

高信頼性組み込みソフトウェア構築技術

北陸先端科学技術大学院大学 情報科学研究科 片山卓也

研究分担者 早稲田大学理工学部 コンピュータネットワーク工学科 中島達夫

研究分担者 京都大学 湯浅太一

研究分担者 国立情報学研究所 中島震

協力企業 NEC, NECエレクトロニクス, 松下電器産業, Nokia, オムロン, オムロンソフトウェア

1. プロジェクトの背景

組み込みソフトウェアは、家電製品、自動車、携帯端末、制御機器などのあらゆる工業製品の心臓部に組み込まれ我々の生活を支えており、その機能や品質はこれら製品や機器の価値を決める最も重要な要素である。特に、ソフトウェアの不具合による経済損失などが報告される現在、高信頼性は組み込みソフトウェアに対する最も重要かつ緊急な要求である。

従来、組み込みソフトウェアはそのサイズが余り大きくなかったこと、また、比較的単純な機能の実現を行えば良かった事などあって、その開発には最新のソフトウェアテクノロジーが用いられてこなかった。しかしながら、現在では、高度なユーザインタフェースや通信機能など製品に要求される機能が高度化すると同時に、利用可能なCPUやメモリなどのハードウェア資源に対する制約が緩和したことなどによって、組み込まれるソフトウェアが大規模化・複雑化し、これまでのソフトウェア開発方法論が十分に機能しなくなりつつある。最新のソフトウェア開発技術を組み込みソフトウェアの開発に投入し、組み込みソフトウェア開発を前進させることが強く求められている。

本プロジェクトでは、このような観点から高信頼組み込みソフトウェア構築の問題を、(1) 構築環境、(2) 実行環境、(3) 実行基盤の3つの視点から総合的に解決することを目的とし、企業との密接な連携のもと、プロジェクト終了時に産業界への技術転移を目標とし、以下の課題に関する研究開発を行っている。

- (1) 組み込み用オブジェクト指向分析設計技術 (北陸先端科学技術大学院大学)
- (2) 組み込みシステム向け基盤ソフトウェア (早稲田大学)
- (3) 組み込み用実時間Java技術 (京都大学)

2. 組み込み用オブジェクト指向分析設計技術

ソフトウェア開発上の問題の多くが分析・設計などの上流工程で作られてこられ、それが全体の生産性や品質に支配的な影響を及ぼしていることはよく知られている。(1)では、組み込みソフトウェアの特性を適切に扱うことが可能で、上流工程段階から正しさを確認・検証しながら組み込みソフトウェアを構築できる方法論とそれを支える環境の実現に関する研究開発を行っている。具体的には、UMLによる記述、モデル検査や定理証明技術による検証、製品系列やアスペクト概念による設計などを可能にする環境の実現を行っている。

3. 組み込みシステム向け基盤ソフトウェア

携帯電話やデジタルテレビ、あるいはこれから登場する先進情報アプライアンスにより様々な新しいサービスが提供されることが予想される。これらの次世代の情報端末や情報アプライアンスには、高度な通信や協調動作、セキュリティ、センシングなど多くの機能が要求されるが、(2)では、これらを実現する高信頼コンポーネントやミドルウェア、それらを動作させるオペレーティングシステムなどの基盤ソフトウェアの研究開発を行っている。

4. 組み込み用実時間Java技術

組み込みソフトウェアでは、一定の限られた時間内にレスポンスを返すこと、また、限られたメモリで動作することなど、動作時間や利用可能資源などに厳しい制約がついているのが普通である。従来はプログラマが細心の注意を払って、職人芸的技術によってこの問題を解決しており、開発コストおよび品質の両面で大きな問題となっていた。(3)では、これらの作業を最大限に自動化し、信頼性の高い組み込みソフトウェアの生産性を高めるための研究開発を行っている。

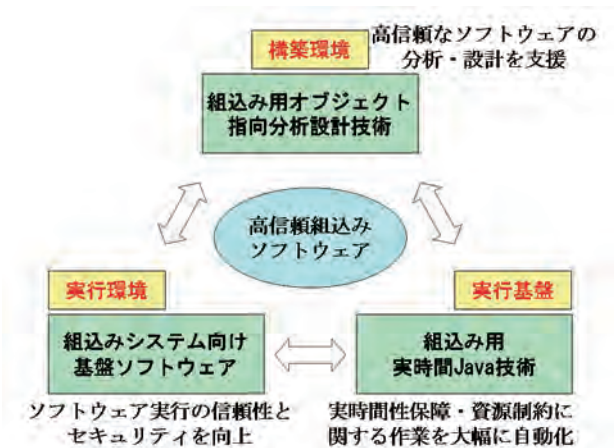


図1 プロジェクトの全体像

組み込み用オブジェクト指向技術

北陸先端科学技術大学院大学 情報科学研究科 岸知二, 青木利晃, 片山卓也

国立情報学研究所 中島震

協力企業 NEC, NECエレクトロニクス

<http://kt-www.jasit.ac.jp/project/esociety>

1. プロジェクトの概要

ソフトウェアの問題の多くは上流工程で作られており、産業界でも設計品質に対する問題意識が高まっている。本プロジェクトでは高信頼な組み込みソフトウェア開発のための分析・設計の手法や環境を研究している。

研究にあたっては、ソフトウェア開発に関する最新の工学的、科学的成果を、実際の組み込みソフトウェア開発に適用することを試みている。例えばオブジェクト指向開発、アスペクト指向開発、プログラクライン開発といった工学的成果や、宇宙・航空・軍事といった分野でのみ利用されてきた高度だが高コストで利用の難しかった形式的手法やシミュレーションなどの科学的成果を、家電製品のような民需分野での組み込みソフトウェア開発へ適用することを目指している。

具体的には、形式的手法のひとつであるモデル検査技術と呼ばれる技術を活用し、UMLを用いた設計の正しさを検証する手法について検討を進めており、今年度はUML設計を行うためのツールの実用性を高めるとともに、それを活用した事例研究、さらに組み込みソフトウェアの検証のために、周期イベントに基づいた並行タスクの振る舞いの検証法や、モデル検査技術のリアルタイム・スケジューリングへの応用について検討した。

2. 2006年度の成果

以下に本年度の主要な成果について報告する。

2.1 UML 検証ツールの開発

昨年度までに開発したベータバージョンとその評価に基づき、ツールの第一版を開発した。開発においては過去のツールの評価結果から、UMLでの設計を行いながら検証を行う一連の作業手順やプロセスの整理を行い、以下のような典型的な検証を行いやすくする機能を強化した。

- ・インクリメンタルな検証: モデルの修正・拡張毎に過去の検証項目を回帰検証する。
- ・シナリオベースの検証: 外部環境をモデル化し、特定のシナリオ下での性質を検証する。
- ・範囲を指定した検証: システム全体だけでなく、その構成要素や部分に対して検証項目を設定して検証する。

またこれらの機能を体系づけるために、設計モデル、検証モデル、各種の指定や定義を体系的に管理するための管理モデルを明らかにし、それに基づきツール機能を関連づけた。現在本ツールはいくつかの企業や大学において評価を行っている。

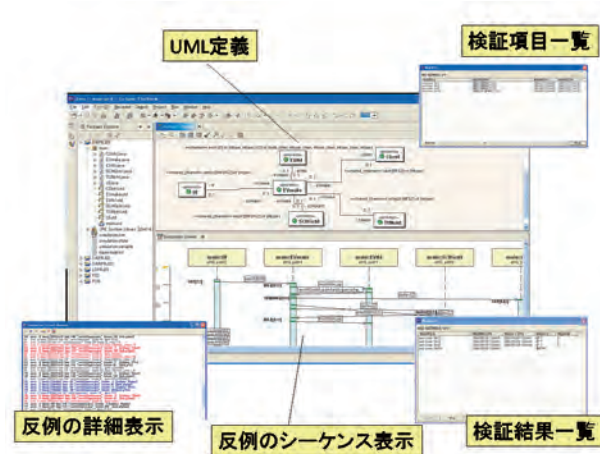


図 1 UML 検証ツールの利用画面

2.2 事例研究

開発したツールを活用しながら、複数のメーカーや研究機関と組み込みソフトウェアの検証について事例研究を進めた。これによって、複数の異なった分野における設計検証の問題を具体的に知ることができ、またそのために必要な設計検証の技術や枠組み、ツールの機能などに関して貴重な知見を得ることができた。

2.3 周期イベントに基づいた並行タスクの振る舞いの検証法

マルチタスクソフトウェアでは周期やデッドラインといった時間に関する性質が重要であり、これまでに提案した手法では扱えなかった。そこで、時間経過について厳密に記述するとともに、周期の取り扱いを可能とするために周期遷移のトリガとなるイベントに周期を割り当てることができるようにした。この手法をCDプレーヤの事例に適用し、アプリケーション側とドライブ側の状態が一致しているかどうかについて、周期を考慮した検証を行い、いくつかの問題を検出した。

2.3 リアルタイム・スケジューリングへの応用

組み込みソフトウェアで重要となるリアルタイム・スケジューリングの問題を扱うため、論理的な振る舞い仕様をベースとし、時間特性を付加したタスク振る舞い記述を対象として、モデル検査法を用いたスケジュール生成の方式を検討した。この方式により「優先度逆転の現象」が検知できること示し有効性を確認した。

組込みシステム向け基盤ソフトウェア

早稲田大学 理工学術院 中島達夫

<http://www.dcl.info.waseda.ac.jp/>

筑波大学 システム情報工学研究科 追川修一

<http://www.real.cs.tsukuba.ac.jp/>

1. はじめに

組込みシステムは今後益々複雑となり、システムの信頼性の向上はより困難な課題となっていく。特に、現状の組込みシステムでは、使用する組込みOSも μ ITRON仕様に基づく比較的単純なものからLinux等のより複雑なものを使用するようになってきている。しかし、既存のRTOS上のソフトウェアをLinuxに移行する場合、それらのソフトウェアを単純に実時間タスクとして実行すると、従来のタイムシェアリングに基づくスケジューリングを前提に作られたアプリケーションの応答性が極度に低下する可能性がある。

また、多くの人に魅力的と思われるアプリケーションは既に提供されてきているため、今後は、各個人の好みに応じてパーソナライズしたアプリケーションの提供が重要となる。その場合、ユーザが気に入ったアプリケーションを必要に応じてダウンロードできるようになる。各アプリケーションのリソース使用量をあらかじめ見積もることが困難なため、あるアプリケーションがリソース使用量を間違えるとシステム全体の動作が不安定となり、使用することが不可能となる可能性も考えられる。そのため、特にダウンロードすることにより後から追加したアプリケーションが使用するリソース量を制限することにより、はじめからインストールされているアプリケーションの動作が不安定にならないようにしないといけない。

本年度は、以上の問題を解決するため、以下に述べる2つのシステムの開発をおこなった。1つ目のシステムは以前から開発を続けているアカウンティングシステムである。2つ目のシステムは、アカウンティングシステムの応用の1つでもある仮想周期実行システムである。

以下、2つのシステムの概要を紹介する。

2. アカウンティングシステム

アカウンティングシステムは、アカウンティングオブジェクトという抽象化の提供をおこなう。アカウンティングオブジェクトは周期と実行時間の2つのパラメータを持つ。アカウンティングオブジェクトをバインドされたプロセス群は、それらが周期毎に決められた実行時間のみを消費することを保証する。

アカウンティングシステムの応用としては、ダウンロードしたプログラムが規定した以上のCPUキャパシティを用いないようにすること、リアルタイムクラスと非リアルタイムクラスのプログラムを柔軟に調停すること、システム全体のオーバロードを検出すること、リアルタイムプログラムのためのCPUリソース予約などが考えられる。

アカウンティングシステムは、極力Linuxカーネルを変更しなくても使用できるようにデザインされている。そのため、Linuxカーネルは頻繁に変更されるにもかかわらず、アカウンティングシステムの

維持を容易におこなうことが可能となる。

3. 仮想周期実行システム

仮想周期実行システムは、長時間実行される実時間スレッドが高い応答性を必要とするユーザインタフェースなどの非リアルタイムスレッドに与える影響を最小限とするためのものである。現状の組込みシステムでは、既存のリアルタイムアプリケーションをLinux上に移植したのもも多く、その場合、非リアルタイムアプリケーションの応答性が大きく低下する可能性が考えられる。仮想周期実行システムは、実時間スレッドの周期を強制的に短くすることにより、以上の問題点を解決する。つまり、長時間実行される実時間スレッドを強制的に短い周期でブロックすることで、非リアルタイムアプリケーションがスケジューリングされる可能性を増大させる。

