

新聞記事からの人物・企業情報の抽出

西野 文人 落谷 亮

{nisino,ochi}@flab.fujitsu.co.jp

ある人物や企業に対して、それらの持つ属性やエピソードを知りたいことがある。そのためには単に名称を認定するだけでなく、実体とのリンク付けや、その実体に関係のある属性・エピソードを抽出することが必要となる。新聞のような、広い読者に向けて新しい情報を正確に伝えるための文章では表記上の工夫により、名称や実体とのリンクでのあいまい性が少なくなるように記述することが心がけられている。そこで我々は、固有名辞書のような大量の辞書を持った形態素解析ベースのシステムではなく、表記の手がかりによるパターンマッチング処理をベースとした人物・企業情報の抽出システムを開発をしている。本稿では人物情報や企業情報などを新聞記事から獲得する際の課題を示すとともに、人物情報抽出実験の結果を示す。

Person and Organization Information Extraction from Newspaper

Fumihito NISHINO and Ryo OCHITANI

Fujitsu Laboratories Ltd.

We are developing a system for the extraction of information(attributes, episodes etc.) on a person or an organization from a newspaper corpus. An essential part of this system is an object identification process, dealing with variety of expressions of the same object.

Considering characteristic of a newspaper article, following strict autographical rules, we adopted pattern matching approach, rather than a morphological analysis-based one, requiring a large dictionary.

In this paper we discuss tasks and problems arising from information extracting process and show the result of an experiment for personal information extraction.

1 はじめに

ある人物に対する情報を調べたいとか、逆に、事象や特技などから対象となる人物を調べたいというようなニーズがある。人物情報だけでなく企業情報についても同様である。このようなニーズに対応して、アンケート調査などをもとに人手で収集・編集された様々な人物事典がある。しかし、このような情報の収集・編集には手間ひまがかかり、日々変化する世の中の情報に追いつけていない。一方、新聞や雑誌などでは、様々な事象に関する最新の情報が記述されており、そこには人物情報や企業情報も豊富に存在する。しかし、単なる全文検索システムだけでは、知りたい情報を手軽に獲得できないことも多い。新聞や雑誌などから人物や企業の情報を手軽に獲得したい、あるいは人物情報・企業情報の事典を作成する支援システムが欲しいというニーズは強い。

技術的な観点で見ると、情報抽出や知識獲得の研究が盛んに行なわれており、テキスト中の人名や企業名を認識する研究 [1] や、辞書作成という観点で人名を獲得しようという研究 [2] が行なわれている。しかし、単なる文の解析や言語処理用の辞書ではなく、事典的な人物情報や企業情報の獲得という観点で考えると、新たな多くの課題がある。ところで、新聞のような、広い読者に向けて新しい情報を正確に伝えるための文章では表記的にあいまい性を少なく記述することが心がけられている。そこで我々は、固有名辞書のような大量の辞書を持った形態素解析ベースのシステムよりも、パターンマッチング処理により有用な情報が得られるのではないかと考えた。

本稿では人物情報や企業情報などを新聞記事から獲得する際にどのような課題があるのかについて検討し、パターンマッチング処理により有用な情報が得られることを実証する。

2 人物・企業情報を獲得するには

2.1 全文検索の問題点

新聞記事から特定の人物や企業に関する情報を文字列ベースの全文検索システムを使って獲得しようとした場合、まずは通常の検索と同様に検索の雑音や洩れ

の問題がある。特に、人物・企業情報の獲得という観点で考えると、以下のような問題・要望がある。

1. その人物・企業に関して関連する部分の情報だけを取り出したい。長い記事の中で、その人物・企業の観点では不要部分が含まれることも多い。
2. 同一情報しか得られない記事が複数存在する。同一情報はまとめた。
3. 表現ではなく実体に対しての検索を行ないたい。すなわち、同姓同名の人物や同一名称の企業による雑音は除去したいし、逆に、部分名(姓だけなど)・表記のゆれ・略称などに対応して情報獲得の洩れを防止したい。

2.2 人物・企業情報を獲得するために必要な機能

上にあげたような要望に応じて、人物・企業情報の獲得を支援するために、以下のような機能を開発することにした。

1. 人物・企業という実体の一属性(名称など)からの検索、あるいは職業・特技、あるいは業種のような特定の観点を指定した検索。
2. その人物・企業に関する基本的な情報の統合・整理。すなわち、注目している人物・企業に対して、基本的な事実情報を抽出し、これらの情報は統合して、その人物・企業の属性として付与する。
3. その人物・企業にまつわるエピソード情報(語録など)の抽出。すなわち、基本的情報以外のものもエピソード情報として関連情報の提示を行なう。

