

# 修士論文要旨

開放環境科学 専攻	学籍番号 80420028	フリガナ 氏名	フカツ ヤスユキ 深津 康行
-----------	------------------	------------	-------------------

(論文題目)

## 英文の依存構造に基づく意味タグ自動付与システムの開発

(内容の要旨)

現在, コンピュータネットワーク上には膨大な文書資源が存在し, ひとびとの知的生産活動を支えている. 一方で, 誰もが自由に資源を公開できる性質上, インターネット上にはさまざまな質の資源が無数に存在し, この中から真に必要なものを見つけ出すのが困難な状況となっている. これに対し, 文書の表す内容を説明するメタデータを付加し, 一定の基準で構造化することでデータを効率的に扱うことが求められている. しかし, 大半の文書は人間にとって自然な言語によって記述されており, これをあらかじめ人が計算機可読な形で構造化することは期待できない. そこで, 構造化されていない自然言語による文書を機械処理可能な形に自動的に構造化するための技術が求められている.

近年, 自然文の最も基本的なレベルの意味を構造化する技術として, 機械学習により文中の語の意味役割を推定する方法が研究されている. 英文の意味解析においては句構造に基づく構文情報を利用することで解析精度を大幅に改善できるが, 句構造に基づく構文情報のコーパスは作成コストが大きく, 英語以外の言語ではあまり整備が進んでいない. そこで本研究では, よりプリミティブで作成コストの低い構文情報である依存構造を用いた意味解析手法を提案する. 依存構造は言語に普遍的な単純な修飾関係を表す構造のため作成コストも低く, 構造化文書の標準記述体系においても構文情報の記述方式として採用されているため, 将来的な多言語での構造化文書の普及・促進に適していると考えられる.

EDR コーパスの文例中の依存関係から語の意味役割を推定する意味解析器を **Support Vector Machine** を用いて作成し, 入力文に意味役割タグを自動付与するシステムを作成した. また, 形態素レベルのわずかな特徴量のみから依存構造解析と意味解析をブートストラップ的におこなう双方の精度を高める解析手法を考案した. 評価実験の結果, 他のコーパスにおける既存の意味解析システムと同程度の解析性能が依存構造という比較的単純な構文特徴のみで実現できることを示した. また日本語文についても同様の意味解析をおこない, 既存研究と同程度の良好な解析精度を得た.

この結果から, 言語依存性の低い構文特徴である依存構造でも十分な意味解析が可能であることを示し, 簡単な構文特徴である依存構造を基盤として多言語にわたって意味解析システムを構築し構造化文書の普及促進をおこなえる可能性を示した.

## Summary of M. Sc. Thesis

School	Student No.	SURNAME, First name
Science for Open and Environmental Systems	80420028	FUKATSU, Yasuyuki
Title Semantic Tagging System based on English Dependency Structure		
<p>Nowadays, enormous document resources exist on computer networks and help people's intellectual production activity. While, because everyone can publish their own resources easily, there are too many documents on the Internet and it comes difficult for a user to find documents he/she really wants. To solve this, it is required that documents be structured by adding some kind of "metadata" to explain the contents. But, most documents are written in natural language and difficult to be added metadata beforehand by the author. So, technology to structure documents automatically is required.</p> <p>Recently, machine learning based semantic analysis of sentence (Semantic Parsing or Semantic Role Labeling) makes rapid advance. In English Semantic Parsing, phase-structure based constituency tree of sentence improves estimation accuracy considerably. But constituency tree corpus has large cost to provide and is not ready in other languages. So, this paper proposed dependency-structure based semantic parsing method. Dependency is primitive syntactic structure of language, so dependency tree has less production cost than constituency tree. By the same token, dependency structure is adopted for a common standard of multi-language document annotation. Because of this, if it is possible to analyze semantics of sentence using dependency tree instead of constituency tree, you can easily construct a semantic parser.</p> <p>In this paper, I invented a machine learning based semantic tagging system using sentence dependency-structure. The system uses Support Vector Machines, learned from dependency relation of sentence in EDR corpus, for estimation of word-by-word semantic relations. Furthermore, I proposed a bootstrap method of sentence parsing that analyzes dependency-structure and semantic relation alternately. These two analyses boost their performance each other. As a result of evaluation experiment, proposed systems achieved comparable accuracy with existing semantic parsing systems that use phrase-structure based rich syntactic features. Additionally, I applied this method to Japanese semantic analysis and achieved fine performance.</p> <p>These results indicate that it is possible to analyze sentence semantics only using dependency structure that is less language-sensitive. And I can show the prospect that dependency-structure based semantic analysis can accelerate multi-language semantic structured document making.</p>		