

機関番号：62615

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

研究成果の概要（和文）：本特定領域に参加する計画・公募研究班で共用するための研究基盤を構築し、研究活動の支援を行った。これにより、限られた経費の中で研究資源の共用を図り研究連携を深める効果を発揮した。具体的には開放型検索エンジン TSUBAKI による大規模コーパスの提供、広域分散コンピューティングテストベッド InTrigger、実世界インタラクション計測分析環境 IMADE、そしてセンサーネットワーク予防医療の実験環境を構築した。

研究成果の概要（英文）：This project implemented a common research infrastructure for all the research groups participating in this priority-area research initiative, accordingly supported all research activities in this initiative. Providing this infrastructure, we succeeded in accelerating shared utilization of research facilities and resources within the limitation of research funding and strengthening the collaboration among research groups. These shared facilities include (a)TSUBAKI: a open search engine for large-scale corpus, (b)InTrigger: Widely-distributed computing test-bed, (c)IMADE: an environment for real-world interaction measurement and analysis, and (d) prototyping for sensor-network based preventive medicine.

交付決定額

(金額単位：円)

	直接経費	間接経費	合計
2006 年度	145,500,000	0	145,500,000
2007 年度	135,300,000	0	135,300,000
2008 年度	134,650,000	0	134,650,000
2009 年度	133,200,000	0	120,200,000
2010 年度	112,100,000	0	112,100,000
総計	660,750,000	0	660,750,000

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

キーワード：スケーラブル計算基盤、情報リンケージ、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、センサー、クラウド、

1. 研究開始当初の背景

本領域では四つの研究項目（柱）および多数の公募研究を並行して進める。各研究グループでは、大型研究設備、超高速ネットワーク、大規模共通情報基盤などを必要とするが、限られた研究経費を有効に活用し成果を最大限にするために、研究の共通基盤を設けるという方針を採ることにした。領域内の研究

グループ間でこれらの資源を共用し、計画研究・公募研究に渡る連携を効果的に確立するために支援班を設け、ここに経費を重点的に配分するという基本方針を立てた。支援班による共通基盤整備は、国立情報学研究所の提供するネットワーク等の活用も勘案し、また大学等に分散する研究拠点にて効果的に共用するための企画調整は総括班および柱長

会議で行い、それに従って支援班で実施するような方針を立てた。

2. 研究の目的

支援班では、領域内の柱に対応して、以下の基盤整備を推進する。(A01) 大規模コンテンツ共通評価基盤の実現、(A02) スケーラブル計算基盤テストベッド構築、(A03) 実世界インタラクション計測分析環境の構築、(B01 および領域全体)以上を総合して社会情報基盤における実証的実験、の四項目である。我々はこれらの研究基盤を総称して「共創プラットフォーム」と呼んでいる。

支援班で供用する設備やコンテンツは、領域内すべての柱の計画研究および公募研究の間で有機的な共用を図る。研究経費の1/4弱を支援班に重点的に割り当て、領域全体で共有する共創プラットフォームを全員の協力により構築・共有することにより、個々の研究者では困難な斬新かつ骨太な実験を可能とすることを目的とする。これは超分散システム上で情報共有をすることを目指すもので、研究項目を越えた連携とイノベーション創出を加速するとともに、若手研究者の積極的かつ献身的な研究活動を強力に支援し、従来にない大規模な実験を効率良く実現することを目的としており、短期間でスケールの大きな先進的な研究成果を創出することを試みた。

3. 研究の方法

A01 柱では、大規模コンテンツ共通評価基盤を実現すべく、開放型検索エンジン TSUBAKI として実現した。また、無線センタネットワークの大規模実験環境として、X-sensor を実現した。これらは、京都大学、大阪大学のチームを中心に関連分野の研究者を結集して構築されてきたものである。

A02 柱では、スケーラブル計算基盤テストベッド構築を目指して、世界的に類を見ない規模(2,000CPU 規模)の広域分散テストベッド InTrigger を整備し、大規模な分散ソフトウェア分野の研究を効果的に促進するとともに、大規模計算資源は他の柱において大量の計算を行うことに活用された。