そのためには以下のような処理が必要である。

1. 実世界の実体の特定

抽出したいのは言語的情報ではなく実体に関する情報である。同一の名称が違う実体を指す(同姓同名など)場合にはそれらの区別が必要であるし、逆に一つの実体を異なる名称で示し

ている（本名・芸名，略称など）場合にはそれらの情報の統合が必要である。

2. 複数の文書からの情報抽出・整理

ある人物や企業についての情報は単一の記事からだけでなく，様々な記事から得られる。そこから得られる情報に対して重複した内容の単一化など要約処理が必要である。

3 処理の方針

3.1 新聞記事からの情報抽出

ここで新聞記事の性質について改めて考えてみることにしよう。新聞の記事というのは一般読者に対して新しい情報を誤解を生じさせないように伝えるものであり，プロの書き手によって記述されている。特に，誰でもが知っているとは限らない人名や企業名は，その正しい名称を正しく認定できるように表記上の配慮がなされているし，さらに実体としての人物や企業を正しく認定できるように，あきらかな文脈以外では，人物に対しては職業など，企業に対しては業種や本社所在地などの属性が記述されている。

新聞制作のための用字用語のよりどころを示すものとして新聞用語集があり，そこには例えば，中点の使い方の一つとして，「人名と肩書を明確に区切る必要があるとき」とし，「誤読の余地がない場合は省く」[3]とされている。これに対して，新聞記事の中では，「豊田章一郎・トヨタ自動車会長（69）の第八代会長就任を正式に決めた。」の後では誤読の余地がないということで，「豊田新会長は～」のように記述されているが，一方で，「電子スピニング法（豊田新・大阪大助手）」のように役職名との間に中点をおくことで人名と肩書とが明確に区別されている。

情報抽出の一般的な手法は，語彙知識や文法規則や背景知識に基づいて，文を形態素に分割し，名称を抽出し，それから事象構造を認定するというようなカスケード型のボトムアップの手法である。これは，文章の読み手の知識に基づいた処理と言え，適当な量の固有名辞書を持つなどにより，文章の解析の観点ではまずまずの精度を得ることができる。しかし未知の情報を抽出したいという観点で考えると，人名や企業名な

どの未知語に対する抽出の精度はそれほど高くはない。

上に述べたように，新聞記者は，どう記述すれば誤解なく一般読者に読んでもらえるかを考慮して記述していると考えられる。このような文章を解析するには，色々な解釈の可能性の中からあいまい性を除去していくという処理よりは，パターンに基づいて解釈するのがいいだろう。これはすなわち，読み手ではなく，書き手の知識を仮定した処理を行なっていると言えよう。我々は，書き手の知識に基づいた表現上のパターンによる情報抽出を行なうのである。

3.2 手がかりに基づくパターンマッチング

我々の目標は，単に人名や企業名のリストを作成するのではなく，それぞれの人物や企業という実体に対してそれぞれが持つ属性を抽出することである。したがって，ある記事から単に人名や企業名だけしか抽出できないならば，それは抽出しても無意味ということになる。例えば，「作家」の人名を抽出したいとして，「作家」という単語を含む文を検索し，その文に含まれる人名を抽出しても，抽出される人名には作家でない人物も含まれてしまうので，これだけでは意味がないのである。

そこで，我々は，職種や業種を手がかりにして人物や企業の情報を，あるいは逆に人物名や企業名を手がかりにして職種や業種の情報を獲得することを考えた。例えば，人物関連の情報表現としての典型的なパターンとして「<職種> の <人名>」があるが，ここから，「作家」，「画家」というような職種の名前を与えることでその職種に関連する人物を抽出し，逆に「五嶋みどり」のような人名を与えることで「バイオリニスト」のような職種情報を得ようというわけである。

我々は前の報告 [4] で，特定の事象に限定した情報抽出としてパターンマッチングに基づいてトップダウンに解析することで事象構造を抽出するシステムを紹介した。パターンマッチングに基づく解析は，パターンの汎用性が問題になるので，一般的な文章における解析ではその再現率が問題になる。人名の抽出ということで考えると，人名は様々な文の中で出現するので

