

N-012

## 歴史データベース検索のユーザインタフェースを考慮した利用度推定 Using Intensity Estimation of Retrieval Fileds on User Interface in Histrical Database

小野田 賢人<sup>†</sup>  
Kento Onoda

安達 文夫<sup>‡</sup>  
Fumio Adachi

徳永 幸生<sup>†</sup>  
Yukio Tokunaga

杉山 精\*  
Kiyoshi Sugiyama

### 1. はじめに

歴史・文化的な資料をデジタル化し、それをデータベース(以下、DB)としてまとめ Web 上に公開するという取り組みがある。これらの DB にはそれぞれ専用の検索インタフェースが存在し、利用者はそれを使用して資料を検索する。これらの検索インタフェースはそれぞれの資料により異なり、主な利用者はそれぞれの資料を専門に扱う研究者である。したがって、資料の検索は研究を目的として行われており、資料により検索行動が異なる。しかし、利用者の検索行動にあった検索インタフェースを設計する方法は確立されていない。

そこで、利用者の検索行動を分析する事で、より適切な検索インタフェースの設計方法を検討する。検討にあたり国立歴史民俗博物館が Web 上に公開している DB の検索ログデータを用いる。

しかし、検索に利用可能な項目が検索インタフェースごとに異なる方法で提示されているため、各検索インタフェース中の項目の提示方法により検索回数が増減していると考えられる。そこで本稿では検索インタフェースが検索回数に与える影響について推定を行ったので報告する。

### 2. データベース「れきはく」

#### 2.1 検索インタフェース

データベース「れきはく」の検索インタフェースを図1に示す。検索は2通りの方法で行われる。1つは、検索語を用いた検索である。プルダウンメニューから検索項目を選択し、それと対応する入力フィールドに検索語を入力し検索を行う。もう1つは、ボタン検索機能を用いた検索である。ボタンを押す事で検索項目に入りうる語の一覧が表示され、それを用い資料を検索する。

プルダウンメニューで選択する検索項目は、検索インタフェースの画面を表示した時にデフォルトで選択されている検索項目とそうではない検索項目がある。また、ボタン検索機能を持つ検索項目と持たない検索項目がある。

#### 2.2 ログデータ

本稿では、2.1で述べた2つの提示方法の違いにより、検索インタフェースが検索回数に与える影響を推定する。

推定には国立歴史民俗博物館が Web 上に公開しているデータベース「れきはく」の検索ログデータを用いる。ログデータは、利用された検索項目、検索語、検索回数がひとまとまりになったものである。本稿では、検索項目ごとの検索回数に対して検索インタフェースが与える影響を推定し、検索項目の利用状況を推定する。対象とする DB は、研究文献データを集めた文献資料 DB 群とする。

<sup>†</sup> 芝浦工業大学, Shibaura Institute of Technology

<sup>‡</sup> 国立歴史民俗博物館, National Museum of Japanese History

\* 東京工芸大学, Tokyo Polytechnic University

### 館蔵資料データベースの検索

検索語を全角で入力し、【検索】ボタンを押してください。(検索語例)

図1 データベース「れきはく」検索インタフェース

### 3. 提示方法による検索回数の増減の推定

#### 3.1 利用度の推定方法

検索インタフェースが検索回数に与える影響を推定するために、その効果を重み付け係数としたモデル式を構築した[1]。モデル式の構築には「共通性のある資料を取り扱う DB において、検索の際に同じ意味合いを持つ検索項目は、DB が異なっても共通の検索のされやすさ(以下、強度)を持っている」という強度の仮説を用いる。そのため、強度は提示方法によらない検索回数であり、ログデータはこの強度に提示方法による重み付け係数が乗算された値とみなすことが出来る。しかし、DB ごとに総検索回数が異なるため、DB ごとに検索回数を正規化し扱う。さらに DB により同じ意味合いを持つ検索項目が異なる検索項目名であるため、各検索項目はメタデータである Dublin Core を用いてグループ化する[2]。

これまでに筆者はモデル式から推定される検索回数と実際の検索回数の二乗誤差を算出し、その差が最小となるように強度と重み付け係数の値を線形計画法によって決定する手法を提案した[1]。

しかし、上記の手法では強度と検索回数を同時に推定する。そのためモデル式が線形計画法に沿ったものとならならず推定を行う事が出来ない。そのため、まずは強度の仮説をもとに重み付け係数の値を推定する。そして推定された重み付け係数を用い、利用度の推定を行う。

#### 3.2 提示方法による重み付け係数の推定手法

重み付け係数は提示方法ごとに与えられる1つの重みであり、ある提示方法を基準とした場合にどれだけの効果があるかという形で表す。また、強度の仮説より DB が異なっても強度は一定であると言える。

