

- 音声・自然言語に関する研究開発プロジェクト
MASTARプロジェクト発足のご案内
～ 一体的に音声・自然言語の研究開発・成果展開を行う拠点の構築に向けて ～

- 2008年5月9日

独立行政法人情報通信研究機構(以下「NICT」という。理事長:宮原 秀夫。)は、平成20年4月、内閣府の総合科学技術会議が選定した社会還元加速プロジェクト「言語の壁を乗り越える音声コミュニケーション技術の実現」の研究を開始し、併せて、機械翻訳、音声対話、言語資源などの音声・言語資源、処理を統合的に研究開発し、持続的な成果展開を推進する新しい枠組みであるMASTARプロジェクトを発足させました。(MASTAR:Multi-lingual Advanced Speech and Text reseARch)

現在、NICTではユビキタスネット社会の実現に向けた総務省の「u-Japan 政策」に基づき、人に優しいICT を介して、すべての人と人との時間や場所などの置かれた条件を問わずに交流でき、新たな「知」や「価値」を産み出せる社会を構築していくための「ユニバーサルコミュニケーション技術」の研究開発を平成18年から推進しています。

さらに、この度、2025 年までを視野に入れたイノベーションの創造のための内閣府の長期戦略指針「イノベーション25」のテーマの一つであり、イノベーションをいち早く社会に還元することを目指した内閣府の総合科学技術会議が選定した社会還元加速プロジェクトである「言語の壁を乗り越える音声コミュニケーション技術の実現」のための研究開発を開始しました。

そこで、多言語音声翻訳、機械翻訳、音声対話などの音声・言語処理を統合的に研究開発し、成果展開を推進するため、新たに、けいはんな研究所内に「MASTARプロジェクト」(リーダ:中村 哲 上席研究員)を発足させました。このプロジェクトでは、産学官が連携しながら、国内外の言語資源、言語翻訳、音声コミュニケーションの研究者が集まり日本を代表する中核拠点として、横断的に多言語音声翻訳、多言語テキスト翻訳、多言語対話応答システム技術の研究開発、人材育成を行います。これまで単一の機関では困難であった音声・言語資源の収集・蓄積を行い、それらを基に、産学官が連携して実運用に結びつけながら、サービスの提供を目指します。

< 広報 問い合わせ先 >

総合企画部 広報室

室長 栗原 則幸

Tel:042-327-6923

Fax:042-327-7587

< 本件に関する 問い合わせ先 >

けいはんな研究所

事務長 高橋 幸雄

Tel: 0774-98-6810

Fax:0774-98-6955

MASTARプロジェクトの狙い

【MASTARプロジェクトとは】

- 異なる言語を話す人々との会話を音声翻訳技術により実現します。特に、ネットワークを使うことで、場所や話題を拡張し、種々の地名、レストラン名などの固有名に対応し、個人に適応した高度な音声翻訳を実現します。
- 1社では翻訳のための外注費用に膨大な費用が必要であったマニュアル・社内文書の翻訳を、MASTARプロジェクトを軸に類似分野の複数の会社が辞書やコーパス*を共有できるようにし、機械翻訳を用いることで、より安価で早い文書翻訳を可能にします。
- 商品の問い合わせなどに関するコールセンタを自動音声対話応答システムにすることにより、24時間に渡るより多くの顧客対応と、さらに経費削減が実現できます。また、自治体などへの多言語の問い合わせ、時間外の問い合わせなどにも対応することが可能になります。
- 機械翻訳、音声翻訳、情報検索、情報要約、教育などの高度な情報サービスの基本となる辞書などの言語資源が、日本では世界的にみて不足しています。そこで、世界規模の言語資源と、全く新しい日々最新の言語資源を標準時配信のように電子的に提供するサービスを開始します。
- 産業界から研究者、技術者を受け入れ、幅広い人材育成を行います。

【具体的な研究開発目標】

1. 総合科学技術会議の社会還元加速プロジェクトの一つに選定されたネットワーク音声翻訳に関する技術開発、社会実験、社会還元を行う。
インターネットを利用し固有名の登録、インターネットを介して翻訳の学習に必要な対訳文を収集、場所、場面に応じた音声・翻訳知識をWEB2.0型で収集、学習、切り替えを行う技術を研究開発し、実際の観光地、自治体などの音声翻訳サービスとして、大規模実証実験を行う世界初のネットワーク型音声翻訳の試みを行う。
2. 産業界とマニュアルなどを対象としたWEB2.0型の機械翻訳サービスを行い、共通辞書、共通コーパスの蓄積、翻訳技術研究のポジティブ成長サイクルを確立する。
マニュアル、自治体の情報のテキスト翻訳を対象に、機械翻訳→人手修正→言語資源蓄積→技術改良→機械翻訳というループを持続的に回し、技術開発を進める全く新しい自然言語処理研究のアプローチを実現する。
3. ユニバーサルコミュニケーションの一環として、自治体などを通じあらゆる利用者へ情報を届けるための音声対話インタフェース技術の技術開発、社会実験、社会還元を行う。
京都における観光情報をはじめ、自治体の情報サービス、自動応答コールセンタ技術として、音声対話→人手書き起こし→言語資源蓄積→技術改良→音声対話の持続的ループを回し、技術開発を進める世界初の自治体多言語音声対話応答サービスの試みを行う。
4. 世界的言語資源を構築し社会に還元する。
高度な情報サービスを可能にする知識／コンテンツまでを大量に含んだ日本語辞書、類語辞書、多言語辞書、さらには種々のコーパスを半自動で作成、日々更新し、電子的にインターネットで配信する世界初の電子的多言語言語資源構築・配信の試みを行う。

(注*) コーパスとは名詞や動詞などの言語的な情報を付与した音声・言語データベースを指す。

