

# NDC を用いた人物ディレクトリの評価実験

片岡 祐輔<sup>†</sup> 浦 芳伸<sup>††</sup> 村上 晴美<sup>†</sup>  
<sup>†</sup> 大阪市立大学大学院創造都市研究科 <sup>††</sup> twinspire

## 1. はじめに

Web 上での人物検索は人気のある検索課題であるが、通常の人名検索では目的の人物を探すのが困難な場合がある。本研究室では、人物検索における人物の理解と識別のために、図書館の分類番号である NDC(Nippon Decimal Classification, 日本十進分類法)を Web 上の人物に付与する手法を提案し、同手法を実装した人物ディレクトリを開発した[1]。先行研究では、NDC の付与に使用する文書として Web ページのタイトル、HTML 全文、検索エンジンのスニペット、氏名を含む前後の文字列を検討したが、どれがよいか確証は得られなかった。本研究では、どの文書を使用するとよいか調べた。

## 2. 先行研究

提案手法では、文書から NDC の相関索引語を抽出し、スコアの高い NDC を人物に付与する。20 の氏名をクエリとして検索エンジンから上位 100 件の HTML ファイルを取得し、人手で同姓同名人物毎に分類したデータセットを使用した。人物は 137 人存在する。

図 1 に試作した人物ディレクトリの二次区分「78(スポーツ・体育)」の画面例を示す。三次区分の一覧と、二次区分に含まれる人物の一覧が表示されている。



図 1. NDC を用いた人物ディレクトリ

## 3. 評価実験

### 3.1 実験 1

タイトル、全文、スニペット、氏名の前後 50 文字、100 文字、200 文字の 6 種類の文書を比較した。提案手法で付与される最上位の NDC に対して、関連度の 5 段階評価を行った。5 を「非常に関連している」、1 を「全く関連していない」とした。タイトルの関連度が最も高かった(3.41)。

### 3.2 実験 2

6 文書それぞれのプロトタイプを作成し、NDC の二次区分(000-990 の 100 種類)を対象に正解率(正解数/人物数)を求めた。タイトルのプロトタイプが最も高かった(39%)。

### 3.3 実験 3

3 種類のプロトタイプ(タイトル、全文、前後 50 文字)について被験者 14 人に対してアンケート調査を行った。NDC の一次区分(000-900 の 10 種類)に登録されている人物の NDC に関して、3 段階評価を行った。3 を「よく当てはまる」、1 を「当てはまらない」とした。タイトルの関連度が最も高かった(2.01)。

試作したプロトタイプの総合的なよさを順位付ける質問でもタイトルが最もよかった(1.50)。タイトルを一位とした被験者からは「無駄な情報が少なくわかりやすい」というコメントがあった。

## 4. まとめ

NDC を用いた人物ディレクトリの評価実験を行った。文書として Web ページのタイトルを使うとよいことがわかった。

## 参考文献

- [1] 浦芳伸, 村上晴美: NDC を用いた人物ディレクトリの開発, 情報処理学会全国大会第 73 回全国大会(平成 23 年)講演論文集, 1, 651-652, 2011.