変換して提供する。

3. 年報文献検索システム

2 で作成した年報文献は非常に大量であり、ユーザが求める情報を必要なときに必要なときだけ有効に利用するためには、その情報を高速かつ簡易に情報を検索するシステムが不可欠である。

そこで、(1)、(2) の作業で作成したページ画像に対して OCR (Optical Character Reader: 光学式文字読取装置) を使用することにより、画像から文字を抽出して、テキストファイルの形で保存する。そのテキストファイルを利用した電子書類の検索する手段として、以下の三つの検索システムを構築した。

- ・ Web サイトで目次を表示して、そこで文献の検索を行う「目次検索」
- ・ タイトル、著者等の情報を年報の文献データベースに問い合わせた結果を元に、文献の検索を行う「カテゴリ検索」
- ・ 「簡易検索」も別途作成したが、これはカテゴリ数を減らした「カテゴリ検索」である。
- ・ 任意のキーワードをユーザに入力させ、そのキーワードについて年報の文章自体をすべて検索した結果を元に文献の検索を行う「全文検索」

これら検索システムは、テストを経て、2003 年 5 月に第 45 号から第 34 号までの年報検索サービスを正式に開始した。

http://www.dpri.kyoto-u.ac.jp/web_j/index_annuals.html

4. 終わりに

平成 16 年度には引き続き、残りの年報と A4 サイズを超える印刷物の電子ファイル化を予定している。今回の COE 研究分担課題を技術室にも採択して下さった関係者の皆様にお礼を申し上げます。