A03 柱では、実世界インタラクション計測分析環境の構築をめざして、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・センサおよび生理計測装置によってリアルタイムで従来より一桁高い精度で計測する環境を実現してきた。特に、協調作業時におけるメディア、運動、生体データを同時に獲得ができる点が画期的であり、インタラクション支援における分野横断的な研究の進展と共通ソフトウェア開発が行われた。

さらに、A03 柱では柱内で行われた研究成果を統合してデモンストレーションする環境として、「ジャーナリストロボット」が実現された。ここに各研究班の成果が持ち寄せられ統合されることにより、成果の分かり易いアピールが可能となっている。

さらに、当初計画にはなかったが、B01 柱においても支援班活動を行うこととし、A01 柱の支援班からセンサーネットワーク、A02 柱の支援班から大規模なデータマイニング処理を行う計算基盤を活用し、しかも医療応用を想定したデータ収集を行う環境をプロトタイピングすることによって、各柱の支援班の基盤を融合する触媒プロジェクトとして機能させた。

更に新たに企画した活動として、米国 Microsoft 社との共同研究の実現である。これは支援班経費を使用しない活動ではあるが、本領域の「情報爆発」という研究コンセプトに対して Microsoft 社が賛同し、平成 22 年度に Microsoft 社の運営するクラウドコンピューティング環境 Azure が本領域の中のプロジェクトに対して提供されるというものである。この提供により、一年当たり、400 万 CPU 時間、20TB のストレージ、毎日最大 150GB のデータ量の転送という潤沢な計算リソースの使用が可能となった。支援班を中心に 7 つのプロジェクトを募り、パブリッククラウドをベースとした情報学研究の課題を検討するという一方で、本領域の研究で開発されたプログラムやデータベースの移植実験が行われる。

4. 研究成果

(1)各柱の年度順による活動成果

①A01 柱 平成 18 年度に開放型検索エンジン基盤 TSUBAKI の構築を開始した。超高速頻出パターンマイニングなどアルゴリズム、ウェブからの意見抽出などの新しい検索技術が提案された。平成 19、20 年度には、深い言語処理に基づくウェブ検索手法、ユーザのフィードバックに基づく再ランキングシステム、情報リンケージによる論文情報検索等に支援班環境が活用され、次世代サーチの研究が推進された。サーチ研究を加速するためのソフトウェアツールの Slothlib を公開し、研究班での利用に供した。また、「Yahoo! ラボ」や@nifty での公開など、実社会での活用に展開した。最終年度として日本語 Web 1.2 億ページをコーパスとして提供し、各班のサーチ機能の評価を行った。一方、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 X-sensor 2.0 として P2P プラットフォーム上への実装と提供を行った。

②A02 柱 平成 18 年度は、広域分散コンピューティングテストベッド InTrigger の構築を開始した。その上で研究成果の効率的な

キャッシュ方式により、1000 ノードでも 2 分程度で仮想クラスタを構築する技法などが実現された。平成 19、20 年度障害の自動検出の枠組みを構築し、MPICH の分散環境利用時に潜む 10 年来のバグの検出に成功した。またベタバイト級の Gfarm 分散ファイルシステムの開発、InTrigger を利用した大規模画像データからの超新星自動発見の国際コンテストなどを行った。平成 21 年度は、広域分散テストベッド InTrigger を更に拡充・機能を強化し、全国で 13 拠点、1300CPU コア以上、200 テラバイト以上の常用テストベッドとし、その上に大規模分散データ処理基盤を実現するための種々の研究成果を実証していった。平成 22 年度には InTrigger の整備は計 17 サイトとなり、研究の取り纏めの環境として最終的な整備を完了した。