パターン解析で十分な再現率が得られるか怪しいところである。しかし、文章中に存在するあらゆる人名を抽出する必要性はないこと、前述したように新聞のような情報伝達のための記事ではあいまい性がない（人物が特定できる）ように付属属性の情報が一緒に提示されていることを考えると、むしろ付属する情報とともに人名を抽出するのがいいのではないかと考えた。

例えば、「竹やぶの近くに住む車いすの画家、はらみちをさん(67)が〜」という文を考えてみよう。形態素解析でこれを正しく解析するのは困難であろう。また、文字種による要素分割をベースとしたもの[1]でも、一般に漢字列やカタカナ列に対しては考慮されるものの、ひらがな名称を処理するのは困難である。しかし、この文は「<職業>、<人名>さん(<年齢>)」のようなパターンとして「はらみちを」を人名として抽出することが可能である。

4 人物情報獲得のための処理

以下では、人物情報を獲得するための処理について述べる。

4.1 職種の獲得

我々の人名の獲得の方針は「作家」のような職業を手がかりに、人物情報を抽出することであるので、まず人名抽出の手がかりとなる職種のリストが必要になる。ここでは既存のシソーラスのほかに、国勢調査や社会学統計などで用いられる職業一覧(約900種)も利用できるし、また「家」や「士」のような接尾辞だけでも手がかりになる。また、ここでは型の強制変換の現象にも対応する。例えば、EDRの概念体系辞書で「バイオリニスト」は楽師という概念の直下にあり、ここには「ピアニスト」のような語が存在し、当然ながら、「バイオリン」や「ピアノ」のような楽器名は存在しない(「リードギター」というのが混在しているが)。しかし、文章中では、「バイオリンのxx氏」とか「将棋のxx九段」のように、本来は職業ではない名詞(楽器名、スポーツ名、学問、作品物、技芸など)が職業のように用いられることがあるので、これらも手がかりとして扱う。

4.2 人名と基本情報の抽出

職種や役職、あるいは特定事象(死亡記事や人事異動、組織情報など)の表現パターンに基づいて人名の抽出を行なう。その際同時に、その周囲に記述されている人物情報(出身地、在住地、生没年、作品名など)も抽出する。ある程度の再現率を得るために、「<職業>の<人名><助詞>」のような適合率のやや低いパターンでも登録しておく。これによって、「一人」、「役割」、「父親」といった一般語や、「ままでいる方」なども<人名>として抽出されてしまうが、このようなものはストップワードとして登録しておき取り除く。

複数の記事から抽出された情報は人名や職種などの観点にしたがって統合し、同一情報は単一にする。

図1に、漫画家に対する人物情報の抽出した結果の一部を示す(括弧内の数字はその情報が抽出された件数である)

4.3 エピソード情報の抽出

職種などを手がかりにしたパターンマッチングに基づく情報抽出によって、その人物の基本的な情報を抽出することができるが、しかしそれだけでは、その人物が持つ別の属性や、語録やエピソードのような付加的な情報は抽出できない。そこで、今度は人名を手がかりとして記事を検索し、その記事を解析する。

全文検索により記事を検索した後、今度は人物名を中心としたパターン表現規則により、人物に関連した情報を抽出する。例えば、「さくらももこ」に関する情報を抽出した結果は次のようになる。

さくらももこ 作品: 「ちびまる子ちゃん」(9) 職業: 原作者(3), 作家(1), 作者(7), 人気漫画家(1), 超売れっ子マンガ家(1), 漫画家(5)