仮想周期実行システムは、2.1で述べたアカウンティングシステムを利用する形で実現されているため、本システムを導入することによるカーネルの変更は極めて少ない。

4. 今後の展開

我々が現在開発したアカウンティングシステムは、日本エンベデッドリナックスコンソーシアムのリソースマネジメントワーキンググループにおける標準化のリファレンス実装として用いられている。また、CE Linuxフォーラムでも実際の家電機器における使用に関して検討が進められている。また、来年度以降は、メモリリソースのアカウンティングについても検討をおこなっていききたい。さらに、最終年度の成果として他のグループの成果との統合を検討していきたい。



組込みシステム用実時間 Java 技術

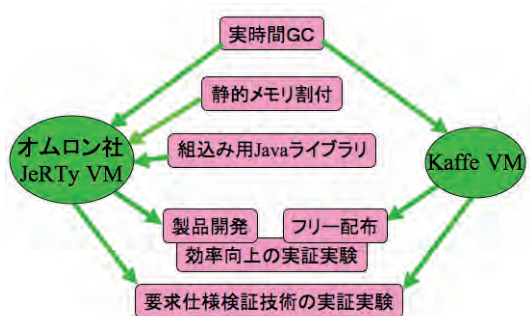
京都大学 大学院情報学研究所 湯浅太一

<http://www.yuasa.kuis.kyoto-u.ac.jp/>

1. 概要

本プロジェクトは、オブジェクト指向言語 Java によって記述される組込み実時間アプリケーションの開発を効率化するための諸技術を開発するものであり、実行基盤の開発、実証実験、要求仕様検証技術との統合、の三つのサブテーマからなる。

実行基盤の開発については、実時間組込みソフト固有の開発コストを軽減するための、自動化機能を備えた Java 処理系実装方式を開発し、試験的実装を行ってきた。具体的には、不要データの回収処理によるアプリケーション実行の中断を回避するための実時間 GC や、限られたメモリ空間を有効に利用するための静的メモリ割り付け、これらに対応できる基本的な組込み用ライブラリ群を、オムロン社の組込み用実時間 Java 処理系である JeRTy VM に試験実装するとともに、フリーの Java 処理系である Kaffe VM にも実装し、公開してきた。さらに、実証実験および徵求使用技術との融合の一環として、プロジェクトで開発したライブラリを中心に、実際にいくつかのソフトウェアについてモデル検査を試験的に実施した。



平成 18 年度は、これらの成果を、携帯電話機に代表されるモバイル端末に適用し、実用化をさぐる活動を行った。

2. モバイル端末における Java VM

携帯電話機でも高度な Java アプリケーションが利用できる時代になってきたが、そのようなアプリケーションの多くはゲームであり、実時間性に対する要求が強い、メモリ空間が数百 KB から数 MB、さらに数十 MB と拡大するにつれて、携帯電話機の Java アプリケーションはさらに高度化する傾向にある。従来は、クリティカルな場面で GC が起きないように、ベンダは、アプリケーションの「適切な」箇所に、GC を強制起動するためのコードである `system.gc()` を手作業で挿入していた。挿入箇所は端末機のモデルに依存し、このために開発コストの増大、開発期間の長期化、人材不足といった問題が生じていた。実時間 GC の実装によってこの状況が解決できることは明らかである。

モバイル端末における Java VM としては、Sun Microsystems 社が提供する KVM/CLDC が主流である。KVM とは、モバイル端

末に適した超小型 (KB クラス) の VM であり、CLDC とは、モバイル端末に適した想定実行環境 (configuration) を、コンパクトな基本的ライブラリによって提供する。この KVM/CLDC が、多くのモバイル端末における Java VM のベースとなっており、特に我が国では、携帯電話キャリア三社すべてがこの VM をベースとして採用している。そこで、本プロジェクトでは、KVM/CLDC に実時間技術を実装することによって、モバイル端末における実用性を検証することとした。

3. 実装の概要

KVM/CLDC は、一括マーク・スイープ方式に圧縮オプションを加えた GC を採用しており、1 回の GC 処理は、ルート走査、マーク、スイープの三つのフェイズからなる。これらの各フェイズを、これまで開発してきた実時間技術を適用して細分化し、実時間化を達成した。本格的な実装は現時点でも継続して作業中であるが、プロトタイプの実装はきわめて少ないコード量によって実現でき、わずか 1 週間を要しただけであった。まず PC 上でテスト及び性能評価を行い、より携帯電話端末に近いプラットフォームとして、ARM アーキテクチャを採用した PDA に移行して性能検討を行った。PC から PDA への移行は単に make を起動するだけであり、本実時間技術がプラットフォームに依存しないものであることがまず確認できた。最終的には実際の携帯電話端末で実行することが是非とも望まれるが、一般ユーザがこれを行うことは不可能であり、今後キャリア等の協力を仰ぐ必要がある。

4. 性能評価

実際の携帯電話で利用されているゲームプログラム等を実行するには、KVM/CLDC に加え、GUI などの機能を提供する高度なライブラリが必要であり、今回は性能評価に用いることができなかった。そこで、メモリを大量消費する Lisp ベンチマークを使って性能評価を行った。結果、最大停止時間を約 1/2000 に短縮することができ、アプリケーションの実行時間にもあまり大きな性能低下はみられなかった。現在作業中の本格的な実装の完成を待って、詳細な性能評価を行う予定である。

プログラム自動解析に基づく高信頼ソフトウェアシステムの構築技術 —高信頼プログラミング言語と高信頼プログラム開発環境—

東北大学 電気通信研究所 大堀 淳

協力企業 沖電気工業(株), 算譜工房(有)

http://www.pllab.riec.tohoku.ac.jp/

1. プロジェクトの背景と目的

今実現しつつある高度情報化社会が、従来通りの信頼性と安全性を確保しながら発展していくためには、社会基盤としての高信頼ソフトウェアを効率よく構築する技術を確立する必要がある。特に以下の2点の実現が求められている。

1. プログラムの誤りを開発の初期段階で検出し高信頼プログラムの効率的生産を可能にする技術,
2. C, Java等の既存言語で書かれたプログラム部品を利用し, 高信頼プログラムを効率よく開発する技術.

これら機能の実現には、従来のソフトウェア開発技術を越えた新たな技術革新を必要とする。その最も有望な基盤が、型理論に基づくプログラムの自動解析技術である。この技術は、システムに甚大な障害を引き起こす恐れのあるプログラムの不整合をコンパイラの段階で自動検出することを可能にする。

本プロジェクトは、型理論を基礎とし、上記2つの機能の実現の中核となる

1. 次世代高信頼プログラミング言語、及び
2. 次世代高信頼プログラミング言語をCやJavaなどの既存言語とともに使いこなすためのプログラム開発ツール

を実現し、さらにそれらを統合し、高機能高信頼システムの効率的な生産を可能にする高信頼ソフトウェア開発環境の実現を目指す(図1参照)。

これにより、ソフトウェア生産のボトルネックとなりつつあるテスト、デバッグ、保守コストの大幅な削減、さらにプログラムエラーに起因する壊滅的なシステム障害のリスクの大幅な低減が実現できると期待される。本プロジェクトでは、我々が保有する、世界的にみても最先端の型理論と型主導コンパイル技術を基礎としてこれら技術を実現する。

2. プロジェクトの具体的目標

以上の目的を達成するために、本プロジェクトでは、以下のソフトウェアを開発する。

1. 次世代高信頼プログラミング言語 SML#.

本プロジェクトの中心は、次世代高信頼プログラミング言語 SML# の開発である。SML# は、既存の高信頼言語である Standard ML を、我々の基礎研究の成果を取り入れて拡張した次世代言語であり、レコード多相性やランク1 多相性などの新しい先端機能を実現するとともに、実用上重要なCなどの既存言語とのデータの相互運用性を関数型言語として初めて実現する。本プロジェクトでは、SML# の言語仕様、コンパイル理論とコンパイル方

式、コンパイラ、実行時処理系のすべてを開発する。

2. 高信頼プログラム開発ツール

SML# を実用システム開発に応用するためには、既存言語とともに SML# を使いこなし、高信頼プログラム開発を効率よく行うことを可能にするプログラム開発環境が必要である。この実現のために、以下のツール群を開発する。

- (ア) SML# のプログラミングを支援する種々のプログラミング自動化ツール
- (イ) Java で書かれたプログラム部品を組み合わせる際の不整合を自動検出するツール
- (ウ) SML# を C, Java, およびデータベースシステムなどの既存システムとともに利用するための相互運用サポートライブラリ。

これらは、ソフトウェア生産の高信頼化に直接貢献するばかりでなく、高信頼ソフトウェアの研究の新たな展開のシーズとなると期待される。特に、SML# 言語は、レコード多相性や相互運用性などの先端機能を世界ではじめて装備した言語であり、高信頼ソフトウェア開発の生産性を大幅に高めるばかりでなく、日本発の最先端高信頼プログラミング言語のリリースによる世界への情報発信と、その波及効果として、我が国の高信頼技術の世界への普及等が期待できる。

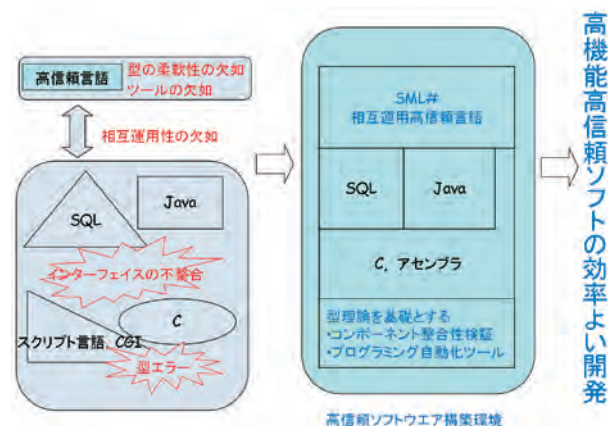


図1. プロジェクトの概要

3. 実施体制

これら先端ソフトウェアを、システム開発現場にて活用できる実用性あるソフトウェア基盤として開発するためには、ソフトウェア開発を行っている企業との緊密な連携が不可欠である。本プロジェクトでは、種々のツールの有用性の現場に即した評価や分析およびプログラム開発を、沖電気工業(株) および算譜工房(有) と共同で行う体制を取る。



4. SML# 言語

以下各項目についてその基本方式およびその実現アルゴリズムを構築し、SML# を、高信頼実用ソフトウェア開発に耐える実用言語として構築し、世界にリリースする。

1. レジスタ抽象機械
2. 抽象機械コードへの型主導コンパイラ
3. 基本ライブラリ
4. 相互運用実行時ライブラリ
5. 最適化処理
6. ネイティブコードコンパイラ

この中で1から3までおよび4の中のCとの連携機能、5の中のソースレベルの最適化が、平成18年度までに完成させリリース予定の項目、上記以外の4と5の項目および6は平成19年度に開発し、プロジェクト終了時点でリリース予定である。

平成18年度は

- (1) 幅広い型に対するシームレスなCとの連携機能
- (2) 新しい関数融合方式による最適化方式の構築と実装
- (3) 種々のプラットフォーム対応のシステムの生成処理

の各項目の開発を行い、大幅に機能アップしたSML# 言語のコンパイラ0.20版を完成させ、平成19年3月に以下のURLにてリリースを行った。

<http://www.pllab.riec.tohoku.ac.jp/smlsharp/>

(1) では、ユーザにスタブ作成やデータ変換などの処理を一切要求せず、SML# の対話型セッションから単に宣言するだけで、Cプログラムを呼び出し、コールバック関数や構造体を含めた幅広いデータを受け渡すことが可能な、他のML系言語処理系の追従を許さない強力で使い易いCとの連携機能を実現している。(2) の新しい関数融合方式の開発では、再帰的なユーザ定義の関数を直接融合可能な方式を開発しその実装実験に成功した。この方式は、世界的に見ても極めて新規性の高いものであり、2007年1月に開催されたプログラミング言語で最も権威と影響力のある国際会議ACM POPLに採録され「Lightweight Fusion by Fixed Point Promotion」の名前で発表した。(3) のシステムの生成処理では、ソースコードを共通部と環境依存部とに切り分け、依存部については、ユーザの計算機環境の特徴や利用可能な機能を自動的に判別して、必要なコードを生成する仕組みを用意した。これによってCygwin, MinGW, Linux, Solarisなどの各種プラットフォーム上で動作するSML#システムを共通のソースコードから生成することが可能となり、SML#を幅広いユーザに提供する環境が整った。

以上の3項目に加え、次年度の分割コンパイル可能なネイティブコードコンパイラ構築に向け、そのコンパイル方式および分割コンパイルの型チェック方式を検討しその概要設計を行った。

5. プログラム開発環境

SML# を、Cや、Java、データベースシステムなどの既存のシステムと共に使用し、効率的な高信頼プログラム開発を支援するツールやライブラリ等を開発する。

1. SML#プログラム開発ツール群

本年度は、SML#コンパイラのデバッグおよびSMLプログラムのデバッグを目的としてデバッガを開発した。デバッガはエミュレーションランタイムとともに動作し、SMLプログラムの任意の場所にブレークポイントを設定して、その場所でプログラムの実行を停止させ、その場所で可視な変数の名称一覧を表示させることや、そのうちの一つを指定してその値を表示することができる。

