「高度メディア社会の生活情報技術」
平成12年度採択研究代表者

辻井　潤一

（東京大学大学院情報理工学系研究科　教授）

「情報のモビリティを高めるための基盤技術」

１．研究実施概要

平成12年度は、今後のプロジェクト遂行のための環境整備を中心に研究を進めた。プロジェクトの中核部を担う言語処理Gは、従来からの研究成果を整備したソフトウェアツール群と必要となる言語資源の整備、ソフトウェアGは、WEB中に分散して存在する情報群に関する情報収集を行う計算機システムのハードウェアの整備とソフトウエアの基本設計、オントロジー Gは、既存のオントロジー構築ツールの網羅的なサーベイ、エージェントGは、実世界での対話データ収集のためのシステム整備を、それぞれ行った。

２．研究実施内容

本年度は、研究実施期間が３ヶ月と短期間であったために、各グループとも今後の研究のための調査、思考実験、予備的実験を中心に研究を行った。具体的には、⑴本プロジェクトを遂行していくための研究設備とシステムを整備すること、⑵これまで各グループが行ってきた研究の成果を本プロジェクト遂行のために整備すること、⑶いくつかの基盤的な要素技術の研究・開発を中心に行った。

以下に各グループごとの研究実施内容を示す。

Ⓐ言語処理 G

平成12年度は研究開始の初年度であり研究実施期間が４ヶ月と短期間であったことと、本研究グループの辻井が、平成12年度終了の日本学術振興会・未来開拓プロジェクトのリーダをしていたこともあり、本年度は、以後の研究を行うためのツール群、および、言語資源の拡充を外注業者を使うことによって行った。具体的には、WWW上に存在するテキストに対し、構文解析・情報検索・情報抽出を行うための基礎的な自然言語処理技術の開発として、⑴自然言語を解析するための文法獲得の研究、⑵特定領域におけるオントロジーの記述枠組の研究、⑶文情報、意味情報等の情報を含むコーパス・リソースの作成、および作成されたコーパスからの情報抽出を容易に行うシステム（TIMS）の作成を行った。まず、⑴に関して、大規模かつ詳細な記述が与えられている英文法として有名なXTAG英文法を、我々が研究対象としているHPSG文法に変換する技術を開発した。この

技術により得られた英語文法はおよそ30万の語彙項目からなり、これをベースにして意味表現を含むより詳細な文法に拡張していくことで広範かつ詳細な文法が開発できると期待される。⑵に関して、意味表現としての生物学分野のオントロジー体系を包摂関係によりとらえる予備的実験を行った。生物学の知識は記述的表現が多く、シンボルによるオントロジー体系記述の良い例題といえる。これらの予備的実験を通して得られた知見をもとにより一般の知識体系の記述へと拡張できることが期待される。⑶に述べたTIMSは、XMLをベースとする知識データベースであり、前述のオントロジー表現を含む知識処理、情報検索、情報抽出を統合して実現するための核システムであり、今後の研究の土台として利用可能な状態になった。

Ⓑソフトウェア・広域分散システム G

ソフトグループは、並列分散情報収集ソフトウェアの設計を行った。設計目標として、以下の４項目をあげる。⑴並列かつ分散してWEBページの収集・索引付けを行う。⑵収集・索引付けに参加するノードを動的に変化させることができ、その際に全体の動作をとめることがない。⑶広域分散環境で、局所性および負荷分散を両立する。⑷収集された大量の生データを、モバイルプログラムを用いて効率良く公開する。

本年度は、各項目に関して、設計や予備的な実験を行った。⑴、⑵に関して、動的なノードの増減をサポートする並列計算モデル（Phoenixモデル）の設計と、予備的な実装を行った。ノードを増減させながら、メッセージ通信に基づく並列計算を、連続的に動作させることができた。また現在、Phoenixモデルよりも高水準な、共有オブジェクト空間に基づいた仮想機械の設計を進めている。また、Phoenixモデルを用いたWEBページの並列収集器の設計を進めている。

⑷に関して、モバイルプログラムによるストレージサーバアクセスのためのAPIを設計し、モバイルプログラムによって処理が高速化することを、予備実験を通して検証した。

Ⓒオントロジー G

オントロジー工学グループでは、協調的オントロジー構築に基づいたオントロジー・アラインメント手法の開発のため、⑴オントロジーの協調的構築およびアラインメント過程の分析、および⑵オントロジー・アラインメントの半自動化のためのツールの開発に関する準備を行なった。⑴では、GMDで開発されたConcept Indexプロトタイプを研究実施場所に実装するためのサーバおよびデータベースの整備を行い、システムの日本語対応化に着手した。また、文書群からユーザが自由選出したキーワード群から共同作業として単一の概念体系を構築する場合と、各ユーザが予めそれぞれ概念体系を構築し、それを融合する場合を比

較する予備調査を行った。これによると、前者では一般に見られるような特徴のない概念体系が構築されるのに対し、後者ではユーザにより場合によっては大きく異なる概念体系が構築され、完全な融合は困難であるという予備結果が得られた。これは今後さらに分析・追加調査を進め、それぞれにおける協調作業の特性を明らかにすることが課題となる。⑵では本年度はアラインメントや融合を試みる半自動化ツールの既往研究のうち、各概念のラベルや属性の形式の類似性を利用する手法を分析し知見を得た。これらはいずれもオントロジーとして抽象化された段階からの処理を想定しているため、それらが生成された際の文脈の情報が十分利用できていない。そのため文書からのオントロジー構築のように文脈が利用できる場合におけるアラインメント手法の調査・分析を行い、ボトムアップ的にオントロジー・アラインメントを行う手法を検討することが課題である。

