# 審 査 の 結 果 の 要 旨

氏名　マムドゥーヘ　ファルーケ　モハメド　ファルガリ

本論文は「An Approach for Semantic Search in the Web including Databases（データベースを含むWebのセマンティック検索に向けたアプローチ)」と題し，英文で記されており，７章から成る．

第１章「Introduction（序章)」では，まず増大を続けるWebデータの活用では，キーワードによる検索だけでなく，次第に意味に踏み込んだ意味的検索（セマンティック検索）が必要とされている背景を述べている．2000年代初頭からW3Cを中心にメタデータの記述を対象として，コンピュータにも意味が理解できるように標準化を計るSemantic Webの活動があったが，これは必ずしも成功しているとは言い難く，最近ではその下位層の記述形式であるRDF(Resource Description Framework: 基本的に(subject, predicate, object)の３つ組構造でsubjectとobject間の関係をpredicateで表す)を用いるLinked Open Dataとしての利用が広がりつつある．これとは別に，Webには関係データベース形式で大量のデータが埋め込まれており，コンピュータによるこの意味的検索と活用も課題となる．RDFデータや関係データベースの関係等を表わす語彙は応用領域毎に個別的に定義されるため（これをドメイン・オントロジーと称する)，異なる領域に亘る利用が困難となる課題が生じる．この課題に対して，オントロジー写像（マッピング）の研究が行われてきているものの，課題は解消されておらず，広範囲での運用を阻害する大きな要因となっている．

本研究は日本で開発されたCDL(Concept Description Language)に着目し，上記課題の解決を図るアプローチを提示している．CDLは自然言語テキスト(英語，日本語，スペイン語，中国語…などいずれの言語も対象とする)が表す概念レベルの意味（深い意味でなく表層に近いレベルの意味）を，共通的な語彙セットと44種の関係ラベルによって共通的に表す記述言語であり，これによってコンピュータが自然言語テキストの意味把握を可能にする（本研究とは直接関係しないが，異なる言語間での意味の疎通も可能にする)．領域依存といった性格を持たず，共通性があるCDLをデータ項目間の関係記述に利用することで，汎用的に意味的検索を可能にする新しいアプローチであることを述べている．

第２章は「Related Work（関連研究)」である．関係データベースのデータをRDFに変換してRDFデータとして検索を可能にするアプローチ，このRDFに加えルールを付加し推論により検索によって得られる範囲を拡大するアプローチ，検索キーワード拡張(keyword expansion)による意味的検索法などについて記している．

第３章は「Converting Relational DB to Semantic Format（関係DBのセマンティック記述形式への変換)」であり，Webに埋め込まれた関係データベース(DB)の大量のデータを，CDL記述に変換し，意味的検索を可能にする以下の方法を示している．まず関係DBスキーマにおいて，表の二つの列名（又はフィールド名）間に存在する関係を人がCDLで記述する．次いで，表の各行（タプル）の大量のデータ項目間を，この関係で関係付けたCDLデータを自動的に作成する．ここで人の関与は列名間の関係のCDL記述だけであり，この作業量は大きくない．領域に依存しない共通性を有する語彙によるCDL記述を用いることで，意味的検索を使用して領域依存のオントロジーの差異に阻害されることなく，検索・利用が可能になる．この関係データベースのデータをCDL記述に変換する支援システムDB2CDLを作成しており，この動作を実験により示している．

CDL記述はRDF形式に変換できることから，この章の後半ではルール知識をRDFデータ集合に付加し，ルールによる推論も併用して検索可能な範囲を拡大する，効率的な実現法を提示している．例えば，論文と著者についてのDBから生成したRDFがあるとすると「人物Aと人物Bが同一論文の共著者なら，AはBを知っている」，「人物Aが論文Yの著者であり，論文Yのトピックが Tなら，AはTに興味を持つ」という一般ルールを付加することにより，知人関係や興味についての情報検索も可能になる．Semantic WebやLinked Open DataでRDFの検索にはSPARQLが標準的に使われているが，ここではルールによる後向き推論でクエリ拡張を行い，これに基づくSPARQLによるRDFデータ検

