

研究種目：特定領域研究

研究期間：2006～2010

課題番号：18049073

研究課題名（和文）情報爆発に対応する新 IT 基盤研究支援プラットフォームの構築

研究課題名（英文）Design and Development of Advanced IT Research Platform for Information Explosion Era

研究代表者

安達 淳 (ADACHI JUN)

国立情報学研究所・コンテンツ科学研究系・教授

研究者番号：80143551

研究分野：総合領域

科研費の分科・細目：情報学・メディア情報学・データベース、知能情報学

キーワード：スケーラブル計算基盤、ソフトウェアライブラリ、大規模コーパス、実世界インタラクション、センサールーム、自然言語処理、生態計測、センサーネットワーク

1. 研究計画の概要

特定領域研究「情報爆発 IT 基盤」で実施されるコンテンツ、ソフトウェアシステム、情報メディアの各研究項目内の研究を支援するとともに、研究項目および計画研究・公募研究に渡るコラボレーションを効果的に進めるため、共有のプラットフォームとして下記の研究環境を構築する。

(1) 大規模コンテンツ共通評価基盤の実現

国立情報学研究所を中心にして進めてきた評価用テストコレクション構築と連携し、本領域での研究展開にあわせ新しく1億件を越える Web 文書のコーパスを構築、様々な言語処理を施し索引付けを行う。この共有基盤の上で、新しい情報検索や情報サービスの研究タスクを設定して研究グループ共有の研究資源として活用する。

(2) スケーラブル計算基盤テストベッド構築

広域分散したクラスタマシンのテストベッドを整備し、仮想マシンによるエミュレーションによる大規模な基盤ソフトウェア研究環境を提供する。分散システムやソフトウェア分野に限らずコンテンツ処理等の研究を効果的に促進する基盤とする。具体的には、全国に分散する拠点に計算機クラスタを配置し、連携させて様々な計算処理を並行処理する環境を提供する。またソフトウェア環境構築の研究用テストベッドとしても使用する。

(3) 実世界インタラクション計測分析環境の構築

マルチメディア情報提示やロボットとの

インタラクションにおける人間の発話や表情、行動をマルチモーダル・センサおよび生体・脳計測装置によってリアルタイムで高精度計測するセンサールーム環境を実現し、人間のインタラクションに関する様々な計測データを採取し共有するとともに、コーパスを作成しそれに基づく研究促進を図る。

(4) 社会情報基盤としての実証

分散システム、コンテンツ共有、センサー計測など多岐に渡る研究成果の統合事例として、生活習慣病に関する医療情報センサーデータの収集解析実験環境を整備し、実証的な評価を実施できるようにする。

2. 研究の進捗状況

支援班活動は、他にはない大規模な共有資源を活用することにより、研究を効果的かつ効率的に支援し、計画研究・公募研究に渡る連携を促している。

(1) 大規模コンテンツ関係 日本語WEB文書1億ページを、開放型検索エンジンTSUBAKIにて公開し領域内外で活用している。ここでは、自然文クエリ内の重要語、冗長語を自動認識する手法を適用するなど、文書内に研究成果の言語処理結果を取り込んで、他の研究グループでの活用を容易にしている。また、100以上のノードを持つセンサーネットワーク実験用テストベッドをX-sensor1.0として公開し、通信やデータ処理プロトコルを実験できる環境を実現した。また、センサーデータ共有機構の提供を開始した。

(2) スケーラブル計算基盤テストベッド関係
広域分散テストベッドInTriggerを整備し、全国
の拠点に計15クラスタ、260ホストを擁して
いる。また、ネットワークの10Gbps化を進め
ているところである。アプリケーション研究
グループ（Webページ発見、天文学の画像処理
等）との連携を深め、研究している連携ソフ
トウェアの実装に反映させるとともに、テキ
ストマイニングなど、他の研究項目の計算処
理に活用している。

