

( 続紙 1 )

京都大学	博士 ( 情報学 )	氏名	佃 洗撰
論文題目	A Study on Web Search and Analysis based on Typicality ( 典型性に基づくWeb検索と分析に関する研究 )		
( 論文内容の要旨 )			
<p>本論文では、典型的な情報の検索、および、非典型的かつ有用な情報 ( 意外な情報 ) のための検索手法の提案および分析を行っている。ユーザがあるドメインの情報を調べるときに、典型的な情報を提示することはそのドメインの概要を知るうえで有用である。また、非典型的な情報や意外な情報を検索し提示することは、ドメインをより深く知ったり、ユーザの興味を喚起したりするうえで有用である。しかし、通常の検索エンジンを用いて典型的な情報や意外な情報を検索するのは困難である。そこで、本論文では、典型性に基づいてWeb検索および分析を行う方法を提案している。情報の典型性を測る際には、社会認知量に基づく典型性、および、データ量に基づく典型性を考慮することで、多様な観点からの典型性の分析を行っている。さらに、認知心理学の分野で行われた典型性に関する研究を基に、類似情報の多寡という観点に基づく典型度と、情報の出現頻度がどれだけ大きいかという観点に基づく典型度を用いて、典型度を求める。そのために、各観点での典型度を求める方法についても提案している。具体的には、本論文では「典型度に基づくオブジェクト集合検索」、「語の認知度と語間の関係の非典型度に基づく意外な情報の発見」、「社会認知量に基づく語間の関係の典型度推定による意外な情報の発見」に関する研究を行い、提案手法・システムの評価を行っている。</p> <p>本論文は全7章から構成されている。その概要は以下の通りである。</p> <p>第1章は序論であり、本研究の背景、アプローチ、および、本論文の研究の全体の概要について述べている。</p> <p>第2章では、第3～6章で取り組んだ各研究課題の関連研究について述べ、本研究との位置づけを述べている。</p> <p>第3章では、オブジェクト集合の典型度を推定する手法の提案を行っている。本章ではオブジェクト集合としてレシピを対象としている。提案手法では、食材間の相性と、カテゴリ内で最も典型的なレシピとの差異を元に典型度を求めている。さらに、社会認知量に基づくレシピの典型度およびデータ量に基づくレシピの典型度を測る手法を提案し、レシピの典型度の正解データとの比較分析を行う。提案手法により、ユーザは例えば「カルボナーラを初めて作るので典型的なレシピを探したい」や「カルボナーラは何度か作ったことがあるので一風変わった非典型的なレシピを探したい」といった意図を反映した検索ができるようになる。</p> <p>第4章では、語の上位下位関係を用いて、入力クエリに対する同位語らしい語および上位語らしい語を発見する手法を提案している。提案手法では、語の上位下位関係を基に作られる2部グラフに対してHITSアルゴリズムを拡張した手法を適用することで、同位語らしさ、上位語らしさを求める。例えば、「落合博満」というクエリに対しては、同位語らしい語として「野村克也」や「イチロー」といった語が、上位語らしい語として「野球監督」や「野球選手」といった語が得られることが期待される。本章で提案した手法は第5章および第6章で使用される。</p>			

第5章では、ユーザが与えた1語のクエリに対して、そのクエリに関する意外な情報を発見する手法の提案を行う。提案手法では、クエリに対して意外度の高い関連語を発見し、クエリと意外度の高い関連語を基に意外な情報を発見する。その際、クエリの関連語の中でもクエリとの関係が非典型的であり、かつ認知度が高い関連語ほど意外度が高いという仮説に基づいて関連語の意外度を求める。たとえば提案手法により、「落合博満」というクエリに対して「ガンダム」という関連語の意外度が高いことがわかり、これを基に「落合博満はガンダムマニアである。」という意外な情報を発見することが期待できる。