索で効率的に実行する．RDF へルールを付加するアプローチはこれまでにもあったが，本研究では推論を含む検索の効率化を図ったことが新規な点となる．このシステム作成し，実験により検索の効率性を示している．

第 4 章は「Semantic Search（意味的検索）」であり，グラフマッチングを基本とする CDL データの意味的検索法を提示している．ここで意味的検索とは，完全なマッチングでなくとも，意味的に近い記述項目も含めて検索出力とすることを指す．CDL 記述のグラフ表記では，ノードは概念（名詞，動詞，形容詞，副詞等の自立語に相当）を表し，ノード間のラベル付きエッジ（又はアーク）は関係を表すことになる．この場合，CDL データ集合から構成されるグラフに対して，検索クエリで構成されるサブグラフのサブグラフ・マッチングを行うことになるが，マッチング条件を緩和するここでの意味的検索は以下のように行っている．ノート対ノードのマッチングにおいては，クエリ・サブグラフのノードの概念を，WordNet（英語のオンライン語彙辞書）での同義語，及び CDL の基本語彙階層から１段上の上位語に置き換えたサブグラフとも（不完全）マッチングを行い意味的検索を実現する．更に，ラベル付きエッジについても，44 種の CDL 関係間の意味的な近さを数値的に規定し，近い関係へ置き換えて（不完全）マッチングにより意味検索を行う．以上によっても検索出力が得られない場合，クエリ・サブグラフの意味的な重要性が低いエッジを一部削除して，条件緩和したグラフマッチングにより検索出力を得る．以上のグラフマッチングの度合いを，概念を表すノードのマッチング度，関係を表すエッジのマッチング度，クエリ・サブグラフから削除した部分の割合から算出し，意味的検索出力のランキングとしている．

第 5 章「Implemented Prototype（実装したプロトタイプ）」では，作成した上記の CDL データ意味的検索システムについて記している．実験は，Wikipedia の自然言語テキストや Web 中の関係 DB からの変換で得た 20 万以上の CDL データを対象にしている．この際，自然言語テキストから CDL への変換はまだ全自動にはならないので，不完全部分を見出して処置する編集システムの開発も行っている．Google による検索と比較し，各種のクエリに対し，本研究のシステムの方が高い検索精度が得られることを実証している．

第 6 章は「RDF Ontology Injection（RDF オントロジー注入）」であり，Linked Open Data の進展で蓄積が進んでいる RDF データを，領域依存オントロジーの差異に起因する課題を克服し異種領域に亘る活用を可能にするため，RDF オントロジー注入のアプローチを提示している．これは各個別の領域で規定された RDF の関係を記述する語彙（オントロジー）に対し，その意味を表す CDL 記述をコメントとして注入し，共通性を有する CDL 記述を介して異なる領域の RDF データ間を橋渡し，領域依存オントロジーの課題を克服する．この効果を実現するために具体的に作成したシステムでは，注入した CDL 記述を持つ RDF 関係記述語彙をインデックスとして管理し，RDF に対する検索システムである SPARQL をこのインデックスを参照してクエリ拡張し検索に用いている．そして，異なる領域オントロジーに基づく 2 種の RDF データ集合（合計で約 125 万の RDF データ）を対象にして，提案アプローチの効果を示している．

第 7 章「Conclusion and Future Work（結論と今後の課題）」では，本論文の研究成果をまとめ，今後の課題に言及している．

以上を要するに，本論文は Web 内に関係データベース形式で埋め込まれた大量のデータ，及び Linked Open Data として蓄積されつつある RDF(Resource Description Framework) 形式のデータを対象にし，領域依存の語彙体系（ドメイン・オントロジー）の差異に起因して阻害されていた異なる領域に亘るデータ検索の課題を，共通性のある語彙を持つ CDL(Concept Description Language)を利用することで克服する新しいアプローチを提示している．加えて，ルールを付加して推論も含めて得られる範囲を拡大する効率的な検索法，グラフマッチングを基本として意味的に近い CDL データを求める意味的検索法を具体的に示している．これらを実現するシステムを実装し，実験により効果を実証しており，今後の意味的検索を中心とする Web 技術の発展への貢献が認められ，創造的実践の観点からも価値が認められる．

よって本論文は博士（情報理工学）の学位請求論文として合格と認められる．