

音声読み上げソフトに適した生活情報検索表示システムの設計： 病院を例に

Design of Retrieval and Listing System for Life Information Optimized for Voice Browser :
A Case of Hospital Retrieval and Listing

南雲 彩花[†]
Ayaka Nagumo

中平 勝子[†]
Katsuko T. Nakahira

北島 宗雄[†]
Muneo Kitajima

1 はじめに

高齢者・障害者を中心とした情報弱者の QOL を向上する手段として、Web 上で提供される生活情報の質向上が挙げられる。今日、高齢者・視覚障害者が Web ページを閲覧する際の補助として多くの端末で音声読み上げ機能の搭載が進んでいるが、未だ音声読み上げソフトへの対応が不十分である Web ページが散見されるのが現状である。特に個人サービスの場合、Web ページの製作者が表現したいデザインとの兼ね合いや、知識・経済的な問題により対応することが難しい場合もある。このように、数ある Web ページを音声読み上げソフトに対応させるのは不可能に近い。そこで、本稿では、その実現に必要な音声読み上げソフトに対応可能な検索表示システムの設計を行う。高齢者・視覚障害者に需要が高いと考えられる病院を例に、診療科ないし問題がある身体の部位を入力すると Web ページから病院選びに必要な情報を抽出し、ユーザが示した項目を読み上げるシステムを設計する。

このようなシステムを設計するには、様々な情報が統一的に書かれていない点が特に問題となっている。Web コンテンツを障害者にとってよりアクセシブルにする方法を定義した W3C が勧告する Web Content Accessibility Guidelines (WCAG) 2.0[1] で規定されているガイドライン 3.1, “読みやすさ: テキストのコンテンツを読みやすく理解可能にすること” に沿う為に、情報の整形を行う必要がある。ここでは、その解決策としてオントロジーを用いて内部的に情報を整形して DB に保存する。

病院を例とした場合、現行の診療科標榜方法は、医療法及び厚生労働省令第三条の二 [2] の規定に従った診療科名の表記例が多く、大衆向けに言い換えた表記揺らぎとの関連付けが難しい。そこで、診療科目や身体の部位等についてオントロジーを作成し、診療科目の関連付け及び表記揺らぎの吸収を行う。

2 システム概要

図 1 に本稿で設計するシステムの概要を示す。本システムでは、Web ページから病院選びに重要となる情報を抽出し、DB にそれぞれ保存する。ユーザが診療科ないし問題がある身体の部位を入力すると、DB から条件に該当する医療機関の情報を取得し、ユーザが示した項目を読み上げる。地域や診療日の条件指定は任意であるが、位置情報の取得が可能かつ地域の指定

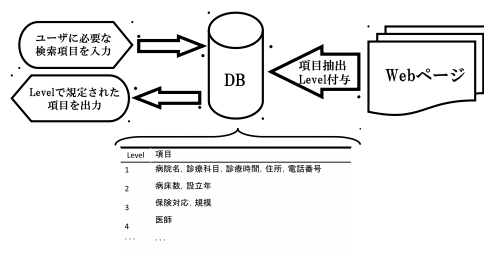


図 1 システム概要

が無い場合、現在地から近い順で並び替えを行う。

病院を探す際に参考となる情報としては、病院名、診療科目、診療時間、住所、電話番号、医師、設立年、規模、保険対応などが挙げられるが、ここで抽出する情報は、必要不可欠であると考えられる、病院名、診療科目、診療時間、住所、電話番号とする。今回は診療時間について、特定の日付のみの休診や時間変更などの動的な情報は抽出対象外とする。また、抽出する情報はテキストで書かれているもののみとし、画像のみによって表記されている情報については行わないものとする。

開発には python および PostgreSQL を使用し、オントロジー構築には protege を用いた。

3 オントロジー構築手法

3.1 オントロジー

診療科目に関連する情報抽出と DB への保存を行う為、また、ユーザが診療科、或いは、身体部位を入力して病院を検索する為に、オントロジーを用いる。また、診療日・診療時間についての情報を整えて DB に保存する為にオントロジーを用いる。

3.1.1 オントロジーの必要性

■診療科目 現行の診療科標榜方法では、医科の場合、単独で診療科名として使用可能な 18 の診療科名に加えて、以下の 4 項目を組み合わせて診療科名とすることが出来る。

- 部位, 器官, 臓器, 組織, 又はこれらの果たす機能
- 疾病, 病態
- 患者の特性 (性別, 年齢を示す名称)
- 医学的処置

2 つ以上の事項の組み合わせも可能である。組み合わせた診療科

[†] 長岡技術科学大学

名を使用する場合、表示形式が3通り存在するが、同分類に属する事項を複数組み合わせる場合は表示形式に制約がある。

このような診療科標榜方法の性質上、診療科の表記が多様である。また、Web上においては厳密に標榜方法に沿わずに、分かりやすさを優先した表現に変更されている場合もあり、多岐にわたる表記となっている。これらの診療科を構造化し、更に表記揺れを吸収する為にオントロジーが必要である。

