

Web 人名検索結果の要約と可視化を目指して

—2010 年代の進捗—

村上 晴美

大阪市立大学大学院工学研究科

harumi@osaka-cu.ac.jp

概要 著者の研究室では、テキストやデータからの人物の理解に関する研究を行っている。その中で、Web 上の人物の理解のために、Web 人物検索における要約と可視化の研究を行っている。研究の目的は、Web 上の人名検索においてユーザによる人物の選択と理解を支援するインタフェースの開発である。先行研究において 2000 年代の研究の進捗報告を行った。本稿では、2010 年代の研究の進捗の概要をまとめる。内容は、NDC 人物ディレクトリ、人間による同姓同名人物の分離、履歴書と地図の表示、Wikipedia 風概要文の作成、件名の付与である。

キーワード Web 人物検索, 進捗報告, 要約

1 はじめに

著者の研究室では、テキストやデータからの人物の理解に関する研究を行っている。その中で、Web 上の人物の理解のために、Web 人物検索における要約と可視化の研究を行っている。先行研究[1, 2]において 2000 年代の研究の進捗報告を行った。本稿では、2010 年代の研究の進捗の概要をまとめる。

2 研究の概要

研究の目的は、Web 上の人名検索においてユーザによる人物の選択と理解を支援するインタフェースの開発である。

Web 人名検索結果を同姓同名人物毎に分離し、人物クラスタ(人物毎の HTML 文書群)を作成する。ユーザによる人物の選択と理解を支援するために、検索結果と人物の要約と可視化を行う。

本研究では「抽出した情報を用いて人物の選択や理解に有用な情報を作成すること」を「要約」と呼ぶ。要約には抽出した情報の統合や、他の情報源の情報の取得を含む。

3 先行研究の概要

2000 年代の主な内容は、初期プロトタイプ、同姓同名人物の自動分離、職業関連情報の抽出、位置情報の取得[1, 2]及び、履歴書の作成[3]であった。この中で、職業関連情報の抽出と履歴書の作成については研究を完了した。

以下では、先行研究以降の進捗を述べる。

4 NDC 人物ディレクトリ

人物に図書館の分類体系である日本十進分類法(NDC)の分類記号を付与する手法を提案した。人物に図書館の分類記号を付与することにより、人物にラベルを付与すると同時に図書館の分類体系を利用した人物ディレクトリを構築できる。関連する語(索引語)と分類記号を関連付ける NDC の相関索引を利用し、タイトル要素のテキストに含まれる索引語の頻度に基づき分類記号を求める。人物クラスタに対して上位 5 件の NDC9 を付与して人物ディレクトリを試作した。20 氏名×100 件の Web 検索結果から得た 152 人物クラスタを対象として、4 手法×6 文書を組み合わせさせた評価実験を行った。相関索引とタイトル要素のテキストを用いて人物に NDC の分類記号を付与する手法の有効性を確認した。

本研究は文献[4-8]において発表した。

5 人間による同姓同名人物の分離

人がどのように Web 上の同姓同名人物を判別するか調査を行った。20 氏名×20 件の Web 検索結果を分けるように教示し、質問紙、プロトコル分析、インタビューを用いて判別プロセスを分析した。キーワード、職業、作品(架空の人物の場合は登場する作品)、顔画像、関連する人名が、人物を識別するために重要であることがわかった。実験結果に基づき同姓同名人物の分離モデルと知識構造モデルを提案した。

本研究は文献[9, 10]において発表した。

6 履歴書と地図の表示

人物の履歴書と地図を表示するシステムを試作した。

提案手法は人物クラスタに対する (1) 先行研究[3]を用いた履歴書の作成, (2) 履歴書中の学歴と経歴から学校と勤務先の抽出, (3) 学校と勤務先から位置情報の取得, (4) 履歴書と地図の表示の4段階からなる. 56 氏名 (主として有名人) × 50 件の Web 検索結果から作成した 56 人物クラスタを対象として, 学校と勤務先の抽出及び位置情報取得の評価実験を行った.

本研究は文献[11-13]において発表した.

7 Wikipedia 風概要文の作成

人物の概要文を作成する手法を提案した. 人物クラスタから属性情報 (氏名のよみ, 生年月日, 没年月日, 出身地, 職業, 所属, 役職) を抽出し, Wikipedia の第一文風の概要文を作成する. 20 氏名 × 50 件の Web 検索結果から得た 80 人物クラスタを対象として評価実験を行った. Wikipedia 風テンプレートと, 抽出された属性情報の組合せにより Wikipedia の第一文風の概要文を作成できることがわかった.

本研究は文献[14, 15]において発表した.