2. SML#サポートライブラリ

本年度は、日本語などの多バイト文字の処理を可能にするサポートライブラリLMLMLを開発した。このライブラリは、関数型言語の機能を生かしたストリーム処理を基本に、種々のコード体系の切り替えやユーザが新たなエンコーディングを定義し拡張できるなどの豊富な機能を備えた実用ライブラリである。

3. Javaでプログラム部品の不整合を自動検出ツール

本年度は、昨年開発した依存性解析関数ツールを実際のプログラム開発工程に適用して評価し、その性能を確認した。

6. まとめと今後の展望

より実用に近づいたSML#コンパイラの0.20版を予定通りリリースすることができた。このコンパイラは、その先進性と実用性から、学会での関心を得、2006年度ソフトウェア科学会大会および日本ソフトウェア科学会PPL2007ワークショップの依頼によりSML#に関する招待講演を行い、広く情報発信を行った。

平成19年度は、残された課題であるSML#のネイティブコードへのコンパイラおよびデータベースやJavaとの相互運用性を実現し、より高機能かつ堅牢になるSML#コンパイラを継続的にリリースし、研究成果の普及を図る予定である。

安全なシステム記述言語および高信頼 OS 記述言語

東京大学大学院情報理工学系研究科コンピュータ科学専攻 米澤 明憲
 協力企業 日立製作所
<http://www.yl.is.s.u-tokyo.ac.jp/e-society>

1. 高信頼基盤ソフトウェアの必要性

高度情報化に伴い、コンピュータウィルスや不正アクセス、情報漏洩などの問題が、社会の安全性を脅かす深刻な事態となってきた。今後、さらなる情報化は不可避であり、コンピュータを利用した社会基盤の安全性を保証することが早急に必要である。

すでに我々の周りには数多くの応用ソフトウェアが稼働しており、その多くに安全性の問題があると考えられている。それら個々の問題を解決していくことは対症療法にすぎない。問題の根本は、これらの応用ソフトウェアが、安全性の問題が社会的に顕在化する以前に開発された基盤ソフトウェア（言語、OS）を利用している事にある。

本プロジェクトでは、型理論をはじめとするプログラムの解析技術によって、既存の基盤ソフトウェアの信頼性を強化することを通して、それらを使用する応用ソフトウェア全体の安全性を向上させることを目標とする。

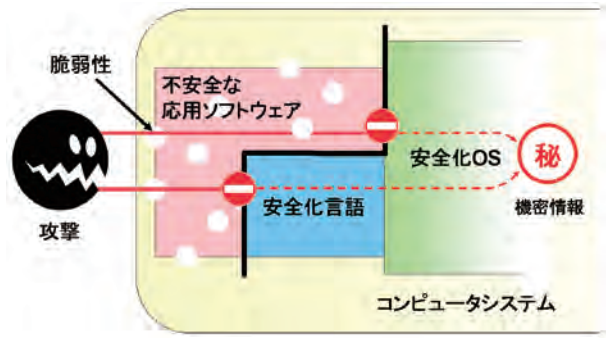


図 1 言語と OS による二重の防御

2. プロジェクトの概要

本プロジェクトでは、次の三つの方向性から上記の目標に取り組んでいる：

高安全 C 言語コンパイラの開発

C 言語はその実行速度やハードウェアよりの記述能力から、Javaなどの新言語登場後の現在もシステム設計では最も一般的であり、多くのアプリケーションにも広く使われている。

しかし、これらの C 言語で書かれたプログラムがバグにより誤動作したり、脆弱性攻撃を受けることで、機密情報を盗まれる情報被害が多発している。これは、C 言語自体にはそもそも情報の機密性という概念がないこと、また、言語によるメモリ保護機構を欠くため、メモリ管理のバグによって容易にセキュリティホールが発生することに問題がある。

この問題を解決し、既存の C 言語で欠かれた基盤ソフトウェアをできるだけ自動的に安全化するため、安全なメモリ管理と機密

情報の流れを監視する情報流解析機構を導入した、対攻撃耐性を持ったコード生成コンパイラ VITC を開発する。情報流解析により機密情報を漏洩する危険がなくなるため、VITC でコンパイルされたプログラムは、従来の攻撃にさらされても安全に実行を継続することができる。

OS 用型付きアセンブリ言語の設計・実装

高級プログラミング言語で使われていた型理論をアセンブリ言語に適用したものである「型付きアセンブリ言語」を、OS の構成要素であるカーネルのメモリ管理機構、スレッド切替機構、割込処理機構、デバイスドライバ等を記述できるように拡張する。この拡張した型付きアセンブリ言語を用いて構成要素を記述すれば、OS のメモリ安全性を保証することが可能となる。また、この手法は、通常の OS のみにとどまらず様々なハードウェアにおいて、それに密着したソフトウェアの安全性保証に適用することも可能である。

定理証明器やモデル検査器による OS 安全性の形式的検証

通常、OS 上では同時に複数のプログラムが実行され、これらのプログラムはメモリやハードウェアなどの計算資源を共有している。これため、この共有計算資源を制御し、タスク分離性（複数のプログラムが互いに不正に干渉されないこと）を保証することは、OS の重要な機能の一つである。

このタスク分離性のような、高度で複雑な安全性を OS が正しく実現しているかどうかを検証する手法として、定理証明器やモデル検査器を用いた手法を研究する。

3. 高安全 C 言語コンパイラの開発

コンピュータウィルスをはじめとするセキュリティホールへの攻撃から C 言語プログラムと機密情報を防御するためには、まず C 言語自身の未熟なメモリ管理機構に由来するメモリ脆弱性に対処できるメモリ安全 C コンパイラが必須であり、既に種々のメモリ安全コンパイラが研究、実装されている。ただし、このメモリ安全性は直接機密情報を守る手段だというわけではない。C プログラムのメモリ脆弱性は機密情報を詐取するための主な足がかりであり、防御することは非常に重要ではあるが、メモリ脆弱性がなくとも機密漏洩を起こしてしまうバグを持つ C プログラムが存在するし、またサーバなどのソフトウェアをメモリエラー発生後も継続動作させる、メモリ安全コンパイルを発展させたエラー忘却計算においても、メモリエラー後のプログラムの実行継続に予期せぬ副作用を生じ、それが機密漏洩を起こす危険性が存在する。

本研究ではメモリ安全性に続く機密保護へのより直接的なアプローチとして、プログラム内の機密情報の流れを追う、型による情

報流解析をメモリ安全コンパイル技術に組み合わせ、既存のCプログラムから、たとえ攻撃を受け不安定となっても機密漏洩を引き起こさないようなコードを生成するコンパイラVITCの作成を目指す。

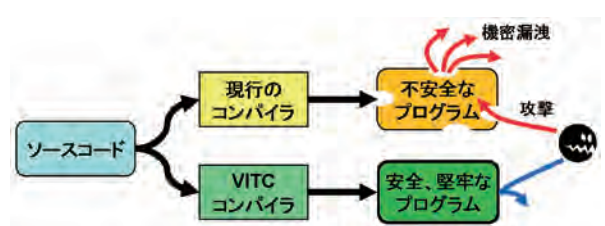


図 2 VITC コンパイラ

型による情報流解析の研究は、JavaやMLなど、理論的に厳正な静的型システムを与えることができる言語系においてもぱら行われ、情報流もほぼ静的に、プログラム実行前に解析することができた。それに対し、C言語の型システムは非常に柔軟ではあるが正確性を欠く。そのため、既存研究と同様の静的情報流解析のみではC言語の表現力を保ちつつ機密漏洩を防ぐことは難しく、実行時の情報流の動的検査が不可欠である。

本年度はVITCコンパイラの完成へ向けて、メモリ安全Cコンパイラ上に静的情報流型システムと、情報流動的検査コード生成器を作成、基本的なCプログラムに機密情報の仕様情報を付加してコンパイルすれば、情報流解析による静的、動的な機密保護を行うコードを生成することが可能となった。また、OSのファイル権限情報を機密度として扱い、保護ファイルから読み込まれた情報は自動的に機密情報として扱うことができるよう、動的情報流検査ライブラリに拡張をほどこした。これらを利用し、thttpdウェブサーバをVITCでコンパイルし、メモリ脆弱性攻撃を受けてもサーバが保護ファイルの情報を誤って漏らすことのないことを実験で確かめることができた。

4. OS用型付きアセンブリ言語の設計・実装

近年の型理論研究の進歩により、多くのアプリケーションプログラムが「強く型付けされたプログラミング言語」(例: Java, C#, OCaml, 等)を用いて既に作成されるようになっている。これは、強く型付けされた言語で記述されたプログラムは、実行時に予期せぬメモリエラーを生じない、ということが保証されるためである。

ところが、コンピュータを動作させる上で最も基礎的で重要なプログラムであるOS(オペレーティングシステム)は、未だ強く型付けされていない言語を用いて作成され続けている。このため、従来OSの安全性を保証・検証することは非常に困難であり、実際、安全性が証明されたOSは(機能が限定された非常に小さなOSを除けば)未だ存在していない。

そこで本研究では、OSの作成に利用可能な、強く型付けされた安全な言語を設計・実装することを目指す。具体的には、OSカーネルの重要要素(メモリ管理機構やスレッド管理機構など)の記述

に適した、強く型付けされた安全なアセンブリ言語を設計・実装する。これを用いてOSカーネルを記述すれば、セキュリティ脆弱性の原因となるメモリエラー(バッファオーバーフロー等)が発生しないことが保証された、安全で高信頼なOSを実現することが可能となる。

本年度は、前年度までに設計した型付きアセンブリ言語(TALK)の理論を、より多くのOSカーネルの構成要素の記述を可能とするために、CPUハードウェアの割込み機構に対応するように拡張した。また、より多くのCPUアーキテクチャへ型付きアセンブリ言語の理論を適用するために、機械語に近い低級言語の型検査器を、CPUアーキテクチャになるべく依存せずに構築するための手法を研究した。

5. 定理証明器やモデル検査器によるOS安全性の形式的検証

OSの重要な機能の一つに、複数のプログラム(タスク)間で共有される計算資源(メモリやデバイスなど)を制御する機能がある。通常コンピュータシステム上では、複数のプログラムが同時に動作しており、これらのプログラムが互いに干渉してプログラムが異常停止したり、あるプログラムの使用しているメモリが他のプログラムによって勝手に読み書きされたりしないことを保証することが重要である。つまり、OSが共有計算資源の制御を正しく行っているかどうかはコンピュータシステム全体の安全性に大きく影響する。誤ったOSの上では、たとえ個々のプログラムが正しく書かれていたとしても、システム全体の安全性は保証できない。

これに対し我々は、定理証明器やモデル検査器を用いて、既存のOSを数理論的に検証する手法の研究を行った。本年度は、複数スレッドが同時に実行されるOSにおいて、プログラム間のタスク分離性(複数のプログラムが相互に不正な干渉を行わないこと)を検証するための、モデル検査器を用いた検証手法を考案した。また、昨年度より引き続き、定理証明器Coqを用いた検証理論を考案し、実際に既存OSのヒープ管理コードのメモリ安全性の検証を行った。

データ収集に基づくソフトウェア開発支援システム

奈良先端科学技術大学院大学 鳥居宏次

研究分担者 大阪大学 井上克郎

協力企業 日立公共システムエンジニアリング, 日立製作所, NTTソフトウェア, SRA 先端技術研究所

http://www.jempirical.jp/

1. はじめに

本プロジェクトでは、エンタープライズ系、組み込み系を問わず、広くソフトウェア開発において、他の科学や工学分野と同様に、計測、定量化と評価、そしてフィードバックによる改善という実証的手法（エンピリカルアプローチ）の実践を目指す。以下では、本年度の主な成果の概要を述べる。

2. ソフトウェア開発データ分析技術の高度化

平成17年度に構築し、オープンソース化したデータ収集システムEPM (Empirical Project Monitor) 等で収集された「ソフトウェア開発データ」のより多様な分析と開発者へのフィードバックを可能とする方式の開発を目指した。具体的には、以下の点に注力した。

①ソフトウェア開発のリアルタイム管理:「プロジェクト遅延リスク検出モデル」をはじめとする、平成17年度に開発した手法に基づく分析結果をリアルタイムに開発者にフィードバックする具体的な方式とツールを開発し、ソフトウェア開発プロジェクトに実際に適用することで評価した。具体的には、フィードバックを補助するツールを開発し、ソフトウェアエンジニアリング技術研究組合(COSE)が実施するソフトウェア開発に適用した(図1)。その結果、ソフトウェアの潜在的品質リスクの指摘やプロジェクトマネージャの状況把握の支援に有効であり、プロジェクト管理方法の一般化、汎用化を進める上でも有効な手段であることが確認された。

