審 査 の 結 果 の 要 旨

氏名　タコア　レネヴェイ　フランシスコ　アントニオ

本論文は「Semi-Automatic Conversion of Natural Language Text into Concept Description Language (CDL)（自然言語テキストの Concept Description Language (CDL) への半自動変換)」と題し，英文で記されており，７章から成る．

第１章「Introduction（序章)」であり，まず Web 空間を中心とする情報量の爆発的増大により，次第に意味に踏み込んだ情報検索や情報利用が必要とされてきているという研究の背景を述べている．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 の活動があるが，これは領域毎の基本語彙の差異などの問題で必ずしも成功しているとは言い難い状況である．これとは異なり，本研究は Web 情報の中核的形態である自然言語テキストを対象とし，その表す表層的意味をコンピュータにも理解できる形で表現する記述言語 Concept Description Language (CDL) を基盤として，上記のような意味計算(semantic computing)を実現するために必要とされる，自然言語テキストから CDL へ効率的に変換する課題に取り組んでいることを述べている．

第２章「Concept Description Language (CDL)」では，日本で開発された CDL について説明している．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（深い意味ではなく表層に近いレベルの意味）を，共通的な語彙セット（20 万語以上）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．領域依存といった性格を持たず，共通性がある CDL は意味的検索を含む今後の広範な意味計算の基盤になり得るものとしている．本研究では立ち入らないが，異なる言語間での翻訳，意味の疎通の役割も果たすものとなる．CDL の関係ラベルの主なものは semantic role（意味役割）であるが，英語に対して採用されている semantic role とは異なり， CDL の semantic role は多言語に共通的なものを採用しており，共通的概念記述言語として好ましいことを記している．CDL は語彙や概念を関係ラベルで結んだものが基本となるが，表現形としてはグラフ表現，テキスト表現，そして RDF (Resource Description Framework) 表現がとられる．

第３章は「Related Work（関連研究)」である．まず CDL に先行して国連大学高等研究所（日本が拠点）で開発され，UNDL 財団（ジュネーブと東京が拠点）で継続的に開発されている Universal Networking Language (UNL)について記している．UNL は多言語翻訳におけるピボット言語（中間言語）を起源とし，主として Web テキストにおける多言語間の情報流通を可能にするものとして設計，開発されてきた．CDL は UNL の基本語彙と関係ラベルとを受け継ぎながら，Web 情報空間で馴染みやすく受け入れやすい記述形式を採り入れている．UNL システムでは，自然言語テキストから UNL への変換はルールベース翻訳技術に基づく自動変換機能が提供されているが，精度が十分でない問題が存在している．自動機械翻訳は進歩しつつあるものの，なお現実に要求される性能との間のギャップは大きい．なかなか超えられない壁であり，本研究で全自動でなく半自動変換のアプローチを探る背景となっている．本研究では，語義曖昧性解消(word sense disambiguation: WSD) を介する半自動変換のアプローチを採るとしている．

第４章は「The CDL Conversion Tool（CDL 変換ツール)」である．自然言語テキストから CDL の自動変換では十分な精度が達成できないことから，人手が関与する半自動変換のアプローチが必要となるが，出来るだけ人間の負担は減らす形態でなければならない．本研究では，WSD を介するそのような半自動変換のアプローチを採り，その構成法，インタ

フェースについて記している．テキストをまず標準的な構文及び係り受け解析器により解析する．次いで，第５章で記す WSD により各単語の語義同定の精度を上げる共に，次段の動詞と関連単語間の意味役割の同定を含む CDL 関係同定の精度を向上させ，人手の関与の割合を減らすことが，ここで半自動変換の主要な考え方になっている．

係り受け関係にある動詞 - 名詞間，名詞 - 名詞間（これらの単語の語義は WSP により定まっているとする）の CDL 関係を識別するのに必要な変換ルールを用意し，CDL 関係を持つ CDL 記述を生成する．WSD の誤り，ルールによる変換誤りにより，誤った CDL 記述も生成されるが，グラフ表示やテキスト表示の視覚的インタフェースを用意し，人手による修正を行い易くしている．

第５章「Word Sense Disambiguation(WSD) based on Context Expansion（文脈拡張による語義曖昧性解消)」では，第４章の半自動変換における変換性能を向上させるため，新規の WSD を考案，開発している．WSD において用いられるのは，対象単語の文脈（周囲の単語の現れ方）の利用である．まず既存の方法として，１文中に現れている各単語語義の辞書での定義文における単語の重なり度合いから正しい確率が大の語義を選択する語彙知識に基づく方法，正しい語義が付与されているテキスト集合から学習した判別関数により判定を行うコーパスに基づく方法を記している．ここでは後者の方法を採るが，CDL では語彙が付与された十分な量のテキスト集合が利用できないことから，コーパスに基づく方法では十分な精度が得られないデータ・スパースネス問題が生じる．そこで，対象語の周辺文脈に出現する単語に関係の深い単語を追加して，拡張文脈単語ベクトルを特徴に用いて判定する方法を考案している．ここで，関係の深い単語は語義が分かった単語と品詞付き単語との共起マトリックスから得る．

第６章は「Experiments（実験)」であり，考案した WSD 及び構築した自然言語テキストから CDL への半自動変換ツールの実験結果を記している．WSD に関しては，名詞と動詞を対象として行っており，拡張する単語の重みの学習は L2 正則化付きの回帰で行っている．判別関数の要素として単に１文に現れる単語とその文脈拡張単語を用いる場合と，２単語の積も 2 次特徴として用いる場合を実験している．積の 2 次特徴を用いる場合には特徴量が非常に増大してしまうため，ここでは上位 50 に制限している．実験により，積の 2 次特徴を用いる場合の方が良い判別性能が得られ，拡張文脈を用いて WSD の性能が向上できることを示している．

構築した半自動変換ツールは２名の被験者によりテストし，WSD 判定結果を確定する適切なスレッショルド値を定め，スレッショルド値以下の単語の語義を表示して人による選択に委ねることによって，全自動変換より高い精度で語義を決定でき，全人手による決定よりも短時間で決定できることを示している．また，この語義決定後のルールによる CDL 関係の決定は約 86%の精度が得られることを示している．

第７章は「Conclusions and Future Work（結論と今後の研究)」であり，本論文の成果をまとめ，今後に必要とされる研究課題に言及している．

以上を要するに，本論文は Web 空間等における意味計算基盤としての自然言語テキストが表す概念意味の共通的記述言語である Concept Description Language(CDL)に関し，自然言語テキストから CDL への精度の高い全自動変換は現状では困難であることに鑑み，文脈単語拡張による新しい語義曖昧性解消(Word Sense Disambiguation: WSD) 法を介する，人手介在が少ない半自動変換法を提案，実装し，構築した半自動変換ツールの効果を実験的に検証している．これは今後の自然言語テキスト情報の意味計算へ向けての貢献が認められ，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．