これにより、漫画家という手がかりだけでは抽出できなかった、作家という属性や、あるいは職種の「漫画家」に対して「マンガ家」という表記のゆれがあることが抽出できる。

さらに、上記では抽出できなかったエピソード情報などの不定型な情報を抽出する。この部分を本格的に

うえやまとち	在住: 福岡 (1)	職業: 漫画家 (1)	
えびなみつる	在住: 東京都世田谷区 (1)	職業: 漫画家 (1)	
おおば比呂司	職業: 漫画家 (3)	出身: 札幌 (1)	
ごとう和	在住: 同市 (1)	職業: 漫画家 (1)	
さくらももこ	作品: 「ちびまる子ちゃん」 (1)	職業: 人気漫画家 (1), 漫画家 (4)	
しりあがり寿	職業: 漫画家 (1)		
すがやみつる	職業: モータースポーツファンの漫画家 (1), 漫画家 (3)		
たけしまさよ	本名: 竹内昌代 (1)	在住: 京都市左京区 (1)	職業: 漫画家 (1)
ちばてつや	職業: 引き揚げ者の漫画家 (2), 漫画家 (5)		
つげ義春	職業: 漫画家 (3)		
はらたいら	職業: 漫画家 (1)	出身: 同県 (1)	
みうらじゅん	職業: 漫画家 (1)		
みずしな孝之	職業: 漫画家 (1)		
みつはしちかこ	職業: 漫画家 (1)		
やなせたかし	職業: 原作者の漫画家 (2), 漫画家 (9)	出身: 同県 (1)	
やまだ紫	職業: 漫画家 (1)		
やまだ哲太	職業: 漫画家 (2)		
ウラジミール・モチャロフ	職業: 人気漫画家 (2)		
エルジー・シーガー	職業: 米国の漫画家 (1)		
コジロー	職業: 漫画家 (1)		
サトウ・サンペイ	職業: 漫画家 (1)		
サトウサンペイ	職業: 漫画家 (5)		
ジャック	職業: 漫画家 (1)		

図 1: 人物情報抽出結果例

行なうのは困難なので、現在は、記事の中から注目する人物が関連する部分を抜粋するにとどめている。「さくらももこ」に対するダイジェストの一部を表 1 に示す¹。

5 実験・評価

今回の実験では、CD 毎日新聞 91-95 の 5 年分の記事に対して、まず、「バイオリニスト」や「画家」という文字列を含む文を抽出し、そこにある表現パターンを分析することで人名の抽出規則の作成を行った。その際、あまり特殊なパターンは規則とせず、一般性のあると思われるものだけを規則化した。そして、規則作成の対象としていた記事とは別のものとして「作詞家」という文字列を含む記事 197 記事を取りだし、人手で作詞家である人物および作詞家でない人物に対してタギングをした正解セットを作

¹本ダイジェストプログラムは仲尾 [5] によるものである

成してもらい、それに対して前述の規則を適用し評価を行なった。

情報抽出の名称認定のタスクでは通常、文に出現するそれぞれの名称表現に対して、それが抽出できたかどうかで、再現率や適合率が求められるが、人物情報データベース作成という立場で考えるならば、各文章ごとの名称の認定が正しいかどうかではなく、必要な情報が獲得できたかどうか、そして取れてきた情報に雑音や誤りがないかが重要である。ここでは、情報が抽出できたかどうかの観点として、何人の作詞家の名前が抽出できたかどうかで評価することにした。

実際に「作詞家」という文字列を含む記事のテストセットから、70 名の作詞家がリストアップされた。なお、ここでは、空白に関する表記のゆれは吸収したが、「阿木耀子」と「阿木燿子」、「岩谷トキ子」と「岩谷時子」、「喜多忠」と「喜多條忠」²は別人物

²記事を読むと、「喜多忠」も「喜多條忠」も「喜田川」を作詞したようになっており、同一人物らしい

910112030	「おどるポンポコリン」センバツ入場曲に
「まる子のテーマで入場なんてうれしいよ」——11日、第63回選抜高校野球大会（毎日新聞社、日本高野連主催）の入場行進曲に決まった「おどるポンポコリン」。子供からお年寄りまで親しまれたレコード大賞曲が、センバツを彩る。…一方、作者のさくらももこさんは「ポンポコリンで行進するなんて夢のよう」。	
910603059	【ベストセラー】ノンフィクション単行本（5月28日・東販調べ）
（1）もものかんづめ（さくらももこ、集英社・850円）…	
921201262	【タレントロジー】私は「ちびまる子ちゃん」——さくらももこ語る
…大人から子供まで熱烈に支持されている「ちびまる子ちゃん」がアニメの正月映画になった。原作者のさくらももこは、素顔を知る機会が少なかったが、今回は積極的にインタビューに応じ、思うことをしゃべってくれた。…	
930507214	【フレッシュ】ドラマと歌でデビュー＝藤川なお美さん
…お店の隣に住む歌の好きな高校生が藤川なお美。…というわけで“フレッシュ”の誕生である。…芸名のお美はドラマの役名だから、名付け親は脚本のさくらももこ。…	

表1: 「さくらももこ」に対する記事のダイジェスト例（抜粋）

として取り扱っている。

これに対して、システムが人物として抽出したのは52件（マッチング処理の段階では56件が抽出され、「父親」、「卵」など4件がストップワードとして取り除かれている）であり、そのうち、人手でリストアップした作詞家と一致したのは51名であり、適合率は98%、再現率は73%であった。ストップワードとして取り除ききれなかったものは「作曲家」だけであった。