②ソフトウェア開発データリポジトリに対するマイニング:EPMやそれに類するシステムや方式で収集された開発データリポジトリを対象として、組織横断的な分析を可能とする技術を開発した。具体的には、汎用的な開発データリポジトリに適用できるツールNEEDLEを開発し、プロジェクト管理データ(規模、品質、対象システムのプロファイル)、EPMで収集した不具合管理票、一般的な不具合管理票、を対象として、開発データリポジトリに適用した。更に、マハラノビス・タグチメソッドやSupport Vector Machineといったパターン認識技術をソフトウェア開発データの分析に応用する具体的な方法について検討した。具体的には、次のとおり。

- ・マハラノビス・タグチメソッドを品質データに適用する方法を検討し、NASA公開の品質データでその有用性を確認した。
- ・Support Vector Machineを官公庁系事務処理ソフトウェアの開発データに適用し、バグが含まれる確率の高いモジュールの判別に有用であることを確認した。

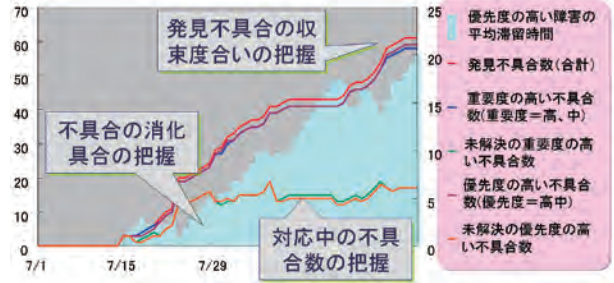


図1 COSEにおけるフィードバック例: 不具合管理票からの品質リスク推定

③ソフトウェア開発プロセス管理とプロセスベンチマーク:JIS X 0141の測定情報モデルに基づいた開発データの定義や実測値を一元管理することで、定量的プロセス管理を支援する方式を開発し、プロトタイプシステムを設計、実装した。具体的には、国内企業で用いられている定量的管理指標セットを対象に、定量的データに基づくプロセス管理のための電子ガイドブックシステムEPDG (Electronic Process Data Guidebook) システムの設計、実装を行った。更に、データ収集やそれに基づくプロセス管理がどの程度厳密に行われたかという観点から、データ分析の信頼性を評価する「プロセスベンチマーク」を開発した。具体的には、開発プロセス定義(開発に関わる作業手順)とEPMからの実測ログ(ツールの実行ログ)を突き合わせることで、プロセス定義の遵守度合いやデータ収集の品質を評価するマイクロプロセス分析手法を提案した。さらに分析ツールを実装し、商用開発データに適用し、その有用性を確認した。

3. 共有技術に基づくソフトウェア開発支援システムの開発

EPM等で収集されたデータだけでなく、平成17年度に開発した分析手法をも含めた共有技術を実現することで、解析・評価データに基づいたソフトウェア開発支援システムの構築を目指した。具体的には、以下の点に注力した。

①ソーシャルネットワーキング技術による開発知識共有:EPM等で収集、蓄積されたソフトウェア開発データに基づいて「各開発者が有する知識」と「開発者間の社会的関係」を抽出し、各開発者が必要としている知識(ノウハウ)を有する人物を推薦する方式を開発し、ソフトウェア開発における知識協創支援のプロトタイプシステムを設計、実装した(図2)。具体的には、ソフトウェアリポジトリから抽出した開発者のプロジェクト所属情報に基づいてネットワーク密度を算出し、豊富な知識を有する開発者に推薦が集中することを防ぐ、負荷分散型の知識共有メカニズムを開発した。

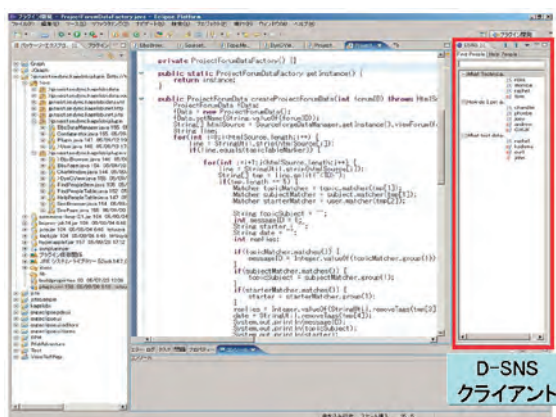


図2 知識協創支援システム画面例

②ウェブサービス技術による分析手法共有: 収集データと分析手法をウェブサービス技術を用いて公開・共有する方式を開発した。特に、収集データそのものは公開せず、分析手法と共に「サービス」という形で公開することで、データに対する匿名性の確保や権限に基づく利用制御の方式を検討した。更に、新たな分析アプリケーションの開発やそれに基づく開発支援を容易に実現できるように、収集データや分析手法を外部アプリケーションソフトウェアへ組込む枠組みについても検討し、既存デスクトップアプリケーションjbirthのWebサービス/Webアプリケーションインターフェースを開発した。

③開発者個人レベルでのデータ収集・分析プラットフォームの開発: 開発者個人レベルでのデータ収集・分析方式を開発し、プロトタイプシステムを設計、実装した。Window上で開発者個人が用いるツールである「Microsoft Office」, 「Microsoft Visual Studio」, 「Eclipse」などから開発データを収集し、個々の開発者が使用するコンピュータに蓄積することができる。また、データ収集・分析の機構を独自に拡張することを容易となっており、導入と運用に必要な工数をできるだけ小さくしている。

4. システムアーキテクチャの変遷分析ツールの開発

EPMに蓄積されたソースコードやドキュメントを用いて、過去の版に対して、現在の版のシステムのアーキテクチャはどのように変化したかを、CCFinder, diffなどのツールを用いて分析し、視覚的に表示する技術の開発を行い、プロトタイプツールの開発を行った。具体的には、次のとおり。

①変遷を表示するには、その前提となるシステムを抽象化したアーキテクチャモデルの定義が必要である。また、時系列の変遷を効果的に表示できるものが望まれる。そのため、種々のアーキテクチャモデルの比較検討を行った。そしてグラフモデルでの表現が相応しいと考え、ソフトウェアのコンポーネントを頂点、その間の利用関係を辺、そして異なる版の間の類似頂点を破線で表すグラフモデルを構築し、アーキテクチャ表現とした。この検討結果は、内部ドキュメントとしてまとめた。

②EPMのデータから定義したアーキテクチャモデルへの変換方式の検討を行った。その際、既存のCCFinderやdiffなどのツールを最大限に利用し、コンポーネント間のクローン履歴関係を求め、対応するコンポーネントを効率的に変換できるようにした(図3)。



図3 コンポーネント間のクローン履歴解析手順

③プロトタイプツールの開発を行った。既存の視覚化ツールで不十分であった開発者情報を表示する機能追加することで、クローンの履歴関係を、開発者ごとに定量的に表示し、クローンの評価を容易にすることが可能になった。

5. おわりに

本年度は、ソフトウェア開発データの分析技術を高度化すると共に、分析結果を開発者にフィードバックする具体的な方式とツールの開発を進めた。

成果の発信にも努め、学術論文や国際会議での口頭発表に加え、産業界向けに技術研究会を2回、国際フォーラムを1回、それぞれ開催し、技術研究会は各80名余り、国際フォーラムは150名余りの参加者を得た。また、先述の開発データリポジトリ分析ツールNEEDLEの利用講習会を開催し、ソフト開発ベンダを中心に13社の参加を得た。更に、EPMや分析技術を、海外に向けてより積極的に公表し、海外の研究者や企業との意見交換や技術交流を図った。一例を挙げると、インドのソフトウェアベンダ企業2社の技術者との技術検討会を開催し、9月にインド・バンガロールで開催されたNASSCOM Quality Summit 2006では招待講演を行った。

来年度はプロジェクト最終年度となることから、今年度の成果を更に発展させると共に、これまで研究開発した技術やツールを産業界へ移転する活動や組織作りについても検討していく予定である。

高信頼構造化文書変換技術

東京大学 情報理工学系研究科 武市正人

<http://www.psdlab.org/>

1. 高信頼構造化文書変換技術の概要

電子的な構造化文書情報の蓄積と効果的な情報利用技術は、インターネットを含む広範な情報の交換・流通にとってきわめて重要な位置を占めている。XMLに代表されるこれらの技術は、発展の著しいウェブによる情報環境において既存の技術の延長線上で実務的に開発されたものであり、事実上の標準となつてはいるがその言語的な概念が十分に整理されているわけではない。このような体系的な処理技術の欠如に起因する問題が情報交換の発展を阻害している。すなわち、既存技術の使い回しによる姑息な対処や人手による個別対応が、信頼性および可搬性に欠ける文書情報を蓄積する一因となっているのであり、この問題を早急に解決することが重要な課題となっている。

本プロジェクトの中核となるProgrammable Structured Document (PSD) は、プログラムの記述を含む文書を対象とすることで、構造化文書処理を効果的に実現しようとするものである。すなわち、PSDにおいては構造化文書をプログラミングにおける構造化データとみなし、プログラミング言語の理論的基盤を適用することによって、安全かつ信頼性の高い処理を実現する。構造化文書はプログラミングにおけるデータ構造と共通の性質を有しており、その処理を行うアルゴリズムの記述には関数型言語が適している。また、処理を実現するプログラムは対象文書に埋め込まれており、これによって文書の高い独立性と可搬性が実現される。

2. PSDの基盤技術

PSD実現のために必要となる基盤技術としては、(i) 構造化文書に必要な準構造データ概念を型として捉える形式的枠組の定義およびデータ型に基づく効率的変換手法、(ii) PSDのための計算機構を組み込んだ構造化文書の実現手法、(iii) 関数型言語におけるデータ型の理論の発展と準構造データに適した型理論の構築、の三つから構成される。(iii)は、さらに代数的プログラム変換(演算)の成果を構造化文書に適用し、自己参照による変換戦略を文書自体に付随させるという演算随伴機構に関する理論を含む。

以上の基盤技術に立脚したPSDを操作するためのPSD処理システムは、文書処理ソフトウェア等による構造化文書の編集から発信にいたる処理を可能とし、構造化文書を変換するためのプログラムの自動生成を実現するものである。このような優れた特徴を持つPSD処理システムは、ウェブを含むオフィス環境における基本ソフトウェアとしても位置付けることができる。

3. PSDのもたらす効果

本プロジェクトでは、構造化文書に対するPSD処理機構と変換規則を体系化し、ソフトウェアの信頼性確保の基礎となる言語仕様・設計仕様を設計し、実用的な構造化文書処理システムを開

発する。その成果として、一般の文書作成者が利用できる高信頼PSD汎用ソフトウェアパッケージ(統合的構造化文書処理システム)とPSD構造化文書を対象とした各種アプリケーション開発用のプログラミングシステム(構造化文書変換プログラム生成システム)を開発する。前者は構造化文書を効率よく作成し、それらを対象としたアプリケーションを半自動的に生成することを目的としている。また、後者は主としてアプリケーションソフトウェアの開発者向けのものである。

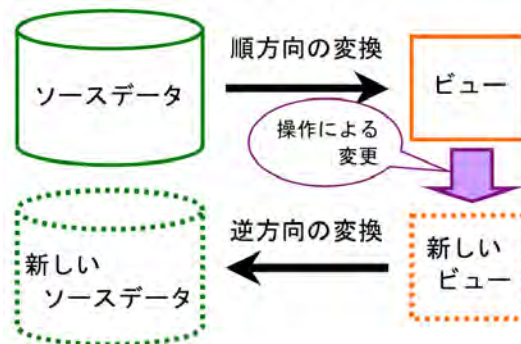
文書に価値を付与するためには作成時から文書を構造化しておく必要がある。統合的構造化文書処理システムは、それ単独または従来の文書処理ソフトウェア等と連携することで、付加価値の高い文書やコンテンツの作成を行うことが可能であり、情報発信のツールとして個人およびオフィスにおける利用が期待できる。特に、オフィス系ソフトウェアにおけるXML関連ソフトウェアの比率は大きくなってきており、構造化文書変換プログラム生成システムの利用は、このような文書の開発の効率を高め、生産性を向上させることになる。

4. 本年度の成果

PSD文書作成ツールの開発

2006年度には、これまでに開発したPSD処理システムを双方向変換機構により再構築し、XML上のアプリケーションに双方向変換手法を適用したシステムを開発した。双方向変換プログラムの記述にはBi-X言語を用い、2005年度に開発したJavaライブラリBi-XJをウェブサービスとして提供することで、双方向変換を自由に利用できる枠組みを実現した。

双方向変換とはソースデータとビューの間の両方向の変換であり、これによりソースの変更がビューに反映されるだけでなくビューの変更もソースに反映させることができる。



双方向変換言語Bi-Xでは、ソースからビューへの変換を与えることで逆方向の変換は自動的に導出され、2005年度にはその処理系をJavaライブラリとして提供した。本年度はこれをウェブサー