⒟エージェント・対話 G

実世界の状況下における身体性を伴う多様な会話による情報モビリティの実現の研究を行うため、テレイグジスタンスと自律行動の両方の機能をもつハイブリッド型実世界エージェント実験システムの装置設計を行った。この実験装置は、図１のごとくマスタ・スレーブ型の構成になっている。スレーブ側のセンサー情報はマスタ側に無線によって伝達される。この情報に基づいてマスタ側からスレーブ側を遠隔操作できる。操作者の表情や声といったマスタ側の情報をスレーブ側に伝えることもできる。さらに、スレーブ側からマスタ側に送られる情報をスレーブ側に搭載されたコンピュータに送り、画像処理、音声処理、自然言語処理やプランニングを行うことによって、スレーブ側がある程度自律的に行動できるようにすることもできる。本装置は、平成13年６月に製造・納入予定である。

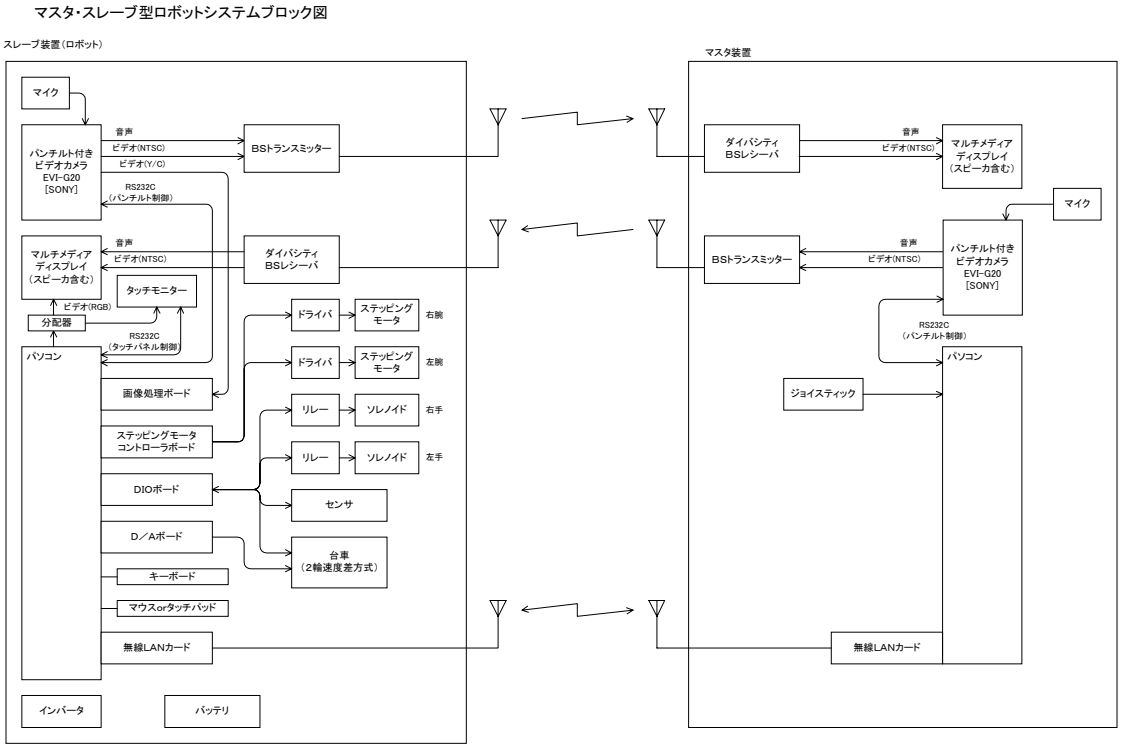

図１：実世界エージェント実験システムの構成

実世界エージェント実験システムを用いて、身体性と社会性を考慮に入れた言語表現のグランディングに関する研究の具体化について検討を行った。その結果、次のような取り組みが有望であることが明らかになった。⑴実世界エージェントを用いて、離れたところにいる人間がさまざまな状況下で会話する実験を行い、会話データの収集と分析を行うことによって、人間が身体性・状況・社会性などを考慮に入れてどのように発話生成と理解を行うか、基礎データを得る。⑵マスタ側の操作ログから実世界エージェントがコミュニケーション行動のモデルあるいはその一部を自動生成する。⑶実世界エージェントに感情モデルやソーシャルインタラクション機能を組み込んで、人間社会のなかに自然に溶け込めるようにする。⑷人工物のどのような特性が人間らしいしぐさや振る舞いとみなされるか、また、人間らしさが情報のモビリティにどのように関係するかを解明する。

以上と並行して、従来から研究開発を行った分身エージェントシステムの会話制御方式を参考に、実世界エージェントのメモリモデルの検討を行い、連想表現を用いた分身エージェントの記憶方式と話題の連鎖による発話会話方式が有望であることを明らかにした。

３．主な研究成果の発表（論文発表）

特になし