(3) 実世界インタラクション計測分析環境関
係 京都大学に、センサールーム環境IMADE
の構築を行い、様々な計測機器や計算機など
を導入し、人の視線の追跡や発話の「間」の
計測など、インタラクション要素の抽出とイ
ベント発見などが可能な環境整備を実施した
。また、それらに必要なソフトウェアライブ
ラリを整備してきた。

(4) 社会情報基盤関係

生活習慣病予防医療実験として、九州地区
においてセンサーを活用した情報収集と解析
実験を実施し計測データの評価ができるよう
な環境整備を行っている。

3. 現在までの達成度

①当初の計画以上に進展している

計画研究 13 件、公募研究 74 件からなる本領
域で、支援班を設けたことは、平成 20 年度
の領域の中間評価でも研究成果創出を促す
効果が大きいとして高い評価を受けた。支援
班により、一つの研究班では購入できないよ
うな機器環境を共有できるように整備し、ま
た、共通の研究基盤となるデータベースやソ
フトウェアを共同開発して共有することが
当初の目的であった。これにより、研究全体
の投資効果が極めて効率的なものとなると
ともに、研究班の間および研究項目を越えた
連携を自然に促進することができ、絶大な効
果を上げることができたと考えている。さら
に予想を上回ったこととして、若手研究者や
大学院生などが研究班や大学を越えて共有
プラットフォームの構築に参画し、研究者交
流の環境にもなるなど、人材育成によい効果
をもたらしている。また共通のデータベース
を使用して実験することにより、研究間での
比較評価も客観的に行えるようになってい
る。これらの効果は、評価助言委員会でも認
められていることであり、研究の実証や成果
の「見える」化でも支援班の構築した研究環
境が貢献している。

4. 今後の研究の推進方策

後半 2 年においては、前半で構築してきた
共有プラットフォームを活用し研究成果創
出を促進することに注力する。集積したコン
テンツに対して様々なマイニングを施したり
、ソフトウェアの整備を行う。また、
InTrigger についてはコア数を引き続き増
大し他にはない程度の規模の実験をできる
ようにしたい。また、インタラクションの計
測では、より豊富な実験データコーパスを収
集し研究への利用が円滑に進むようにする。

まとめとして、地域医療を想定した社会情
報基盤の実験を実用レベルに近い形で行い、
総合的に評価できるような共有環境とする
ことを目指す。

5. 代表的な研究成果

〔雑誌論文〕（計 8 件）

(1) 相澤彰子、高須淳宏、深川大路、高久雅
生、安達淳 “アカデミックリンケージ：膨
大な学術情報へのアクセスを支援するリン
ケージ基盤” 情報処理 Vol.49 No.8 2008.
pp.935-938

(2) 黒橋禎夫、新里圭司 “TSUBAKI：深い言
語処理を特長とするオープンサーチエンジ
ン基盤” 情報処理 Vol.49 No.8 2008.
pp.931-934

(3) 大島裕明、中村聡史、赤星祐平、中田秀
基、喜連川優、田中克己 “SlothLib/EaRDB：
マイサーチエンジン開発環境” 情報処理
Vol.49 No.8 2008. pp.927-930

(4) 喜連川優、安達淳、他 “情報爆発時代に
おける情報管理・融合・活用基盤” 人工知能
学会誌Vol.22 No.2 2007. pp.215-221

〔学会発表〕（計 2 件）

(1) Masaru Kitsuregawa. “Challenge for I
nfo-plosion.” The 18th International Co
nference on Algorithmic Learning Theory
(ACT2007) 2007.10 Tohoku University Se
ndai Japan

〔図書〕（計 0 件）

〔産業財産権〕

○出願状況（計 0 件）

○取得状況（計 0 件）

〔図書〕（計 0 件）

〔その他〕

ホームページ

[http://www.infoplosion.nii.ac.jp/info-p
losion/html/S/](http://www.infoplosion.nii.ac.jp/info-plosion/html/S/)