ビスとして提供するBi-Xサーバを開発することにより、任意の端末からのこの処理系の利用を可能にした。このBi-Xサーバを利用することで、これまで開発を行ってきた(株)ジャストシステム社のxfyシステムの上のアプリケーションで双方向変換の利用が容易になった。

本年度は、Bi-Xサーバを利用したアプリケーションとして、AmazonやCiteSeerのデータベースを閲覧・編集を行うアプリケーションを試作した。後述のVu-Xシステムも、Bi-Xサーバを利用したアプリケーションの一つである。

PSD 双方向変換生成ツールの開発

本年度においては、双方向変換に基づく実用的なアプリケーションの実現に向けて、ビュー上のインタラクティブな操作によって、元のソースデータの更新や、変換プログラムの生成を支援するシステムを設計した。また、これらの機能を利用したXMLベースのウェブサイト更新機構Vu-Xを開発した。

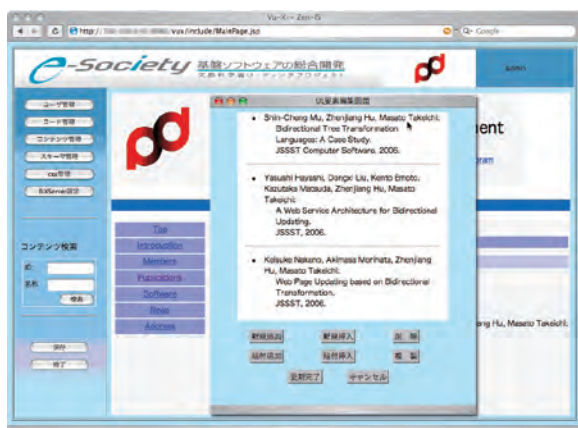
一般にウェブサイトは複数のウェブページから成り、それらが持つ情報の整合性を保持するにはXMLデータと変換の組合せによって各ページを自動生成する方法がよく知られている。しかしながら、各ウェブページの修正や更新に対応するXMLデータの編集箇所を見つけることが難しく、XMLデータへアクセスできる人が限定されてしまうという問題がある。Vu-Xシステムでは、双方向変換の枠組みを利用することでこの問題を解決した。また、Vu-Xは先述のBi-Xサーバを利用しており、特別なアプリケーションをインストールすることなく、標準のウェブブラウザ上でウェブページの編集ができ、その変更は自動的にXMLデータに反映されるため、必然的にサイト全体の情報の整合性を維持することが可能になる。

Bi-X言語によるプログラミングは、ポイントフリーと呼ばれる形式による記述を要求するため、一般のプログラマが変換を記述するには高度な技術や豊富な経験が必要となる。そこで、汎用的なXML処理言語XQueryで記述されたプログラムをBi-X言語による双方向変換プログラムへ翻訳する枠組みを提供することにより、双方向変換プログラムを容易に得られる環境を実現した。

また、Bi-X言語の処理系のない環境でも双方向変換の枠組みの導入を可能にするために、Bi-X言語で記述された双方向変換プログラムから順方向と逆方向の変換を表すXQueryプログラムに翻訳するツールを開発した。これにより、XQueryの汎用的な処理系を用いてBi-Xの特長である双方向変換機構を利用することが可能になった。

5. おわりに

2006年度は、双方向変換技術に基づくPSD文書処理システムの開発を行った。特に双方向変換をウェブサービスとして提供するBi-Xサーバを作成し、2007年度以降の実用的な高信頼構造化文書変換技術開発の基盤を構築した。



PSD 双方向翻訳ツールの開発

2006年度には、XML処理言語として広く使われている言語XQueryで記述されたプログラムをBi-X言語による双方向変換プログラムに翻訳するツール、およびBi-X言語による双方向変換プログラムをXQueryプログラムに変換するツールを開発した。

高信頼 WebWare 生成技術

名古屋大学 大学院情報科学研究科 阿草 清滋

研究分担者 和歌山大学 システム工学部 鯉坂 恒夫

研究分担者 愛知県立大学 情報科学部 山本 晋一郎

協力企業 富士通研究所 富士通ソフトウェアテクノロジーズ 野村総合研究所

<http://www.agusa.i.is.nagoya-u.ac.jp/research/webware/index.xhtml>

1. はじめに

本研究の目的は、WebWareの信頼性と安全性の向上のためにエンジニアが行う作業の支援と、デザイナーとエンジニアの協調作業の支援が可能な統合的WebWare開発環境を構築することである。

本年度は、WebWareのテスト支援、解析技術、作成支援システムにおいて以下の成果を得た。

2. WebWareのテスト支援

多くのWebWare開発では、分業による工数の削減を目的として、Strutsに代表される「MVCモデルに基づくWebアプリケーションフレームワーク（以降、単にWebフレームワークと呼ぶ）」が利用される。しかしながら、WebフレームワークにおけるViewであるJSPの単体テストは煩雑で工数のかかる作業であるため、JSPの単体テスト支援は高信頼なWebWare開発に重要である。昨年度までに、Webフレームワークをオブジェクトワークスに限定した上でJSP単体テスト支援ツールを開発し、実プロジェクトへの適用実験により我々のアプローチの有用性を確認した。本年度は昨年度までのアプローチをより一般化し、特定のWebフレームワークに依存する箇所をホットスポットとして差し替え可能なJSP単体テスト支援フレームワークを開発した。また、JSPのデバッグコスト削減を目的としてカスタマイズ可能なJSPコーディングチェッカを開発した。

JSP単体テスト支援では(1)テストケース生成、(2)テスト自動実行、(3)証跡保存を行う。Webフレームワークと独立したテストケース生成を可能にするため、WebWare設計モデルを策定しWebWare設計モデルからのテストケース生成機能を実現した。テスト自動実行ではテスト対象ページを表示するために、バックエンドデータのセットとWebフレームワークのController操作を行わねばならない。これらの実現にはWebフレームワークとの連携が必要不可欠なため、JSP単体テスト支援フレームワークにおいてホットスポットとして差し替えできるように基底クラスを作成した。また、テスト自動実行や証跡保存でのWebブラウザ操作に関しても、ブラウザ操作の基底クラスを作成した。このように、設計モデルの策定とホットスポット用基底クラスを作成することにより、JSP単体テスト支援に必要な機能とテスト支援プロセスを提供するJSP単体テスト支援フレームワークを開発した。本フレームワーク上で動作するStruts用のホットスポットを作成し、Strutsを利用したWebWareのJSP単体テストにおいてもJSP単体テスト支援ツールと同程度の支援が行えることを確認した。

JSPからWebフレームワークの機能を利用する際には、Webフ

レームワーク毎に特有の規則に従いタグライブラリを利用する必要がある。タグライブラリの埋め込まれたJSPは、サーブレット上でタグライブラリが展開され、JavaServletに変換され、コンパイルされ、動作する。このため、特にWebフレームワーク上での利用を前提にしたJSPに不具合が発生した場合、そのデバッグは困難である。コーディング規則はWebフレームワークによって異なるため、Webフレームワーク毎のコーディング規則の差異に柔軟に対応できるようなJSPコーディングチェッカが望まれている。本年度はコーディング規則が容易にカスタマイズ可能なJSPコーディングチェッカのプロトタイプとして、XPathを利用したコーディング規則記述に基づくJSPコーディングチェッカを作成した。プロトタイプは、JSPの細粒度解析結果であるXMLに対してXPathを適用し要素の取得を試み、その結果に応じてJSPがコーディング規則に準じているか否かを判定する。実際のプロジェクトで利用されているコーディング規則の3/4以上がプロトタイプで検査可能であることを確かめることにより、本手法におけるアプローチの妥当性を確認した。

3. WebWareの解析技術

WebWareのユーザインタフェースであるページは、JSPなどのページ生成プログラムの出力であり、原理的には無限通りのページが生成され得る。このためページ生成プログラムに対するテストが十分に実施されたかの判断が難しく、ページ生成プログラムのテスト達成度を測定するメトリクスが求められている。本年度は、WebWareを構成するページ生成プログラムのテスト達成度を測るメトリクスとして、生成されるページの構造に着目した基準に基づくカバレッジを提案し、提案するカバレッジの測定手法を開発した。昨年度までの成果であるWebWare解析ツールに対して、テストの実行結果から提案する基準に基づくカバレッジの測定機能を実現した。

ページにおける重要な要素は、表や箇条書きといった構造とその属性、表示される内容、異なるページへのリンクである。そこで、各ページからこれらの要素を特徴量として抽出し、特徴量が等しいページは同じページであるとみなすことで、ページ生成プログラムから生成され得るページを有限個に分類する。カバレッジは、テストによって生成されたページが、生成され得るページから得られる分類のうち、何種類を占めたかを求めることで得られる。また、ページの特徴量に基づいてページの等しさを判断するための基準として、5段階の基準を定めた。最も緩い基準ではページに含まれるリンクの遷移先のみを特徴量として利用し、最も厳しい基準では特定

のHTML要素と属性、リンクの遷移先、表示されるテキストを特徴量として利用する。

昨年度までの成果であるWebWare解析ツールにカバレッジ測定機能を実現した。カバレッジ測定の対象となるページ生成プログラムはJSPである。測定機能はWebアプリケーションのソースコードとテストの実行結果を入力とし、個々のJSPに対するテストのカバレッジを表示する。また、WebWare解析ツールのページランジションビューで表示されるJSPファイルを全網羅/部分網羅/テスト未実施に応じて色分けする(図1)。

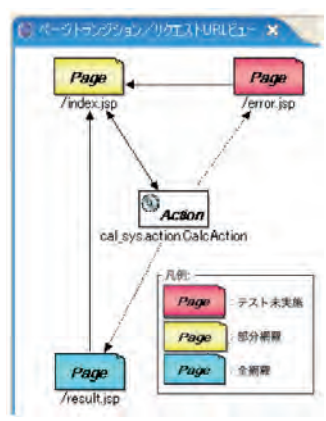


図1 ページランジションビュー上でのカバレッジの表示

4. WebWare の作成支援

WebWareにおいて、クライアント上でのユーザインタフェースを制御する代表的な技術として、JavaScriptとCSSがある。これらの技術は近年、非同期通信でのページ更新を実現するAjax技術として注目されている。本年度はWebWareのクライアントサイド技術の作成支援として、非同期通信でのページ更新を考慮したAjaxアプリケーションの解析手法を提案した。また、スタイルシート記述の保守支援としてCSSのリファクタリング技術を提案し、リファクタリングツールを開発した。

Ajaxアプリケーションの解析手法では、従来の非同期通信を含まないJavaScriptアプリケーションとの違いを明確にし、クライアントサイド内におけるページの状態遷移を捉える手法を提案した。さらに、サーバサイドでのページ生成だけでなく、クライアントサイドでのページ更新を考慮したデータフローモデルのための基盤技術を開発した。具体的には、非同期通信に用いられるXMLHttpRequestオブジェクトの利用パターンに着目し、イベントハンドラから、XMLHttpRequestオブジェクトのコールバック関数を経由して、要素に付けられたid属性などを利用して操作対象となるプロパティを特定する手法を開発した。

WebWareでは画面デザインの保守管理も重要である。デザイン(見た目)の制御はCSSで行われるが、開発の進行に伴い記述が複雑になり、保守性が低下する。このため、CSS記述の可読性向上を図るリファクタリングについて研究を進めた。リファクタリング支援機能として以下の機能を検討した。

- ・セレクトア間の依存関係を考慮した宣言の複製
 - 一つのCSSの各セレクトア間における依存関係を求め、プロパティごとに依存関係の親となるルールから子であるルールへ宣言を複製する。これにより暗黙に宣言されていたプロパティの明示化が可能となる。
- ・宣言の集合の等価変換
 - プロパティの指定には、paddingなどのように、上下左右の各プロパティを別々に宣言する方法(展開形)と一度に宣言する方法(短縮形)がある。展開形と短縮形の両方により指定可能なプロパティの宣言の集合を相互に変換する。
- ・同一セレクトアの合成
 - 同一のセレクトア名に対する指定が分割して記述されている場合に、セレクトアの詳細度に対応させる合成を行う。

これらの機能を組み合わせて、暗黙に宣言されているプロパティを明示し、プロパティの記述スタイルを統一することで高い保守性を獲得するCSSリファクタリングツールを作成した。

5. おわりに

本年度の成果として以下を得た。

- ・WebWareのテスト支援
 - JSP単体テスト支援フレームワーク
 - カスタマイズ可能なJSPコーディングチェッカ
- ・WebWareの解析技術
 - ページ生成プログラムのカバレッジ測定ツール
- ・WebWareの作成支援
 - 非同期通信でのページ更新を考慮したAjaxアプリケーションの解析手法
 - CSSのリファクタリング支援ツール