抽出できなかった人名としては以下のようなものがあった。

1. 表の中での人名

作詞家（シングル）
順位 売上数（万）
…

（3） 317 稲葉浩志

新聞記事中には、オリコンランキング、高額納税者リストのような表が存在する。表は表として認識してそれに対応する処理が必要であり、現在はまだ対処していない。

今回の評価の中で、表の中でしか出現しなかった人物が9名おり、表を対象記事からとりはずして考えるなら、再現率は84%である。

2. 職業表現のないもの

- 喜多條さん作詞の「神田川」は、～

「作詞家」とは明示されていないが、「作詞」という動作が示されている。作詞をしたからといって作詞家であるとは限らないが、作詞家の可能性は高い。

3. 職業情報の存在する文にある人物情報が不完全なもの

- 一つは、み群杏子作の「恋ごろのアドレス」。み群はNHKのみんなの歌三十周年記念コンクール最優秀賞を受賞した作詞家でもあり、戯曲はこれが初めて。

2文めの「み群」が1文めの「み群杏子」を指しており、これによって「み群杏子」が作詞家であることがわかる。

4. 1文だけの処理では情報が完全でないもの

- 今週の「私の選んだ文庫ベスト3」の選者は、伊集院静氏。1950年山口県生まれ。CFディレクターや作詞家などとして華々しく活躍した後、～

主語の補いが必要である。

以上のような抽出失敗はあるものの、従来の形態素解析や文字種区切りをベースとした手法ではあまり考慮されていなかったひらがな名の人名なども抽出でき

ており、新聞記事のようなあいまい性を少なく書いた文章ではパターンマッチングに基づく手法は有効であると言えよう。

人物や企業に対して職種や業種が明示されるのは、死亡記事や葬儀参列記事など、その記事ではその職種や業種に関する話題が主テーマではないときである。このような場合には、職種や業種によらず共通の表現として情報の抽出が行なえる。したがって、他の職種や業種をもとに作成した規則がそのまま利用できる。一方、記事が、その職種や業種の問題に深く関連する場合には、一般常識の範囲ということで、情報が省略され、また、それぞれに応じた表現パターンが存在する。例えば、会社社長は人物の情報としてより会社の情報の属性として社長名が付与されていることが多いし、国会議員や各自治体の首長では選挙に関する情報が存在するという具合である。今回の実験では「作詞家」として抽出したが、「作詞」という動作で捉えるならば、個別のパターンが必要であろう。そのためには、人物一般の基本属性とともに、各職種を階層化してそれぞれに応じた重要属性の定義及びそれぞれの職種に応じた特徴的記事とその表現パターンの分析が必要である。

6 今後の課題

今後取り扱わなければならない課題としては以下のようなものがある。

6.1 記事全体からの処理

「<人名>さんは～<職>。」のようなパターンでは、人物情報としては姓しかとれないことも多い。これはこの文以前にこの人物の情報が記述されていて、この文で付加的に職業情報を示しているからである。記事全体から人物情報を抽出する必要がある。

6.2 照応処理

各人の情報をデータベースに格納するには、表現ではなく具体的な実体を指し示す記述がなされなければならない。「同」に対しての照応処理が必須である。また、「同」以外にも、「地元」とか「親友」で

の「地元」とはどこであるのか、「親友」とは誰の親友なのかの処理が必要である。

6.3 並列表現への対応

並列表現からも人物や企業に関する情報を抽出することができるが、並列表現に対する属性の付与の仕方には様々なバリエーションがあり、その取り扱いには注意が必要である。以下に注意が必要な並列表現の例を示す。

- 作家の司馬遼太郎、三浦朱門、大江健三郎、三浦哲郎、村上元三、杉森久英、河野多恵子、大庭みな子、加賀乙彦、衆議院長の桜内義雄、元経済企画庁長官、高原須美子、映画監督の熊井啓、俳優の森繁久弥の各氏。
- 【ピアノ】野原みどり【バイオリン】諏訪内晶子、西沢和江【チェロ】山本裕康【作曲】権代敦彦

6.4 属性情報の整理

現在、属性として得られている情報は整理されていないものである。作品や在住などの一部の情報は属性として抽出しているが、どのような情報を属性として抽出するかも検討が必要である。また、名前に関する属性、人物に対する属性、職種に対する属性のように属性を区別することは必須である。また抽出されてきた属性値に関しても、同義語や上位下位語のまとめなどの整理が必要である。

