

(続紙 1)

京都大学	博士（ 情報学 ）	氏名	加藤 誠
論文題目	A Study on Information Retrieval based on Relational Analysis (関係分析に基づく情報検索に関する研究)		
(論文内容の要旨) 本論文では、知らないドメインの情報を知っているドメインのクエリ、または、検索対象を例示することで検索する、関係分析に基づく情報検索を提案している。語彙の制限や検索対象に対する知識不足のために、知らない分野の情報を検索することは困難である場合が多い。そこで、本論文では、知っているドメインのクエリ、または、検索対象を例示することで、知らないドメインの情報を検索する方法を提案している。この実現のためには、オブジェクト間の対応を発見する必要があり、また、対応は構造（関係）の類似性に基づくため、オブジェクト間の関係、および、オブジェクトとドメインの間の関係をモデル化する必要がある。そのため、これらの関係をモデル化し、関係の類似性を測る方法についても提案を行っている。具体的には、本論文では「クエリ関係に基づく構造化クエリ推薦」、「語関係に基づくクエリ修正」、「抽象具体語関係を利用したWeb画像検索」、「ドメイン-オブジェクト関係に基づく例示検索」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。 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。 第1章は序論であり、本論文の研究の背景、本論文におけるアプローチ、および、本論文の研究の全体の概要について述べている。 第2章では、第3～6章において説明するそれぞれの研究課題に特に深く関連する従来研究について整理し、本研究との位置づけを議論している。 第3章では、クエリ推薦を構造化して提示し、ユーザの検索行動を効率化する方法について提案を行っている。例えば、「ニコン」というクエリが入力された場合に、システムは「ニコン」に関するクエリ推薦「ニコン coolpix」と共に、関連するエンティティ「キャノン」に関するクエリ推薦「キャノン ixy」を並べて提示することで、異なるクエリ間のクエリ推薦の対応関係を明示し、関連エンティティ間の比較を容易に行えるようにすることを目的としている。本章では検索エンジンに蓄積されたログを用いることで、クエリ推薦間の対応関係を発見する手法を提案し、検索行動に大きな影響を与えることを示している。 第4章では、語間の関係の類似性を用いて、異なるドメイン間で対応する語を発見する方法を提案している。クエリは3語で構成され、例えば、「京都」、「八つ橋」、「北京」を与える。このとき、出力xとしては「京都」と「八つ橋」の関係、「北京」とxの関係が類似するようなx、例えば、「月餅」が期待される。本章では、共起する語と言語パターンを利用することによって関係をモデル化する方法を提案している。			

第5章では、抽象語による画像検索を抽象語からの具体語集合変換によって改善する方法を提案している。現在のWeb画像検索エンジンは具体的な語によるクエリに対しては十分高い適合率を示すが、一方で抽象的な語によるクエリに対しては高い適合率を得ることができない。そこで、本章では、ソーシャルタグ情報を用いて、抽象的な語を具体的な語に変換し、抽象的な語の代わりに具体的な語によってWeb画像検索を行うことによって、その精度を改善する方法を提案している。

第6章では、ドメイン-オブジェクト関係に基づく例示検索を提案している。例示検索は、例を与えることで似た情報を検索する手法であるが、例示するドメインと検索を行うドメインが著しく異なる場合、適合する情報が与えられた例と必ずしも類似するとは限らない。そこで、本章では、ドメイン-オブジェクト関係を分析し、入力された例のそのドメインにおける役割を特定することによって、異なるドメインにおける例示検索を可能にしている。

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。

(論文審査の結果の要旨)

本論文は、関係の類似性により異なるドメイン間の対応を発見することにより、知っているドメインのクエリや検索対象を入力として、知らないドメインの情報を検索する手法を提案したものである。

検索が困難である状況には、ユーザが適切なクエリを入力できない場合と、システムが与えられたクエリに対して適切な検索結果を出力できない場合がある。前者には事前知識が欠けているために適切なクエリが入力できない場合が含まれ、本論文では主にこの問題について取り組んでいる。

学位申請者は、知っている分野の情報を例示することで、事前知識があまりないような知らない分野における検索方法を提案している。さらに、それら提案手法について、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. 異なるクエリに対するクエリ推薦の対応関係を発見し、構造化クエリ推薦インタフェースを提案し、検索エンジンのログデータによってこれを実現する方法を提案した。提案システムの評価は、ユーザ実験によって行われ、特に比較検索タスクにおいてユーザの効率的な検索に寄与していることを定量的に示している。
2. 語の関係性をWeb検索エンジンの索引を用いてモデル化し、与えられた a , b , c という語に対して、 a と b の関係、 c と d の関係が類似するような語 d をもっともらしい順に出力する方法を提案した。提案された関係のモデル方法を評価するために、複数の関係に対して提案手法を適用し、特に語共起と言語パターンを組み合わせた手法が最も良いという知見を得ている。
3. 現在のWeb画像検索エンジンでは抽象的な語による精度は低く、一方で、具体的な語による精度は高いことを事前実験で示した上で、抽象的な語を具体的な語に置き換え検索することでその精度を改善する手法を提案した。実験においては、特に概念、属性、動作などの抽象語において精度が改善できたことを示している。
4. 異なるドメインにおける例示検索において、あるドメインにおけるオブジェクトのドメイン-オブジェクト関係を考慮する事により、より適切な特徴空間において類似度を計算する方法を提案している。評価実験では、所属するドメインを考慮する事が必要であることを示し、また、既存の手法と比較することによって提案手法の有効性を示した。

以上、本論文は、関係を分析しその類似度を利用する情報検索手法について提案した。申請者は、本研究を通して、これまでの情報検索では考えられてこなかった関係の類似という概念を利用することで、異なるドメイン間の対応を発見する方法を提案しており、さらに、提案手法の有効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24年8月9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。