本研究で開発されたソフトウェアの一部は以下から入手可能である。

- ・ <http://www.agusa.i.is.nagoya-u.ac.jp/research/webware/index.xhtml>
 - ・ <http://www.sapid.org>
 - ・ <http://sdlab.sys.wakayama-u.ac.jp/jsanalyzer/>
- また、協力企業の製品にも組み込まれている。



インターネット上の知識集約を可能にするプラットフォーム構築

早稲田大学 理工学術院 村岡洋一

<http://www.muraoka.info.waseda.ac.jp/~katayamail/e-society/>

1. はじめに

インターネット上のWebサーバから発信される情報量は、本プロジェクトにより収集した144.5億のWebページをもとに推定すると、2007年3月時点で11,365万台のWebサーバから合計で500億ページに上ると推測される。

このような膨大なWeb上には、人間が一生かかっても学ぶことのできない情報、知識、ノウハウが凝縮されていると言っても過言ではない。これまで、我々は、このような膨大な情報を活用するために、GoogleやMSN、そしてYahoo!に代表される検索エンジンを利用してきた。各種情報の検索はもちろんのこと、辞書の代わりとしての利用、適切な翻訳語を探すための利用など、その用途は広範囲に渡る。

こうした検索エンジンに対して、最近注目されているのが、分析エンジン(Analysis Engine)である。分析エンジンとは、Webページを含む膨大な情報の中から有用な情報を見つけ出すことを目的としたエンジンである。このような分析エンジンとしては、1999年から研究プロジェクトとして実施されているIBMアルマデン研究所のWebFountainが有名であるが、対象とするWebデータの規模は20億ページに留まる。これに対し本プロジェクトでは、世界中のWebページから100億を越えるデータを対象とし、かつ、1ヶ月以内に更新される新鮮な情報を利用することにより、規模において世界一を達成すると共に、こうした膨大な情報からの知の創出を目指している。

2. 研究課題と進捗状況

本プロジェクトでは、商用・研究用を通じて世界最大となる120億ページを対象に、(1)平均して1ヶ月以内の新しいデータに更新することを可能とするWWWクローラを開発すると共に、(2)利用者の検索目的に応じて必要となる情報を抽出する知識フィルタリング技術の開発を目指している。

4年目である2006年度は、まず、2006年7月末まで網羅的な収集を継続し総収集ページ数を144.5億ページまで増加させた。さらに、解析対象として有用と判断できる日本語で記述されたWebページを抽出し、日本語ページに対して1ヶ月単位での更新を行った。知識フィルタリングの技術開発においては、具体的なアプリケーションとして「e企業調査プロトタイプシステム」を設定し実験を行うと共に、ブログや掲示板に対する解析を進めた。

以下に2006年度の成果概要を示す。

2.1. 世界最大の144.5億URLのWebデータ収集

表1に本プロジェクトで収集したWebページの内訳を示す。総収集ページ数は14,456,201,906ページに上る。世界中で発見したWebサーバ数は約13,468万台であり、内8,116万台の収集

を完了した。しかし8,116万台の内、実際に収集できたのは5,548万台であり、2,568万台(収集済サーバの約32%)は既にIPアドレスが存在しない等の理由からアクセスができなかった。また、256万台(収集済Webサーバの3%)については、/robots.txtによりクローラによるアクセスが禁止されていた。残り約5,000万台については未収集である。

表1に示すように、ドメインによってWebサーバあたりの平均Webページ数が大きく異なる。従来研究では、1台あたりの平均Webページ数は200ページ前後と考えられていたが、日本やドイツでは、これを大幅に上回るページ数が確認された。この主な原因は、バックエンドにDBを持ち各種情報を提供するサーバが両国で増加しているためだと考えられる。

表1 収集済Webページの内訳 (2006/7/31時点)

ドメイン	収集済Webページ数	割合	平均ページ数 / サーバ
com	7,140,485,020	49.4%	339
net	1,532,341,309	10.6%	169
jp (日本)	897,042,638	6.2%	609
de (ドイツ)	758,922,114	5.2%	459
その他	4,127,410,825	28.6%	185
合計	14,456,201,906	100%	261

次に、計画に従い解析に有用なWebページとして、日本語で記述されたWebページの抽出を収集済の144.5億のWebページから行った。そして、日本語ページを1ページでも保有するWebサーバに対して1ヶ月毎の更新収集実験を行った。その結果を表2に示す。

表2 日本語を含むWebサーバの更新収集

年月	対象Webサーバ数	前月からの増加サーバ数	前月からの更新割合
2006/09	867,672	-	-
2006/10	1,147,080	279,408	81%
2006/11	1,250,964	103,884	80%
2006/12	1,291,390	40,426	79%
2007/01	1,312,852	21,462	77%

実験結果から、世界中に存在する日本語ページを1ヶ月以内の新しいデータに更新するためには、29台の一般的なPC(2.4GHz, 1MBメモリ規模)があれば十分であることを確認した。また、1ヶ月毎の更新収集を通して、毎月約80%のページが更新(新規出現を含む)されていることが判明した。

なお、図1に示すように、富士通株式会社と共同で開発した分散収集型クローラに対し、現在どの地域のWebページを収集しているかをリアルタイムで表示する機能を追加した。

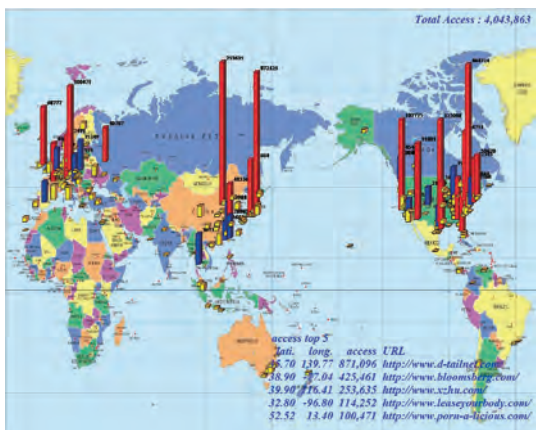


図1 収集先リアルタイム表示システム（富士通との連携）

2.2. 知識フィルタリング技術の開発

知識フィルタリング技術の開発では、具体的なアプリケーションとして、Web上のあらゆる情報から国内の特定企業に関する情報を集約して表示する「e企業調査プロトタイプシステム」を設定し実験を行うと共に、ブログや掲示板の解析を進めた。図2にe企業調査プロトタイプシステムの外観を示す。



図2 e企業調査（アクセラテクノロジーとの連携）

本システムは、Webページの情報をもとに、当該企業のホームページの特徴後や静的ページ数、動的ページ数、被リンク数、外部へのリンク数、サイト内リンク数、画像点数、さらには2005年度までに開発を進めた技術を用いてWeb上の評判情報、関連サイトなどを自動的に解析し表示する。

図3に解析例を示す通り、実際に企業を対象に解析を行ってみると電気系の同業種企業間においてもホームページの作りが大きく異なり、N社は静的なコンテンツ中心の従来型のWebページ、F社はユーザビリティを高めるためにサイト内リンクを多くしていることが理解できる。

このように、企業のホームページのリンク数や関連性の分析により、企業活動のある程度推測できることがわかった。特にインターネット上で盛んに活動している企業ほど、分析結果がその企業の特徴を示しやすいことがわかった。今後は企業活動の数値化精度の向上を目指す予定である。



図3 N社とF社のホームページ解析

3. 実施体制

本プロジェクトは、プロジェクトリーダー岡洋一（早大・理工）、サブリーダー山名早人（早大・理工）のもと、富士通株式会社、アクセラテクノロジー株式会社、国立情報学研究所等と共同で実施し、技術移転を進めている。

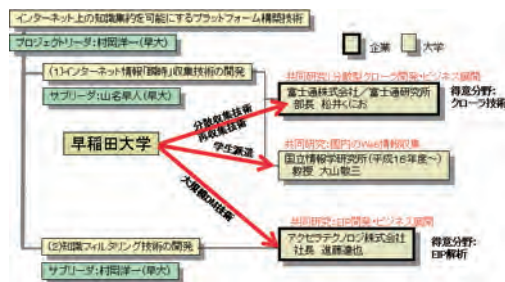


図4 実施体制

これまでの技術移転の成果として、図5に示す2製品への技術移転が進んでいる。



図5 技術移転

先進的なストレージ技術

東京大学 生産技術研究所 戦略情報融合国際研究センター 喜連川 優
協力企業 日立製作所, NTTプラットフォーム研究所
<http://www.tkl.iis.u-tokyo.ac.jp/project/e-society/>

1. はじめに

情報通信技術の革新が進む中で、ITシステムの一層の高信頼化、高性能化、管理容易化が求められている。とりわけ9.11テロ以降、災害時にもデータを失うことなく業務を継続させることができるディザスタリカバリ (DR) 技術が注目を集めている。また、爆発的に増加し続けるデータの高速な検索と低コストでの管理を実現するためには、ストレージの高性能化・管理容易化が必須である。本研究では、サーバ上のデータベースアプリケーションとの融合による次世代ストレージ技術 (Storage Fusion) の確立を目指す。戦略的競争力を実現すべく、上記の課題の解決に焦点を絞り研究を進めている。平成18年度は、安価なシステムコストで、広域災害時のデータ保護を実現する高度DR機構、各種知識を利用し大幅にI/O性能を向上させるストレージ超高速アクセス機構、ストレージとDBMSの性能ボトルネック検出を支援する管理コスト低減機構を開発した。

2. 高度ディザスタリカバリ機構

1) ログ転送によるリモートサーバレス方式のDBヘッド化技術の開発 地震など広域災害に対応したDR方式として、ログのみを同期リモートコピーで転送し、DBを副サイトでのログ適用により回復するログのみ同期転送方式を日立製作所より製品化した。さらに、平常時の機器コストを削減するために、副サイトでのログ適用をDBサーバではなく、プロセッサやメモリなどのリソースに制約のあるDBヘッド (DBアプライアンス) により行うこと目指し、ログ適用処理の高度化技術の開発、及び評価を行った。この結果、ログの並べ替えを行った後、順に並列適用プロセスに振り分ける方式により、最大で5.2倍速でのログ適用を達成した (図1)。本技術により、高速ログ適用を間歇的に実行し、その間に検索処理を行う副サイト検索サービスなど、副サイトの有効活用が可能である見通しが得られた。

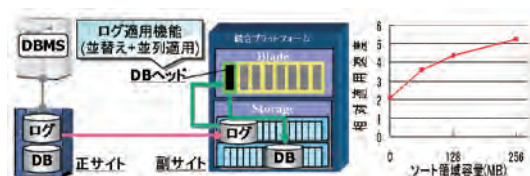


図1: システム構成 (左) とログ適用性能評価結果 (右)

3. ストレージ超高速アクセス機構

1) ストレージリソースとI/O性能の見積・予測技術の開発 システムの安定運用や複数処理間の性能調整を目的に、物理ドライブの稼働率が閾値以下となるように先読みI/Oの発行を制御する先読み絞り込み技術の開発、及び評価 (図2、図3) を行った。この結果、本技術によって目標の物理ドライブ稼働率で動作させるこ

とが可能となり、システムの安定運用や複数処理間の性能調整を実現する見通しが得られた。

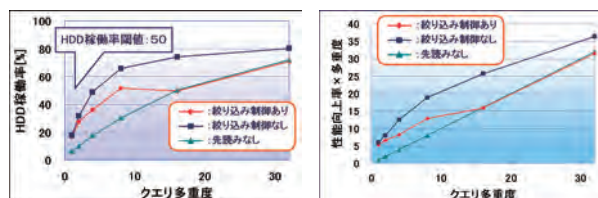


図2: HDD稼働率評価結果

図3: 処理性能評価結果

2) I/Oスケジューリング調整・複数DBMS調整技術の開発 I/Oスケジューリング調整・複数DBMS調整技術の実現方式を検討し、処理情報・モニタ情報から期待性能発揮に必要なリソース量 (稼働率含む)、I/O応答時間を予測し、制御する方式を考案した。

4. ストレージ管理コスト低減機構

1) ボトルネック検出技術の開発 DBシステムの性能障害対策において極めて困難とされる排他待ち原因の検出を支援する排他待ち原因検出支援技術の開発、及び効果の検証を実施した。この結果、本技術が排他待ちの大本の原因の特定に要する工数の削減に効果があることを確認した。また、性能障害の原因となるデータベース構造劣化に関して、これまで不可能であったリアルタイム解析を可能とする構造劣化可視化ツールを開発した。(図4)

2) DBMS構成最適化技術の開発 過去に発生した性能障害事例データベース、及び事例における診断手順を定型化したルールに基づく診断支援技術の基本設計を実施し本方式による性能障害対策支援方式を考案した。