③A03 柱 平成 18 年度から、支援班として開発したインタラクション計測分析環境 IMADE ルームを用いて、多人数、マルチモーダルな対話状況記録をコーパス化するためのツールを開発した。平成 19、20 年度には IMADE ルームを用いてさまざまな会話状況を設定し、会話実験記録の収集、分析を行った。平成 21、22 年度には、柱全体の研究成果の結集を図るため、計画班および支援班を中核とするタスクフォースを設け、情報ネットワーク社会と人とを結ぶインタラクションシステムとして、情報コンシェルジュとジャーナリストロボットの開発をさらに進めた。これらのシステムが備えるインタラクション機能の設計には、IMADE ルームで得られた成果が応用された。実世界インタラクション計測分析環境として京都大学のセンサールーム環境 IMADE におけるインタラクションコーパスの閲覧・分析・ラベリング環境を更に整備し、ソフトウェアパッケージを実現した。さらにこれをインタラクションマイニングへと展開できるような準備を整えた。

④B01 柱 センサーネットワークを用いた地域予防医療サービス実験をターゲットにおいて、平成 18 年度から実際のフィールドにおいて実証を開始した。これは制度的課題や組織・人材面の課題も積極的に考察しようとする点に特色がある。理論的基礎を固めることに注力し、安全・安心な情報生活空間に関する研究等をまず行った。平成 19 年度の特筆すべき成果として、センサーネットワークを用いた地域予防医療サービス実験がある。実証的研究体制を組織し、予防医療サービスを可能にする高度 IT 基盤技術の具現化、制度設計、政策提言による成果の社会への反映等に取り組んだ。平成 22 年度には、センサーネットワーク予防医療に関する研究は、新たな健康管理手法として多くの企業等から高い評価を得ることができた。生活習慣病予防医療実験として、センサーを活用した情報収

集のため、行動情報収集システムを作成し被験者を集めてデータ収集活動を実施した。

(2)支援班で行われた活動の特徴

支援班で供用する設備やコンテンツは、領域内すべての研究項目に係る計画研究、公募研究の間の有機的な結合を図る点に第一の特徴がある。

これによって、例えば、A03 柱において蓄積した計測データを、A01 柱との共同研究によって実世界インタラクション・コーパスとしてコンテンツ化し、さらに京都大学に設置されるサーバを介して A02 柱の計算基盤テストベッド InTrigger での大規模処理アルゴリズムを共同開発するなどの活動として実施された。従来は、メディア情報処理、ロボット工学、生体計測などの個別分野において断片的に取り組みられてきたインタラクションに関する研究の統合的な推進、分野を超えた連携が可能となった。

情報爆発に伴い、例えば Web 評価用テストコレクションの規模が、過去 2,3 年で 100G から 1T バイトへと 10 倍の規模に切り替わるなど、研究の遂行に必要な処理コストは、現状でも、通常一研究室で運営できる設備の限界に迫っている。支援班が提供する共通基盤を積極的に利用することで、研究経費の効果的な運用を図りつつ、個々の研究課題が各々において独自性を持つアイデアを発展させることが可能となり、領域全体としての多様性を保ち、格段の研究の発展を促す効果が期待された。これが第二の特徴である。

本研究によるコーパスの作成やテストベッドの構築は、単なる共通基盤整備にとどまらず、先進的な研究の創出やインセンティブ強化の役割を担っている。

スケーラブル計算基盤テストベッドは、従来、各拠点の計算基盤を独立に整備することにとどまりがちであった大規模計算機設備の整備方式を改め、当初から数多くの研究者間の共有研究基盤とすることを目的とした、新しい計算資源整備のモデルである。これにより、従来どんな一拠点でも不可能であった規模の実験、先進的な広域ネットワーク研究のインセンティブ強化、研究成果の研究者コミュニティでの共有、普及促進などの様々な効果が現れた。これが第三の特徴である。

評価基盤としてのテストコレクションや実世界インタラクション・コーパスの作成は、多くの研究グループで学術的に利用できる共通の学術資産として活用できるものであり、また当領域の研究ターゲットに即して設定されるタスクは先端的な研究フロンティアを提示し分野の研究を先導する役割も持っていると考えている。

