

Title	ワールドワイドウェブからの住所録の自動生成
Author(s)	津田, 朋樹
Citation	
Issue Date	2000-03
Type	Thesis or Dissertation
Text version	author
URL	http://hdl.handle.net/10119/1359
Rights	
Description	Supervisor: 佐藤 理史, 情報科学研究科, 修士

ワールドワイドウェブからの住所録の自動生成

津田 朋樹

北陸先端科学技術大学院大学 情報科学研究科

2000 年 2 月 15 日

キーワード： 住所情報, 住所録, ワールドワイドウェブ, 欠落情報の補完, 情報抽出.

これまで特定施設や店舗などの住所や電話番号といった情報（住所情報）を知るためには、人に尋ねるか、自分でその施設の住所情報が記載されているパンフレットや広告を手に入れるか、もしくは電話帳などで探すしかなかった。また、タウンページなどで図書館や病院などの種類（カテゴリ）から、そのカテゴリに属する複数の施設や店舗の住所情報を収集することができる程度であった。しかし、近年のワールドワイドウェブ（以下、ウェブと略記）の発展に伴い、現在ではウェブ上から多くの住所情報を入手することが可能になった。誰もが簡単に情報を発信、受信できるウェブ上には、今後ますます様々な情報が増え続けていき、住所情報を取得する上でも非常に有用なメディアになっていくことが容易に推測される。

ウェブ上の住所情報が掲載されているページは、大別すると以下の 2 種類に分類することができる。

1. 施設、店舗などの紹介ページ

特定の施設、店舗などを紹介する目的で作成されたページ。

2. 住所一覧ページ

図書館、病院などのカテゴリによって分類された住所一覧表（住所録）が掲載してあるページ。

いずれのページも、ウェブ上に点在しているため、ユーザは知りたい住所情報が掲載されているページを探すのに労力を要する。さらに、ページによって情報の記述形式が一樣でないため、ユーザ側としては利用しづらい面も多い。

これらの問題点を解決するために、本研究では、あらかじめウェブ上のそれらのページから住所情報を自動収集し、一つの住所情報データベースとして統合することにより、

ユーザが必要とする情報のみの住所録を自動生成して提供するシステムを提案する。これによりユーザは、必要とする住所録や住所情報を、必要なときに必要な表示形式で素早く手に入れることが可能になる。また、情報の収集も機械処理によって行なわれるため、ウェブ上で常に増え続ける大量の情報を、人手による労力を必要とせずに収集することができる。

本システムを作成する上で実現しなければいけない課題の1つとして、カテゴリ情報を持った住所情報を収集しなければいけないという点が挙げられる。これは、カテゴリ別の住所録を作成する上では必要不可欠な情報であり、ユーザが情報を検索する上でも重要な役割を果たすものだからである。カテゴリ情報を持った住所情報を収集するにあたり、どのような情報源から住所情報を収集するかを考える必要がある。住所一覧ページは、住所一覧表（住所録）の見出しとしてカテゴリ名が用いられている可能性が高く、比較的容易にカテゴリ名を抽出することが可能である。それに対し、施設、店舗などの紹介ページから、対象とする住所情報のカテゴリを抽出するのは非常に困難である。ウェブページ内から店舗名とそのカテゴリを抽出する研究も行なわれているが、これには非常に膨大な外部知識から構築した辞書が必要であり、実用的ではない。

以上の理由より本研究では、住所一覧ページを情報源としてカテゴリ情報を持った住所情報を収集する手法を提案する。そしてこの手法により、今まで人手で行なっていた、情報をカテゴリごとに分類する作業を容易に機械処理することが可能であることを示す。

本システムでは、まず、オフラインでカテゴリ別に住所情報を収集し、それをデータベースに格納する。こうして作成されたデータベースから、ユーザーの要求に従って住所録を生成する。本システムは次の3つの部分から構成されている。

1. 住所情報収集モジュール

カテゴリ名を入力として、ウェブ上から住所情報を収集し、データベースに登録する。

2. 住所情報データベース

住所情報収集モジュールで収集した住所情報を格納するデータベース。

3. 検索モジュール

ユーザの検索要求に応じて、住所情報データベースを検索し、検索結果を一覧表の形式（住所録）で表示する。

また、本システムでは、カテゴリ情報を持った住所情報を収集するだけでなく、ページ内の情報を元に欠落情報の補完も行なう。

本システムは、ウェブ上から様々なカテゴリの住所情報を収集できるように汎用性を持たせて作成されており、任意に選択した30カテゴリに関してウェブ上で住所情報の収集を行なったところ、約32,000件の住所情報を収集することができた。その住所情報の適合率は79%であった。また、システムのカテゴリ判定率は90%であり、正しくカテゴリ判定された住所情報の適合率は88%であった。

本システムの作成により、ウェブ上から必要な情報のみを効率良く収集する方法と、収集した情報をユーザの要求する形式に編集して提供する方法を示した。今後、さらにインターネットが発展するに伴い、このような収集した情報をユーザのニーズに合わせて編集し、提供するシステムの需要は、ますます増えてくるものと思われる。その実現方法を示した点で本研究は意義のあるものであったと言える。