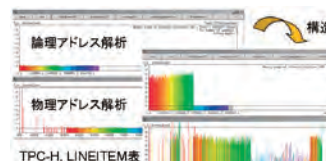


図4: 構造劣化分布の可視化

5. まとめ

ストレージとデータベースアプリケーションとの融合により、高信頼かつ高性能で管理が容易なデータ格納プラットフォームを実現する研究開発を実施した。高度DR機構においては、ログのみ同期転送方式を日立製作所より製品化したほか、機器コスト削減を目指したDBヘッド化を実現するために、限定されたリソースの下でも確実なログ適用を実現する高度化技術の実現方式検討と評価を行い、5.2倍の適用速度を達成した。ストレージ超高速アクセス機構においては、処理に対応するリソース量・I/O処理性能の詳細制御を実現する見通しを得た。ストレージ管理コスト低減機構においては、排他待ち原因検出支援技術の開発及び効果の検証を実施し、排他待ちの大本の原因の特定に要する工数の削減に効果があることを確認したほか、リアルタイムの構造劣化可視化ツールを開発した。また、DBMS構成最適化技術の実現方式検討を実施し、性能障害対策支援方式を考案した。

先進的な Web 解析技術

東京大学 生産技術研究所 戦略情報融合国際研究センター 喜連川 優
協力企業 三菱電機
<http://www.tkl.iis.u-tokyo.ac.jp/project/e-society/>

1. はじめに

Web上では企業や省庁、個人による情報発信が刻々と行われており、近年では実世界の様々な事象が網羅的かつ即時的にWebに反映されるようになってきている。実世界とWebの間には一種の転写構造が形成されつつあり、サイバー社会の構造を把握し、その変化を追跡することは、実社会に起こる事象の背景や予兆を探る上で極めて有効と考えられる。

本プロジェクトでは、Web上の社会知の高度利用を可能にする新たなWeb解析システムの創出を目的とし、日本国内Web情報の過去から現在に至る履歴を蓄積したWebアーカイブを構築し、リンク解析技術およびテキスト解析（自然言語処理）技術を核にWebの空間構造および時系列変化の分析を可能にするSocio-Senseなるシステムの開発を進めている。本年度は特に、企業における消費者動向のタイムリな把握など、実応用に対する有効性の実証に向け、要素技術の統合を進めた。

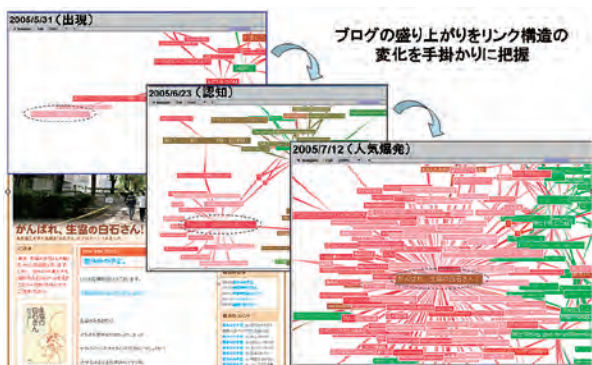
2. Webアーカイブ基盤

Web解析の基盤として日本語Webページを中心とするアーカイブの構築を進め、約8年分の情報を蓄積するに至った。Webアーカイブはスナップショット単位での蓄積からWebページ毎に最適な時間分解能で蓄積する方式へと移行しており、Webの変化を最短1日単位で捉えることが可能になっている。また、追加された情報を直ちに利用できるよう、リンク解析およびテキスト解析との連携も進めた。

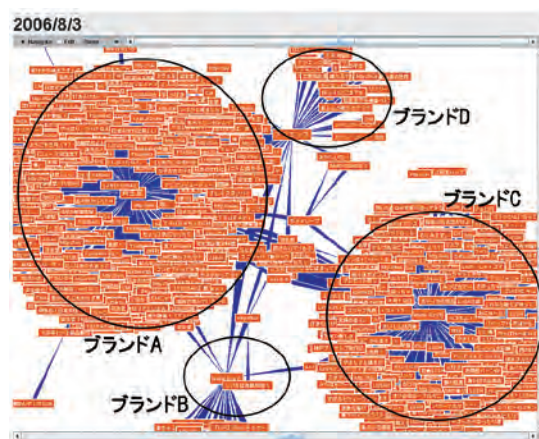
3. Web時空間分析

3.1. 消費者動向分析への応用

ブログの流行により、個人からの情報発信が一段と容易になると共に、企業や従来メディア発の情報にない迅速性や真実味を持った情報源として個人の意思決定に大きな影響を及ぼすようになった。企業にとっても消費者の動向を把握し、より親密な情報発信を行うツールとして、ブログの重要性が増している。



左下の図はある人気ブログの周辺リンク構造を、その出現から人気爆発までWebアーカイブを用いて時系列的に再現したものであり、サイバー空間上での情報の伝播・普及過程を読み取ることができる。また、下図はある商品カテゴリの競合ブランドについて書き込みを行っているブログ間のリンク構造を分析した結果である。



これらの手法のマーケティングにおける有効性を実証するため、当該分野の専門家（専修大学新井教授、株式会社電通）と共同で分析を進めており、例えば広告費とブログ書き込み数の相関などに関して、いくつかの興味深い知見が得られている。

3.2. Webテキスト解析の高度化

Web解析を企業等が利用する際には、当該企業やその製品・サービスに関する客観的な事実よりも、主観的な意見や評価などの記述が重要となる。このような評価情報テキストをWeb上のテキスト情報全体から自動的に抽出するには、まず、評価を記述するときに使用される言語表現（「素晴らしい」「ひどい」など）を登録した辞書が必要となる。精度の高い辞書の作成には人手がかかるが、従来、Web上の多様な言語表現に対応しきれないという課題があった。そこで本年度は、少数の規則に基づいてWeb文書集合全体から評価表現辞書を自動構築する手法に取り組み、大規模な辞書を構築することに成功した。今後は、この辞書を用いて抽出した評価情報テキストの解析に取り組んで行く。

4. まとめ

これまで開発してきたWebアーカイブ基盤、リンク解析技術およびテキスト解析技術を統合したWebマイニング実証システムの構築を進め、企業の消費者動向分析等の応用に対する有効性実証に着手した。今後は、実証実験をさらに進めて行くと共に、そのフィードバックに基づいて各要素ならびにシステム全体の高度化を図る予定である。

ユーザ負担のない話者・環境適応性を実現する自然な音声対話処理技術

奈良先端科学技術大学院大学 情報科学研究科 鹿野清宏

研究分担者 京都大学 河原達也, 名古屋大学 武田一哉, 和歌山大学 河原英紀, 奈良先端大 猿渡洋, 名古屋工業大学 徳田恵一,
立命館大学 西浦敬信

協力企業 松下電器, 旭化成, 日立製作所, 松下電工, ASTEM, オムロン

<http://cif.iis.u-tokyo.ac.jp/e-society/database/index.html>

1. プロジェクトの目標

携帯電話, 携帯端末, PCの入出力, カーナビ, 家電制御, 秘書ロボットなどを, 誰でも容易に利用できることが望まれる. これらを実現するには, 人と機械との自然な対話として, 音声認識・合成技術が有望である. 本格的に利用されるためには, 頑健かつ高精度の音声認識基盤ソフトウェアを開発して, かつ廉価に利用できるようにすることが重要である. この音声認識・合成技術の基盤ソフトウェアの普及により, 誰でもが気軽に, 情報機器の利用ができるようになり, デジタルデバイドの社会問題の軽減, IT市場の活性化につながる.

プロジェクトでは, 大語彙連続音声認識プログラム, 話者環境適応プログラム, ハンズフリー音声認識プログラム, ハンズフリー音声収録DSP, 大語彙連続音声認識プログラムのマイコンへの実装を行う. 音声合成では, 多様な音声合成プログラムを作成する. これらのプログラムは, 単に開発するだけでなく, 実環境での応用システムでの実証試験による評価も行って改善をはかる. さらに, 音声認識技術の利用法のノウハウも蓄える.

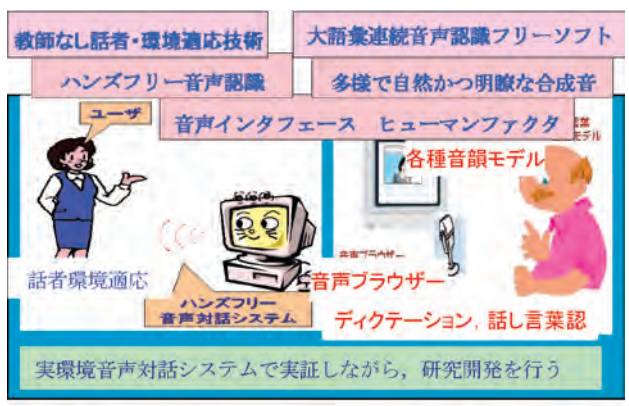
当初の計画に付け加えて, 静かな音声メディアとして発見した「非可聴つぶやき声 (NAM:Non-Audible Murmur) の音声認識・合成の研究開発も進める.

2. プロジェクトの概要

音声認識を本格的な商用化につなげるには, 頑健な音声認識システムとして, 以下の技術を研究開発することが必要である. (i) **利用環境およびユーザに対する負担をかけない適応技術**, (ii) **高精度連続音声認識プログラム**および音声認識システム研究開発ワークベンチ, (iii) **マイクを意識しない自然なハンズフリー音声認識技術**が必要となる. これらのソフトウェアを研究開発し, 廉価に誰もが利用できるプログラム, あるいはDSP/マイコンとして提供する. さらに, 開発したソフトウェアを用いて, (iv) **人と機械の音声対話の実証実験**を行い, ソフトウェアの評価を行うとともに, 利用に関するノウハウを蓄積する. 音声合成では, (v) **多様な声質の実現**が重要であり, 高精度音声分析合成系 STRAIGHTを用いた声質変換プログラムを開発する.

上記に加えて, 非可聴つぶやき声の音声認識 (無音声認識) を (i) で, 変換合成 (無音声電話) を (v) で研究開発を行う. さらに, (ii) では, 新しく開発された音源分離 (BSS:Blind Source Separation) アルゴリズムによるハンズフリー音声認識も行う. (iii) では, マイコンで動作する大語彙連続音声認識プログラムの実行速度を実時間処理にまだ高める.

本プロジェクトの研究開発の概要を図に示す. さらに, 研究開発項目ごとの成果の状況および予定を表にまとめておく.



人にやさしい自然な音声インタフェース

研究開発の成果の状況および予定

研究開発項目	成果の状況および予定
ユーザ負担のない話者・環境適応	教師なしオンライン話者適応 (H18.3) 非可聴つぶやき声 (NAM) 認識 (H19.3)
大語彙連続音声認識ソフトウェア	話し言葉認識モデルとプログラム (H18.3) 大語彙連続音声認識プログラム Julius (H18.3) マイコン SH-4A Julius (実時間動作) (H19.3)
ハンズフリー音声認識	ハンズフリー音声収録 DSP (H19.3) BSS 音源分離オンラインプログラム (H18.3)
実環境音声対話システムの構築	たけまるくん音声情報案構築キット (H16.7) 自動車内音声認識プログラムキット (H19.3)
多様な声質の音声合成ソフトウェア	オンライン音声変換プログラム (H18.3) 高精度音声変換プログラム (H19.3) HMM ベース音声規則合成システム (H15.10) 無音声電話プログラム (H19.3)

(青字で書かれた成果は, 当初計画になく, 追加した成果と予定)

3. H18年度の進捗

ハンズフリー音声認識では, 電子情報通信学会の論文賞, IEEE/IROS Best Application Paper Awards などを受賞した. 音声対話システムでは, たけまるくんシステムの開発と運用が評価され, 情報処理学会山下記念研究賞を受賞した. 話し言葉の認識では, 日本音響学会粟屋潔学術奨励賞を, NAMによる個人認証で暗号とセキュリティシンポジウム SCIS2006論文賞を受賞し

た。鹿野は、これまでの研究が評価され、IEEE Fellowに推挙された。これまでの研究が大いに評価され、多くの受賞につながった。音声合成でも、音声分析合成システムSTRAIGHTのデファクトスタンダードとしての地位がさらに強固なものとなるとともに、音声テキスト合成システムHTSの汎用性と優秀さが国際コンペティションなどで実証された。