さらに、今回の研究で実現した研究基盤は、各々が国内研究の拠点として重要な役割を果たすことが期待され、国際的にもユニーク

な位置づけを持つものである。

(3)支援班活動の効果

本領域において 60 を超える研究班の連携を深め、共同研究を加速するために、支援班の実現した共創プラットフォームは大いに貢献した。従来情報分野の研究では、研究環境は個々の研究者が整備するのが通例であった。今回、本領域では情報学においても大規模な研究基盤が必要であること、このような基盤が提供されると研究が加速されること、そして何よりも先端的研究を実施するための基盤整備が重要であることが実証された。以下に詳述するように、それぞれの研究基盤を実現するには、実際に先端的研究を実施していることがまず求められ、それを共有することにより、よい連携が生まれ発展していくことが分かった。基盤整備と先端的研究とは不可分である。今後とも、我が国の情報研究を振興するにはこのような環境がきわめて重要である。

さらに、今回 A02 柱で整備された分散計算基盤 InTrigger は、現在脚光を浴びているクラウドコンピューティング環境の先鞭を付ける構想であったことは明白である。学術研究用に特化したクラウド環境をいち早く提案したことが特筆すべき成果と言える。

平成 22 年 3 月には情報処理学会全国大会にて、支援班プラットフォームのデモ等の企画を行い、領域全体の成果のアピールを行った。また、米国マイクロソフトリサーチとの間での共同研究の枠組みを作り、同社がクラウドリソース Azure を本支援班に提供するプロジェクトを行った。Azure の資源を利用する本領域内のプロジェクトを 7 件選定し、すでに培われてきた研究成果の移植やクラウドでのインテンシブな研究環境の課題を明らかにすることも含めて、多方面からの研究の基盤を本支援班で提供した。

5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

[雑誌論文] (計 11 件)

1. 喜連川 優: 情報爆発をチャンスに ぶっとんだ発想をだそう 「グーグル超え」が必要
日経コンピュータ, 12 月 9 日号, 2009.
2. 相澤 彰子, 高須 淳宏, 深川 大路, 高久 雅生, 安達 淳: アカデミックリネージュ: 膨大な学術情報へのアクセスを支援するリネージュ基盤, 情報処理, Vol.49 No.8, pp. 935-938, 2008. 査読有
3. 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, pp. 931-934, Vol.49 No.8, 2008. 査読有

4. 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, pp. 927-930, 2008. 査読有
5. 角 康之, 西田 豊明, 坊農 真弓, 來嶋 宏幸: IMADE: 会話の構造理解とコンテンツ化のための実世界インタラクション研究基盤, 情報処理学会誌, Vol.49, No.8, pp.945-949, 2008. 査読有
6. 須藤 修: 情報爆発時代における知識社会形成ガバナンス, 人工知能学会誌, Vol.22 No.2, pp. 235-240, 2007. 査読有
7. 松山 隆司, 西田 豊明, 國吉 康夫: 情報爆発時代におけるヒューマンコミュニケーション基盤, 人工知能学会誌, Vol.22 No.2, pp. 229-234, 2007. 査読有
8. 松岡 聡, 柴山 悦哉, 近山 隆, 中島 達夫, 田浦 健次朗: 情報爆発時代における安全・安心 IT システム基盤, 人工知能学会誌, Vol.22 No.2, pp. 222-228, 2007. 査読有
9. 喜連川 優, 安達 淳, 田中 克己, 下條 真司, 黒橋 禎夫: 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, pp. 215-221, 2007. 査読有
10. 喜連川 優, 松岡 聡, 松山 隆司, 須藤 修, 安達 淳: 情報爆発時代に向けた新しい IT 基盤技術の研究, 人工知能学会誌, Vol.22 No.2, pp. 209-214, 2007. 査読有
11. Hideo Saito, Kenjiro Taura, Takashi Chikayama: Collective Operations for Wide-Area MessagePassing Systems Using Adaptive Spanning Trees., International Journal of High Performance Computing and Networking (IJHPCN), 2006. 査読有

[学会発表] (計 14 件)