8 件名の付与

国立国会図書館の提供する件名標目表である NDLSH を人物に付与する手法を調査した. NDLSH を人物に付与することにより, ゴミの少ないキーワードを付与すると同時に関連語を用いた探索的な検索が可能となる. (a) 検索ランキング, (b) HTML 文書内の位置, (c) 同義語の使用, (d) 文書頻度の使用の組合せについて, 7 節と同様にして得た 80 人物クラスタを対象とした 405 (5 × 9 × 3 × 3) パターンの実験を行った. 実験環境においては, (a) 上位 10 件, (b) 氏名の前後 100 文字 (i.e., 200 文字), (c) 同義語の重み 0.5 倍, (d) 文書頻度の使用の組合せがよかった.

また, 日本図書館協会が提供する件名標目表である BSH4 においても同様の実験を行った. (a) 上位 10 件と (c) 文書頻度の使用が良いことが共通した.

本研究は文献[16-18]において発表した.

9 おわりに

NDC 人物ディレクトリについては研究を完了した. 完了していない研究については 2020 年代の課題として研究を続けている. 本稿作成時点においては, 概要文の作成と件名の付与の研究を行うと同時に, ネットワークインタフェースの試作とキーワード抽出の研究を行っている.

謝辞

本研究は JSPS 科研費 19K12718 の助成を受けたものです.

参考文献

- [1] 村上晴美, 上田洋: Web 人名検索結果の要約と可視化を目指して, 2009 年度人工知能学会全国大会 (第 23 回) 論文集, 2009.
- [2] Murakami, H., Ueda, H., Kataoka, S., Takamori, Y. and Tatsumi, S.: Summarizing and visualizing Web people search results, ICAART 2010, Vol. 1, pp. 640-643, 2010.
- [3] 上田洋, 村上晴美, 辰巳昭治: Web 上の人物理解のための履歴書作成, 人工知能学会論文誌, Vol. 25, No. 1, pp. 144-156, 2010.
- [4] 浦芳伸, 村上晴美, NDC を用いた人物ディレクトリの開発, 情報処理学会第 73 回全国大会講演論文集, Vol.1, pp. 651-652, 2011.
- [5] Murakami, H. and Ura, Y.: People search using NDC classification system, ESAIR 2011, pp. 13-14, 2011.
- [6] Murakami, H., Ura, Y. and Kataoka, Y.: Assigning library classification numbers to people on the Web, AIRS 2013, LNCS, vol 8281. Springer, pp. 464-475, 2013.
- [7] 片岡祐輔, 浦芳伸, 村上晴美: NDC を用いた人物ディレクトリの評価実験, 電子情報通信学会 2013 年総合大会情報・システムソサイエティ特別企画学生ポスターセッション予稿集, pp. 34, 2013.
- [8] 村上晴美, 浦芳伸, 片岡祐輔: Web 上の人物への図書館の分類記号の付与と人物ディレクトリの開発, システム制御情報学会論文誌, Vol. 29, No. 2, pp. 51-64, 2016.
- [9] 三宅悠生, 村上晴美: 人は Web 上の同姓同名人物をどのように判別しているのか, 電子情報通信学会第二種研究会資料 (第 19 回 Web インテリジェンスとインタラクション研究会), pp. 73-76, 2011.
- [10] Murakami, H. and Miyake, Y.: How do humans distinguish different people with identical names on the Web?: a cognitive science approach, CIKM 2012, pp. 2475-2478, 2012.
- [11] 王爽, 浦芳伸, 上田洋, 村上晴美: Web 上の人物履歴情報の地図上への表示, 電子情報通信学会 2012 年総合大会情報・システムソサイエティ特別企画 学生ポスターセッション予稿集, pp. 99, 2012.
- [12] 唐春亮, 王爽, 上田洋, 村上晴美: Web 上の人物履歴情報の地図表示システム, 2013 年度人工知能学会全国大会 (第 27 回), 2013.
- [13] Murakami, H., Tang, C., Wang, S. and Ueda, H.: Vitae and map display system for people on the Web, IEA/AIE 2014, LNCS vol 8482. Springer, pp. 348-359, 2014.
- [14] 村上晴美, 小西利宗, 浦芳伸: Web 上の人物の概要文の作成, 2016 年度人工知能学会全国大会 (第 30 回), 2016.
- [15] Murakami, H., Konishi, T. and Ura, Y.: Generating Wikipedia-like biographical sentences from Web people search results, IIAI-AAI 2017, pp. 992-993, 2017.
- [16] 下倉雅行, 村上晴美: Web 上の人物への NDLSH の付与, 2017 年度人工知能学会全国大会 (第 31 回), 2017.
- [17] Shimokura, M. and Murakami, H.: Assigning NDLSH headings to people on the Web, AIRS 2017, LNCS, vol 11292. Springer, pp. 189-195, 2018.
- [18] 下倉雅行, 村上晴美: Web 上の人物への BSH の付与, 2018 年度人工知能学会全国大会 (第 32 回), 2018.