そのため、DB 内の2つの検索項目の関係は、DB 内で提示方法が同じであれば DB が異なっても同一のものとなる。

また異なる DB でも同名の検索項目が同じ提示方法であれば、DB 内の 2 つの検索項目の関係は同一のものとなる。この関係を式(1)に示す。

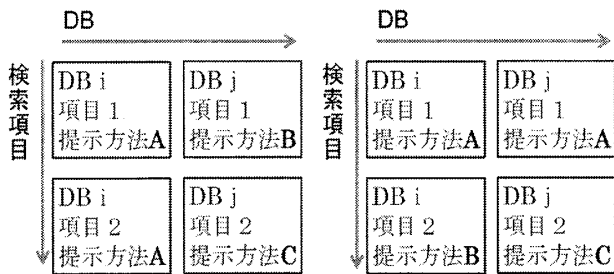
$$W(ij)X_j / W(il)X_l = W(kj)X_j / W(kl)X_l \quad (1)$$

X: 強度  
W: 重み付け係数  
i,k: DB 名  
j,l: 検索項目名

式(1)より図2のような検索項目の組み合わせにおいて、提示方法 B に対する提示方法 C の重み付け係数と提示方法 C に対する提示方法 B の重み付け係数を推定する事が出来る。

図2(a)のように、DB<sub>i</sub>で同じ提示方法である項目1と項目2を除算した値は、DB<sub>j</sub>で同じ提示方法である項目1と項目2を除算した値と同一の値をとる。そのため、DB<sub>j</sub>の項目2が提示方法Bであった場合の検索回数を推定する事が出来る。推定された検索回数とDB<sub>j</sub>の項目2が提示方法Cであるログデータの検索回数を比較し、提示方法Bに対する提示方法Cの重み付け係数を推定する。同様に提示方法Cに対する提示方法Bの重み付け係数を推定する。

図2(b)のように、2つのDBの項目1が同じ提示方法であれば、DB<sub>i</sub>の項目1と項目2を除算した値と、DB<sub>j</sub>の項目1と項目2が提示方法Bである場合の除算された値が同一となる。そのため、項目2が提示方法Bである場合の検索回数を推定する事ができ、推定された検索回数とDB<sub>j</sub>の項目2が提示方法Cであるログデータの検索回数を比較し、提示方法Bに対する提示方法Cの重み付け係数を推定する。同様に提示方法Cに対する提示方法Bの重み付け係数を推定する。



(a)DB 内で提示方法が同じ (b)DB 間で提示方法が同じ

図2 重み付け係数を推定できる4つの検索項目の組み合わせ

### 3.3 重み付け係数の推定

3.2の手法により重み付け係数は複数算出されるため、同じ提示方法の重み付け係数は相乗平均をとる。今回は、デフォルトで表示されておらず、ボタン検索機能を持たない提示方法を基準とした。結果を表1に示す。また、重み付け係数の対数を取り、その標準偏差と相加平均で重み付け係数のばらつきを評価する。評価指標として標準偏差を相加平均で除算した値を扱う。結果を表2に示す。

推定された重み付け係数より、検索項目がプルダウンメニューにデフォルトで表示されている場合、非表示の場合と比べ検索されやすいという事がわかる。また重み付け係数のばらつきも0.5前後に収まる値を示している。

この重み付け係数をもとに、それぞれの検索項目の平均的な利用度を推定する。推定は3.1で述べたモデル式から推定される検索回数と実際の検索回数の二乗誤差を求め、その差が最小となるように強度の値を線形計画法によって決定する[1]。算出された利用度を図3に示す。

図3から Description が多く利用されているという結果が得られた。しかし、この検索項目は推定に用いたDBのうち1つのDBしか保持していない。そのため推定に利用するDB全体を見た場合に保持数の少ない検索項目の取り扱いについて検討する必要がある。

表1 推定された重み付け係数

	デフォルト表示	非表示
ボタン有り	21.3	9.59
ボタン無し	22.5	1

表2 重み付け係数のばらつきの評価指標

	デフォルト表示	非表示
ボタン有り	0.54	0.09
ボタン無し	0.26	—

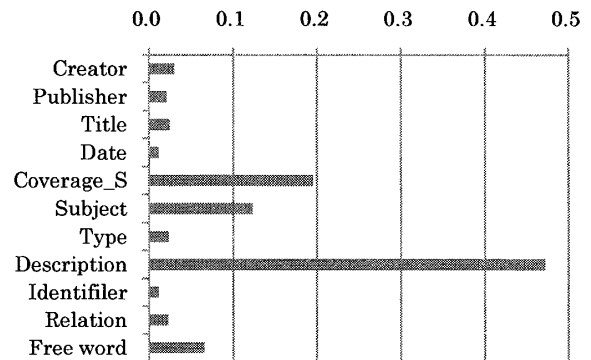


図3 推定された各検索項目の利用度

### 4. おわりに

本稿では、歴史・文化に属するDBの検索インタフェースが検索回数に与える影響の推定を行った。また、推定を行うために強度の仮説について検討した。そして、提示方法ごとの重み付け係数を推定した。最後に、推定された重み付け係数をもとに各検索項目の利用度の推定を行った。

今後は、今回推定に含めなかったDBを利用して重み付け係数の推定を行い、推定手法の適用範囲を検討する。

#### 参考文献

[1] 小野田賢人, 安達文夫, 徳永幸生, 杉山精 “歴史データベースの検索インタフェース設計のための利用度推定法の検討”, 第8回画像ミュージアム研究会, (2010).  
[2] 安達文夫, 鈴木卓治, “歴史データベースの Dublin Core へのマッピングとその課題”, 情報処理学会研究報告, (2006).