研究項目ごとにH18年度の進捗をまとめておく。

- (i) 教師なし話者適応プログラム： オンライン教師なし話者適応プログラムを実現し、良好に動作することを確認した。騒音の大きい駅にエージェントタイプとロボットタイプの音声情報案内システム「キタちゃん」と「キタロボ」を設置して、1年間運用し、良好な結果が得られた。非可聴つぶやき (NAM) の秘話性を利用した話者照合で高い照合率を確認した。
- (ii) 大語彙連続音声認識ソフトウェア： Webから話し言葉調の文を自動選択することにより、話し言葉言語モデルの構築ツールを作成した。日立と共同で、SH-4Aマイコンに、大語彙連続音声認識プログラムJuliusの実装を進め、2万語の連続音声認識の字湯時間動作を実現した。その他、Juliusの機能強化、性能強化を進めた。
- (iii) ハンズフリー音声認識： 実時間ブライント音源分離 (BSS) 処理のDSP実装、高精度化を行った。ハンズフリー音声対話システムを、旭化成と共同で開発し、DSPモジュールに空間スペクトルサブトラクションアレー (SSA) を実装して実現した。BSS処理を雑音推定に用いることにより、SSAの改善としてBSSAを考案した。
- (iv) 音声認識のフィールドテスト： 多くの対話システム「たけまるくん」「キタちゃん」「キタロボ」、京大博物館案内システム、京都市バス案内音声対話システムなどを設置して、音声データの収集に基づくノウハウの蓄積を行ってきた。また、ユーザのPCの音声対話システムを利用してインターネットから楽曲をダウンロードするシステムを利用して、車内を含む多様な環境下で性能を評価した。自動車内環境で集録した大量の音声データを用いて雑音処理の評価を行い、ハンズフリー自動車内音声認識システムを作成した。
- (v) 多様な音声合成プログラム： 高品質なオフライン版音声変換プログラムをC言語で実装した。HMMを用いた韻律制御モデルを改善した新バージョンのHTSを公開するとともに、昨年度から引き続き国際コンペティションを開催した。統計的手法による音声モーフィングアルゴリズムを改良して、非可聴つぶやき (NAM) から通常音声やささやき声への変換に適用して、有効性を確認した。

教師なし話者・環境適応技術

奈良先端科学技術大学院大学 情報科学研究科 鹿野清宏

<http://spalab.naist.jp/>

1. ねらい

人と機械との自然な対話を実現するためには、高精度な音声認識技術が必須である。とくに、利用環境とユーザへの適応技術が必要となり、かつ、ユーザに負担をかけない適応技術が望まれる。このような環境と話者への適応を実現のために、教師なし話者・環境適応技術に関して研究開発を行う。

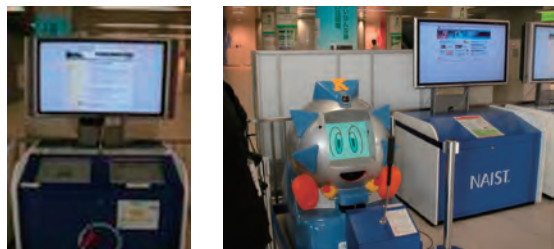
2. 研究の概要

雑音環境下で、任意の1文発声に基づくユーザに負担をかけない教師なし話者適応アルゴリズムの研究開発を進める。具体的には、雑音に頑健な音声認識アルゴリズムとHMM (隠れマルコフモデル) 十分統計量に基づく教師なし話者適応アルゴリズムの研究開発を行う。最終的に、オンラインで動作する話者・環境適応プログラムを完成させる。さらに、新しい静かな音声メディアとして、つぶやき声 (NAM) の認識 (無音声認識)、声を出さない電話 (無音声電話) の研究を行う。

3. H18年度の進捗

非可聴つぶやき声による個人認証の研究で、「暗号と情報セキュリティシンポジウム」の論文賞を受賞した。たけまるくん音声情報案内システムで、情報処理学会の山下記念研究賞を受賞した。H18年度における進捗を以下に示す。

- (i) オンライン教師なし話者適応プログラムを、不特定話者モデルのデコーダーと話者適応モデルのデコーダーを並列に実装することによって、オンライン動作を実現した。
- (ii) 騒音の大きい駅にエージェントタイプの音声対話システム「キタちゃん」と、ロボットタイプの「キタロボ」を設置して、1年間運用して、良好な動作を確認した。
- (iii) 実環境音声データを選択するアルゴリズムを提案して、不特定音韻モデルの学習での有効性を実証した。
- (iv) 外部に聞こえないつぶやき声 (NAM) の認識および無音声電話の研究を継続した。NAMマイクと音声モーフィングを利用した発話障害者音声の研究を行った。その他、NAMが他人に聞こえないとの特質を利用したキーワード発声を利用したNAM個人認証の研究を行った。



駅に設置した音声情報案内システム
「キタちゃん」(左)とキタロボ(右)

大語彙連続音声認識プログラム

京都大学 学術情報メディアセンター 河原達也

<http://www.ar.media.kyoto-u.ac.jp/>

1. ねらい

ユーザに負担のない自然な音声対話を実現するには、音声認識システムが、できるだけ広範な話者層や言い回し、特に話し言葉に対応できる必要がある。このような大語彙連続音声認識を行うオープンソースのプログラムを開発する。また、組み込み機器にも利用できるように、マイコンへの実装も行う。

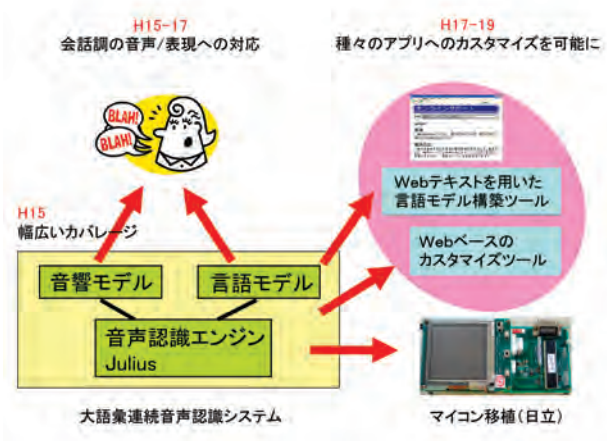
2. 研究の概要

これまでのディクテーションシステムは主に成人の読上げ音声を対象としていたが、音声対話を指向して発展させる。音韻モデルについては、多様な話者および対話調の音声に対応できるようにする。言語・発音モデルについても、話し言葉に対応できるようにする。さらに、音声認識エンジン Julius についても音声対話システム向けの機能強化を行うとともに、マイコンに移植できるような効率化を図る。

3. H18年度の進捗

平成18年度における研究進捗を以下にまとめる。

- (i) 話者適応に適した不特定話者の音韻モデル (SAT 及び VTLN) を構築し、大語彙連続音声認識において評価を行った。
- (ii) Webテキストから話し言葉調の文の自動選択により、音声対話システム向けの言語モデルを構築するツールを作成した。
- (iii) 音声認識エンジン Julius の機能・性能強化 (音声分析の充実など) を行った。
- (iv) SH-4マイコンにおける Julius の実装を改善し、2万語の連続音声認識の実時間動作を実現した。



で3ヶ月にわたり運用を行った。

上記に加えて、ソフトウェア普及のための講習会を行った。

ハンズフリー音声認識

奈良先端科学技術大学院大学 情報科学研究科 猿渡洋

<http://spalab.naist.jp/>

1. ねらい

ユーザに負担をかけない自然な音声入力系として、ハンズフリー音声認識システムを構築する。特に、マイクロフォンアレー (図2参照) による音声収録技術に着目し、音声認識性能の向上を目指す。また、コンパクトかつ廉価なマイクロフォンアレーアルゴリズムの開発を行う。

2. 研究の概要

ユーザからの距離1m以下で高性能に動作するハンズフリー音声認識システムを、8チャンネル以下のマイクロフォンアレーを用いて開発する。認識性能は、1m離れた音声入力で、従来の接話マイクとほぼ同等の認識性能を目指す。さらに、マイクロフォンアレーのコストを20分の1以下にするため、ハンズフリー音声収録用 DSP を開発する。

3. H18年度の進捗

平成18年度における研究進捗を以下に示す。

- (i) 空間スペクトル演算アレー SSA (Spatial Subtraction Array) に関して、非線形処理歪に対応したマッチド音響モデルの構築を行った。また重み付き CSP 法に基づく実時間方位推定・発話検出法を組み込んだPC実装を行った。
- (ii) DSPモジュール上に実時間 SSA 処理系を実装し、ハンズフリー音声対話デモシステムを構築した (図参照)。
- (iii) 音源間の独立性のみに基づいて分離を行うブラインド音源分離 (BSS) 処理の DSP 実装・高精度化を行った。
- (iv) BSS 処理を雑音推定に用いることにより、素子誤差等に頑健な SSA アルゴリズム改良を行った。
- (v) 駅にマイクロフォンアレーおよびロボット音声対話システムを設置し、実環境における雑音データを収録した。



図 マイクロフォンアレー及び SSA 用実時間 DSP モジュール

音声認識システムのフィールドテスト

名古屋大学 情報科学研究科 武田一哉

<http://www.sp.m.is.nagoya-u.ac.jp/>

1. ねらい

ユーザにとっての音声対話システムの性能は、認識性能だけでなく、様々なヒューマンファクタに支配されている。開発したプログラムやモデルを実環境下で運用することで、開発成果の検証を行うとともに、運用結果からユーザ負担の少ない音声対話システムの設計指針を得る。

2. 研究の概要

開発した音声認識プログラム、各種モデルを用いて、実環境下での音声対話システムのフィールドテストを行い、開発成果の検証を行うとともに、多様なアプリケーションを効率的に作成する方法や、音声認識利用のヒューマンファクタに関するノウハウを蓄積する。

3. H18年度の進捗

平成18年度における研究進捗を以下に示す。

(i) 楽曲ダウンロードサービスの運用と評価・改良

音声対話を用いてインターネットから楽曲をダウンロードするシステムを利用して、室内だけでなく自動車内を含む多様な利用環境下で音声対話サービスを運用し、実環境下で音声対話システムの性能を評価した。特に、音響信号処理・パターン認識・言語モデルといった基礎技術以外の、未だ十分定式化されていないヒューマンファクタの把握につとめ、システムの総合的な性能を改善した。

さらに、公共施設による情報案内システム及び駅構内での案内システムの運用を継続して行いデータ収集を続けた（奈良先端大）他、質問生成機能を持つ観光案内システムを博覧会会場で運用し（京都大学）、その有効性を検証した。

(ii) フィールド収集データを用いたモデル学習

システムの運用を通じて多様な環境下での実音声を集め、収集した発話を用いて音響モデル、言語モデルの性能を改善した。

(iii) 自動車内音声認識プログラムの作成

フィールドシステム運用により収集されたデータを利用するとともに、雑音処理手法の評価を行い、自動車内で運転中にハンズフリー、アイズフリーで利用可能な、自動車内音声認識システムを作成した。

多様な音声合成プログラム

和歌山大学システム工学部 河原英紀

<http://www.wakayama-u.ac.jp/~kawahara/>

1. ねらい

本プロジェクトでは、機械と人間との対話を自然なものとするために、人間のように多様な声質を有する音声を合成することの出来るプログラムを開発する。

2. 研究の概要

本プロジェクトでは、我々の開発した高精度音声分析合成系 STRAIGHTを用いた多様な声質の合成音声を作成する。話し手による声質の違いや話し手の感情や話し方による声質の違いを取り出すためのデータベースを整備し、解析結果を利用してプログラムを開発する。開発するプログラムは、高品質で精密な声質の付与が可能ではあるが非リアルタイムのものと、品質には制限があるがリアルタイムで動作するものの二系統とする。

3. H18年度の進捗

本年度も研究計画に従い下記の項目について研究を推進した。

- (i) オフライン版の変換プログラムをまとめた。
- (ii) 音声変換プログラム応用を促進するための支援ツールおよび、設計のためのデータを整備した。
- (iii) 変換用プリセットデータを作成した。
- (iv) オンライン変換プログラムを評価した。
- (v) 隠れマルコフモデル (HMM) を用いた韻律制御モデル用の音声特徴パラメータの検討を行った。
- (vi) HMMを用いた韻律制御モデルを中心とした音声合成システム (HTS) の新バージョン (図) を公開するとともに、国際コンペティションを開催した。
- (vii) 統計的手法による音声モーフィングアルゴリズムを改良し評価した。

昨年度に引き続き、プロジェクト成果の普及に努め、デファクトスタンダードとしての地位を強固なものとした。HTSの汎用性と優秀性が証明されたことを特筆することができる。

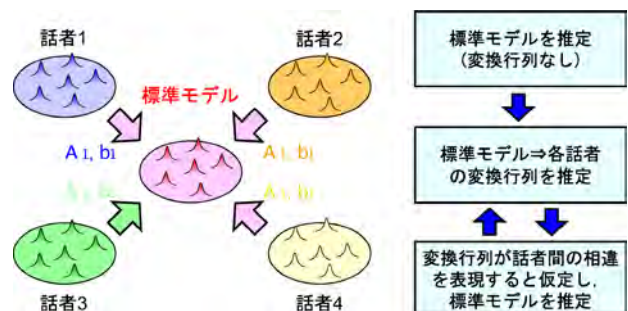


図 HTS 新バージョンにおける話者適応法