1. Yasuyuki Sumi, Masaharu Yano, and Toyoaki Nishida: Analysis environment of conversational structure with nonverbal multimodal data, The Twelfth International Conference on Multimodal Interfaces and the Seventh Workshop on Machine Learning for Multimodal Interaction (ICMI-MLMI 2010), Beijing, China, November 2010. 査読有
2. Satoshi Matsuoka: Clusters, Clouds and Commerce, IEEE Cluster Computing Conference 2009, Sep 3, 2009, New Orleans, US. 査読有
3. 義久 智樹, 神崎 映光, 原 隆浩, 石 芳正, 寺西 裕一, 下條 真司: 複数拠点統合型センサネットワークのためのモバイルエージェントを用いたデータ収集システム, 電子情報通信学会技術報告

- (USN2009-24), July 17, 2009, 京都府相楽郡.
4. Akimitsu Kanzaki, Takahiro Hara, Yoshimasa Ishi, Tomoki Yoshihisa, Yuuichi Teranishi, and Shinji Shimojo: A Sensor Network Testbed Integrating Multiple Networks, MDM 2009, May 18, 2009, Taipei, Taiwan. 査読有
 5. Akimitsu Kanzaki, Takahiro Hara, Yoshimasa Ishi, Tomoki Yoshihisa, Yuuichi Teranishi, and Shinji Shimojo: X-Sensor: a Sensor Network Testbed Integrating Multi-Networks, DMIEW 2009, Mar. 16, 2009, Fukuoka, Japan. 査読有
 6. Keiji Shinzato, Tomohide Shibata, Daisuke Kawahara, Chikara Hashimoto, Sadao Kurohashi: "TSUBAKI: An Open Search Engine Infrastructure for Developing New Information Access Methodology", Third International Joint Conference on Natural Language Processing, pp. 189 – 196, 2008. 査読有
 7. Keiji Shinzato, Daisuke Kawahara, Chikara Hashimoto, Sadao Kurohashi: "A Large-Scale Web Data Collection as a Natural Language Processing Infrastructure", In Proceedings of the 6th International Conference on Language Resources and Evaluation (LREC08), pp. 2236 – 2241, 2008. 査読有
 8. 黒橋 禎夫: 構造的言語処理による情報検索基盤の構築と情報分析, 第1回情報アクセスシンポジウム (IAS), Oct 21 2009, 北海道大学
 9. Keiji Shinzato, Tomohide Shibata, Daisuke Kawahara, Chikara Hashimoto, Sadao Kurohashi: TSUBAKI: An Open Search Engine Infrastructure for Developing New Information Access Methodology, Third International Joint Conference on Natural Language Processing, Hyderabad, India, Jan 200 査読有
 10. 神崎 映光, 原 隆浩, 中山 浩太郎, 若宮直紀, 下條 真司: 複数拠点統合型センサネットワークテストベッド X-Sensor の設計と実装, 情報処理学会 DPS ワークショップ, Nov. 1, 2007, 石川県加賀市.
 11. Mayumi Bono, Yasuyuki Sumi, and Toyoaki Nishida: Towards achieving complex medical engineering to understand conversational dynamics, 2007 IEEE/ICME International Conference on Complex Medical Engineering (CME2007), Beijing China, May 2007. 査読有
 12. Masaru Kitsuregawa: Challenge for Info-plosion, The 18th International Conference on Algorithmic Learning

Theory (ACT2007), Tohoku University, Oct 2007. 査読有

13. Takayoshi Shiraki, Hideo Saito, Yoshikazu Kamoshida, Katsuhiko Ishiguro, Ryo Fukano, Tatsuya Shirai, Kenjiro Taura, Mihoko Otake, Tomomasa Sato, Nobuyuki Otsu: "Real-Time Motion Recognition Using CHLAC Features and Cluster Computing", The 3rd IFIP International Conference on Network and Parallel Computing, 2006. 査読有
14. Yuuki Horita, Satoshi Ito, Kenji Kaneda, Takuya Nanri, Yasuyuki Shimohata, Kenjiro Taura, Mihoko Otake, Tomomasa Sato, Nobuyuki Otsu: "High Precision Gait Recognition Using a Large-Scale PC Cluster", The 3rd IFIP International Conference on Network and Parallel Computing, 2006. 査読有