■診療時間 各医療機関やその診療科のWebページ上において、診療日・診療時間の表記方法は多様である。例を挙げると、診療時間並びに、診察日の休診日どちらか或いは両方の曜日を列挙している場合や、表形式で診療日と休診日の曜を示している場合もある。更には“第一、第三月曜は休診”等、同じ曜日でも週数によって異なる場合もある。また、特に総合病院では、受付時間と診察時間が異なることが多い。時間帯を示す際の表記揺れ等も含め、これら診療時間の情報を整えてDBへ保存する為にオントロジーを使用する。

3.1.2 オントロジー構築の方針

■診療科目 図2に診療科についての、オントロジーの構成概要を示す。医科の診療科標榜方法に則った表記については、組み合わせを含む診療科名を、それぞれ単独で診療科名として使用可能な20診療科のサブクラスとして構造化を行う。また、a,b,c,dの項目それぞれについても、身体部位の区分や、表記揺れ及び類義語が多い為、構造化及び表記揺れの吸収を行い、組み合わせを含む診療科名と関連付けを行う。

診療科標榜方法に沿わない表現については、各分野においてWeb上で多用される表記の調査を行い、それぞれ診療科標榜方法に則った表記に関連付ける方針である。

■診療時間 W3Cが公開する、OWL時間オントロジー[3]に沿うように構築する。

3.2 抽出方法

病院名・診療科目・診療時間・住所・電話番号についての情報は、それぞれ正規表現を用いて抽出する。住所については、住所の定義に合致するテキスト情報が無く、郵便番号の定義に

合致するテキスト情報がある場合、住所検索APIを用いて住所を取得する。

4 オントロジーを踏まえたDB設計

図3に本稿で設計したDBのER図を示す。尚、ここで図3における“センター”は、地域における中核的な機能・役割を持っていると各都道府県等が認定した医療機関の他、特に中規模・大規模病院等の各医療機関で内部的に使用する部門名も含む。曜日属性は、月曜日から日曜日までの曜日と祝祭日を格納する。休診日の属性：時間帯は、午前休診・午後休診・終日休診をそれぞれ区別する為に使用する。

5 今後の課題

DBを設計したことで、音声読み上げソフトに対応させる為の情報整形がやりやすくなった。課題として、本稿ではテキストで記載されている情報のみを抽出の対象としているが、診療時間等が画像のみで記載されているWebサイトも少なくない。今後、画像化されている文字からもテキストを抽出可能にしていきたい。また、今回は、病院を探す為に必要不可欠であると考えられる病院名、診療科目、診療時間、住所、電話番号に絞り込んで情報抽出を行った。しかし、医師、設立年、規模、保険対応等、他にも病院選びの参考となり得る情報が挙げられる。これらの情報も抽出できるよう機能を拡張していきたい。

参考文献

- [1] Web Content Accessibility Guidelines (WCAG) 2.0, W3C 勧告 2008年12月11日, <https://waic.jp/docs/WCAG20/Overview.html>
- [2] 厚生労働省法令等データベースサービス, 医療法施行令, <http://www.hourei.mhlw.go.jp/hourei/html/hourei/contents.html>
- [3] OWL時間オントロジー, W3C 勧告 2017年10月19日, <http://www.asahi-net.or.jp/~ax2s-kmtm/internet/owl2/REC-owl-time-20171019.html>

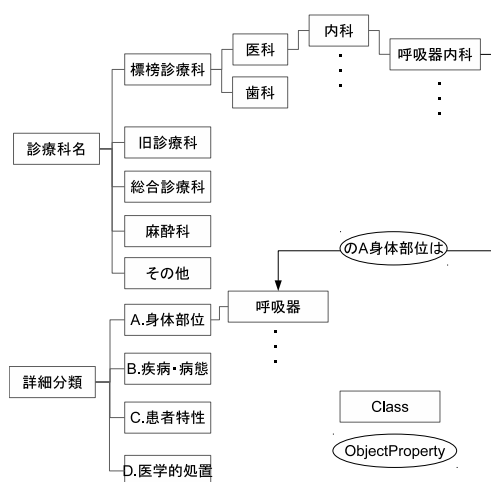


図2 オントロジーの構成概要

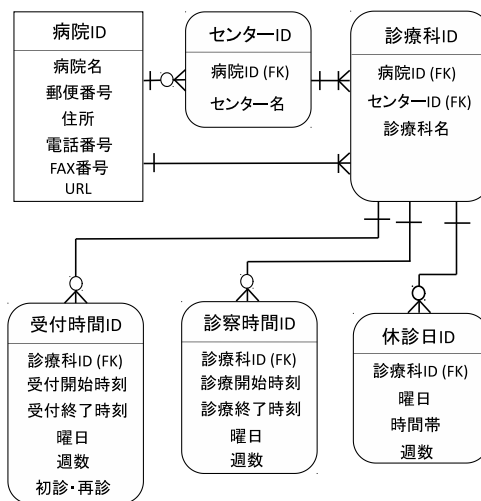


図3 DBのER図