属性情報の整理の別の問題としては、情報の変化への対応がある。すなわち、得られた情報の中には、年齢や役職、在住地のように時間とともに変化する情報があり、その情報がいつの時点のものなのかを保持しておく必要がある。ただし、保有している情報を単に提示したのでは、人間にとっては今度は逆に煩わしい。内部で保持する情報とは別に、外部に示す時は必要情報だけを取り出して提示するというような工夫も必要であろう。

6.5 情報の一般性の判定

利用目的によってはある職種や業種の代表的な人物や企業を知りたいことがある。その場合には、抽出された情報が代表的なものか、あるいは一般性を持つ

ものかどうかを判定することが必要である。例えば、「バイオリニストのxx」はxxの職業がバイオリニストであることを示しているが、「バイオリンのxx」という表現では、xxがバイオリニストであるとは限らない。この区別は必要である。

話題としている分野において代表的な人物や企業は、その話題の文章の中では一般読者の常識の範囲ということで、その職種や業種が明示されないことが多い。分野の違う話題からは「AのX」のような表現でXに対する属性Aを取り出せるが、これはAの代表的なものにXがあるということを示しているわけではない。代表例は「 X_1, X_2, \dots, X_n などのA」のような例示表現から取り出すというようなことが必要である。

また、記事の中には漫画やドラマに出てくる登場人物・企業のような架空のものが記述されていることもあり、これも判定をしておきたい。以下の例は、1)は漫画に出てくる登場人物について述べたものであり、2)はドラマ中の人物(括弧内は俳優名)である。

1. お隣に住む流行作家の伊佐坂先生
2. 西浅草署の刑事、高杉(松方)は、推理作家の妻法子(森口)に連れられ浅草のほおずき市へ。

6.6 人名検索におけるもれの防止と雑音の除去

単に姓名の文字列だけで検索したのでは検索もれが生じる。旧字への対応(例: 関沢義, 関澤義), 表記のゆれへの対応(例: 貴の花, 貴ノ花, 貴乃花), 姓だけ(役職などが後接することも多い), 姓と名の間の空白などの表記バリエーションへの対応, 芸名だけでなく本名での検索あるいは姓名変更への対応も必要である。

雑音の除去も考慮しなければならない。「林弘」という文字列の検索では「小林弘一」も検索されてしまうというような言語レベルの処理だけでなく、実体とのリンクの際での雑音(同姓同名)にも対処することが必要である。所属や役職は変更, 兼務があるので, 生没年のような絶対的な基準での判定が必要である。

7 おわりに

本稿では情報を明確にして提示する代表的な文章として新聞記事を対象にして、そこから人物情報や企業情報を獲得する際の問題点について検討した。特に、特別な知識がない読み手に対して情報を明確にするために書き手があいまい性のないように記述している文章では、ボトムアップな分析をしてあらゆる可能性を確かめるよりは、特に大規模な辞書を持たない表現のパターンマッチングに基づく手法によって精度の良い人物・企業情報が抽出できることを示した。今後、より広範囲の職種・業種の人物・企業情報の抽出、より多い属性の抽出、新聞に限らず雑誌などからの情報の獲得について考えていくことが必要である。

謝辞 (財)計量計画研究所の木田敦子研究員には正解データは作成いただくとともに、様々なコメントをいただいた。ここに感謝の意を表します。また、本研究では、CD 毎日新聞 91-95 版を利用した。コーパスの利用を許可していただいた毎日新聞社、およびコーパスの利用に関して尽力された方々に深く感謝いたします。

参考文献

- [1] 福本淳一, 下畑光夫, 梶井文人: 固有名詞抽出における日本語と英語の比較, 情処研報, NL126-15, pp. 107-114 (1998).
- [2] 久光徹, 丹羽芳樹: 辞書と共起情報を用いた新聞記事からの人名獲得, 情処研報, NL118-1, pp. 1-6 (1997).
- [3] 毎日新聞社(編): 新版 毎日新聞用語集, 毎日新聞社 (1996).
- [4] 西野文人, 落谷亮, 木田敦子, 乾裕子, 桑畑和佳子, 橋本三奈子: トップダウンなパターン解析に基づく情報抽出, 情処研報, NL124-13, pp. 95-102 (1998).
- [5] 仲尾由雄: 見出しを利用した新聞・レポートからのダイジェスト情報の抽出, No. NL117-17, pp. 121-128 (1997).