[図書] (計 1 件)

1. Takahiro Hara, Vladimir I. Zadrozny, Erich Buchmann, Wireless Sensor Network Technologies for Information Explosion Era (Book Series: Studies in Computational Intelligence), Springer-Verlag, 2010

[産業財産権]

- 出願状況 (計 0 件)
- 取得状況 (計 0 件)

[その他]

ホームページ等

<http://www.infoplosion.nii.ac.jp/info-plosion/html/S/>

6. 研究組織

(1) 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

(2) 研究分担者

田中 克己 (Tanaka Katsumi)

京都大学・大学院情報学研究科・教授

研究者番号: 00127375

西田 豊明 (Nishida Toyoaki)

京都大学・大学院情報学研究科・教授

研究者番号: 70135331

國吉 康夫 (Kuniyoshi Yasuo)

東京大学・大学院情報理工学系研究科・授

研究者番号: 10333444

須藤 修 (Sudoh Osamu)

東京大学・大学院情報学環・教授

研究者番号: 10179286

黒橋 禎夫 (Kurohashi Sadao)
京都大学・大学院情報学研究科・教授
研究者番号：50263108
原 隆弘 (Hara Takahiro)
大阪大学・大学院情報科学研究科・准教授
研究者番号：20294043
松岡 聡 (Matsuoka Satoshi)
東京工業大学・学術国際情報センター・教授
研究者番号：20221583
田浦 健次郎 (Taura Kenjiro)
東京工業大学・学術国際情報センター・教授
研究者番号：20221583
建部 修見 (Tatebe Osami)
筑波大学・システム情報工学研究科・准教授
研究者番号：70357432
棟朝 雅晴 (Munetomo Masaharu)
北海道大学・情報基盤センター・准教授
研究者番号：00281783
廣津 登志夫 (Hirotsu Toshio)
法政大学・情報科学部・教授
研究者番号：10378268
松原 仁 (Matsubara Jin)
公立はこだて未来大学・システム情報科学部・教授
研究者番号：50325883

(3)連携研究者

下條 真司 (Shimojyo Shinji)
情報通信研究機構・連携研究部門・上席研究員
研究者番号：00187478
千葉 滋 (Chiba Shigeru)
東京工業大学・情報理工学研究科・教授
研究者番号：80282713
湯浅 太一 (Yuasa Taichi)
京都大学・情報学研究科・教授
研究者番号：60158326
松山 隆司 (Matsuyama Takashi)
京都大学・情報学研究科・教授
研究者番号：10109035
近山 隆 (Chikayama Takashi)
東京大学・新領域創成科学研究科・教授
研究者番号：40272380
近堂 徹 (Kondo Toru)
広島大学・情報メディア教育研究センター・助教
研究者番号：90437575
河野 健二 (Kono kenji)
慶應義塾大学・理工学部・准教授
研究者番号：90301118
岡本 正宏 (Okamoto Masahiro)

九州大学大学院・農学研究院・教授
研究者番号：40211122
合田 憲人 (Aida kento)
国立情報学研究所・リサーチグリッド研究開発センター・教授
研究者番号：80247212
鎌田 十三郎 (Kamata Jyusaburo)
神戸大学・工学部・助教
研究者番号：20304131
喜連川 優 (Kitsuregawa Mararu)
東京大学・生産技術研究所・教授
研究者番号：40161509
山名 早人 (Yamana Hayato)
早稲田大学・理工学術院・教授
研究者番号：40230502
中村 豊 (Nakamura Yutaka)
九州工業大学・情報科学センター・准教授
研究者番号：40346317
小林 広明 (Kobayashi Hiroaki)
東北大学・サイバーサイエンスセンター・教授
研究者番号：40205480
中島 浩 (Nakajima Hiroshi)
京都大学・学術情報メディアセンター・教授
研究者番号：10243057