第6章では、社会認知量に基づく典型度と、データ量に基づく典型度に差がある情報は意外であるという仮説の基、意外な情報の発見を目的としている。本章では特に語間の関係の典型度に着目する。上記の仮説を検証するために、クラウドソーシングを用いて社会認知量に基づく関係の典型度および関係の意外度を取得する。さらに、語の認知度および同位語を考慮することで社会認知量に基づく典型度の推定手法を提案する。最後に、社会認知量に基づく典型度、データ量に基づく典型度、オブジェクトの認知度を用いて情報の意外度の推定を行う。

第7章では、本研究で得られた研究成果をまとめ、さらに今後の展開について述べている。

注) 論文内容の要旨と論文審査の結果の要旨は1頁を38字×36行で作成し、合わせて、3,000字を標準とすること。  
論文内容の要旨を英語で記入する場合は、400～1,100 wordsで作成し  
審査結果の要旨は日本語500～2,000字程度で作成すること。

(続紙 2)

(論文審査の結果の要旨)

本論文は、典型的な情報の検索、および、非典型的かつ有用な情報（意外な情報）の検索を行うための手法の提案・分析を行っている。情報の典型性を測る際には、社会認知量に基づく典型性、および、データ量に基づく典型性を考慮することで、多様な観点からの典型性の分析を行っている。

学位申請者は、認知心理学の分野の典型性に関する研究の知見を基に、似たものの多さという観点に基づく典型度と、出現頻度の大きさという観点に基づく典型度の推定手法、および、それに基づく意外な情報の発見手法を提案している。さらに、提案手法の有効性検証のための評価実験を行い、興味深い結果を提示している。

具体的には、本論文の研究によって得られた成果は以下のように要約される。

1. オブジェクト集合を対象として典型度の推定手法を提案し、社会認知量およびデータ量に基づく典型度推定手法を提案した。各提案手法の評価は、評価者によって作成された典型度の正解データとの比較によって行われ、提案手法およびデータ量に基づく典型度が正解データと高い相関があることを定量的に示している。
2. 語の上位下位関係を用いて2部グラフを構築し、入力クエリに対する同位語らしい語および上位語らしい語を発見する方法を提案した。評価実験ではクラウドソーシングを用いて正解データを作成し、提案手法では比較手法よりも高い精度で同位語らしい語および上位語らしい語を求めることができていることを示した。
3. 与えられたクエリに関する意外な情報を発見する手法を提案した。提案手法では、クエリの関連語の中でもクエリとの関係が非典型的であり、かつ認知度が高い関連語ほど意外度が高いという仮説に基づいて関連語の意外度を求めている。評価実験では、提案手法および比較手法で発見された意外な情報の意外度を評価者に回答してもらうことで定量的な評価を行った。実験の結果、上記の2つの仮説は意外な情報を発見するうえで有用であることを示した。
4. 社会認知量に基づく典型度とデータ量に基づく典型度に差がある情報に対して、人は意外であると感じることを、クラウドソーシングを用いた評価実験により示した。また、社会認知量に基づく典型度の推定手法を提案し、提案手法の有用性を定量的に示した。さらに、社会認知量に基づく典型度、データ量に基づく典型度、語の認知度を用いて、語間の関係の意外度の推定を行った。その結果、社会認知量に基づく典型度の分散が大きいカテゴリでは高い精度で情報の意外度の推定ができていることを明らかにした。

以上、本論文は、典型的な情報の検索手法、および、意外な情報の検索手法について研究を行ったものである。申請者は、本研究を通して、これまでの情報検索では十分に考慮されていなかった情報の典型性という概念を利用することで、典型度に基づく検索および分析の方法を提案した。さらに、提案手法の有効性を示すために詳細な実験を行うなど、学術上寄与するところが少なくない。よって、本論文は、博士（情報学）の学位論文として価値あるものと認める。また、平成26年8月8日に論文内容とそれに関連した事項について諮問を行った結果、合格と認めた。

注)論文審査の結果の要旨の結句には、学位論文の審査についての認定を明記すること。更に、試問の結果の要旨（例えば「平成 年 月 日論文内容とそれに関連した口頭試問を行った結果合格と認めた。」）を付け加えること。

Webでの即日公開を希望しない場合は、以下に公開可能とする日付を記入すること。  
要旨公開可能日： 年 月 日以降