

年報 第 9 号

2007 年度

Annual Report No.9

2007



東京大学情報基盤センター

Information Technology Center
the University of Tokyo

目 次

巻頭言	1
PART 1 概況	3
組 織	5
組織図	5
現員	5
教職員	6
平成19年度中の人事異動	11
東京大学情報基盤センター運営委員会委員名簿	14
予 算	15
収入・支出	15
外部資金	15
補助金等	17
平成19年度科学研究費採択状況	17
平成19年度受託研究費受入状況	18
平成19年度共同研究費受入状況	19
平成19年度 政府系委託費受入状況	19
平成19年度奨学寄附金受入状況	20
PART 2 研究活動報告	21
情報メディア教育研究部門	23
山口 和紀	25
システムの管理コスト低減とデータ分析手法についての研究	
柴山 悦哉	27
セキュアでデペンダブルなソフトウェアのためのテストの自動化に関する研究	
田中 哲朗	32
ゲームプログラミングに関する研究	
尾上 能之	35
情報システムを用いた教育研究支援	
関谷 貴之	40
教育支援システムの開発と運用	

丸山 一貴	50
web 閲覧支援とデバッグ支援の研究	
図書館電子化研究部門	53
中川 裕志	55
機械学習とそのテキスト処理への応用	
二宮 崇	63
自然言語構文解析とソーシャル辞書	
吉田 稔	71
検索支援のためのテキストマイニング技術	
—同義語・接続語の高速検索と Web 文書レイアウト解析—	
清田 陽司	78
図書館情報資源と Web 情報資源の融合	
～Web 2.0 時代の電子図書館の実現に向けて～	
清水 伸幸	87
知識発見と機械翻訳の自然言語処理要素技術	
—同義語獲得、係り受け解析、機械学習テンプレートライブラリー	
王 玉馨	92
Web とテキスト情報処理についての研究	
鍋島 彰崇	100
レポート文書の文のカテゴリ分け	
キャンパスネットワーク研究部門	107
若原 恭	109
ネットワークの設計・運用・保守・管理の合理化を図る技術	
—アドホックネットワーク、ネットワークセキュリティ、車車間通信、次世代衛星ネットワーク、 及び惑星間インターネット—	
中山 雅哉	117
広域分散環境の高度基盤技術に関する研究	
加藤 朗	121
ネットワーク運用に関する研究	
小川 剛史	129
放送型データ配信を用いた情報システム構築に関する研究	
中村 文隆	136
高帯域遅延積環境における高効率転送	
中村 誠	138
教職員のための情報共有サイトの開発と高品質なインターネット中継システムの研究	

関谷 勇司	141
DNS の安定運用に向けた分散計測システムの設計と構築に関する研究	
スーパーコンピューティング研究部門	147
金田 康正	149
スーパーコンピューターに基づく大規模数値計算に関する研究	
石川 裕	160
次世代コンピュータシステムの研究開発	
ーオープンスパコン、グリッド、ディペンダブルシステムー	
佐藤 周行	170
セキュリティを保證する言語処理系の研究	
Optimization Verifying Compiler, Document Carrying Authorization, and LoA of Identities	
佐藤 文俊	176
創薬・バイオ新基盤の技術開発へ向けたタンパク質反応全電子シミュレーション・システムの研究開発	
黒田 久泰	188
4 倍精度演算ライブラリの高速度および疎行列ライブラリの自動チューニングの効果に関する研究	
松葉 浩也	198
スーパーコンピューターのための高速通信機構	
片桐 孝洋	207
ソフトウェア自動チューニングおよび高性能数値計算ライブラリの研究と HPC 教育	
吉廣 保	220
タンパク質量子化学計算支援統合環境の研究開発	
PKI プロジェクト	227
西村 健	229
大学における PKI の整備と既存システムとの連携	
 PART 3 教育・サービス活動報告	233
情報メディア教育部門	235
情報メディア教育部門概要	237
教育用計算機システム運用報告	238
学内組織向けメールサーバ(MAILHOSTING)運用報告	248
WEB PARK サービス運用報告 (2007 年 4 月ー2008 年 3 月)	250
遠隔講義支援サービス運用報告 (2007 年 4 月ー2008 年 3 月)	263
CFIVE 運用報告 (2007 年 4 月ー2008 年 3 月)	269
図書館電子化部門	275
図書館電子化部門概要	277

デジタルコンテンツサービス	278
学術情報リテラシー支援	282
図書館関係システム運用・管理	288
キャンパスネットワークキング部門	293
キャンパスネットワークキング部門概要	295
東京大学キャンパスネットワークシステム(UTnet3)の運用管理	298
セキュリティ対応	305
東京大学情報システム緊急対応チーム(UT-CERT)との連携	307
学内ソフトウェアライセンス	309
スーパーコンピューティング部門	311
スーパーコンピューティング部門概要	313
スーパーコンピューティング業務	314
先端の大規模計算シミュレーションプログラム利用サービス	327
—文部科学省先端研究施設共用イノベーション創出事業【産業戦略利用】—	
PKI プロジェクト	335
PKI プロジェクト	337
PART 4 その他	341
平成19年度委員会委員等	343
平成19年度講習会・セミナー	347

巻 頭 言

東京大学情報基盤センター長 米 澤 明 憲

少し遅くなりましたが、東京大学情報基盤センターの年報をお届けいたします。この年報にセンターの活動が報告されている平成19年度には、今まであまりしていない経験をセンターがしました。

1つには、文部科学省の学位授与機構による大学評価のための資料作成が本センターにも必要になりました。この作業の過程で、本センターの研究・教育・業務活動が外からどのように見られているのか、どのように見てほしいかなどをより真剣に考えることになりました。お届けする年報は、ある意味でその過程での検討結果が現れているとも言えます。

2つ目は、スーパーコンピューティング部門の目覚ましい進展です。平成20年6月より稼動しておりますT2Kオープンスパコン（東大版）はセンターの教員陣が設計仕様を策定し調達した超並列型のスパコンです。このような方式で全国共同利用のためのスパコンを導入することは、日本ではこれまでなかったことです。この新スパコンの導入のための建物、運用方式、などの準備ために本部門が費やした時間と労力は多大なので、この部門の教職員の苦労は計り知れないものです。しかし、このことは本年報の中にはなにも反映されておりません。さらにこの部門では、7大学センターと共同で文部科学省先端研究施設共用イノベーション創出事業「産業戦略利用」を開始し、20社近い民間会社にスパコンを利用してもらうことになりました。この事業開始の準備と運用に関わる作業も大変大きいものですが、この年報にはこの事業の概要が少し述べられている程度です。

3つ目は、図書館電子化部門の活躍です。この部門は東大内の総合図書館等のサービスの高度化にこれまでも大きく貢献してきました。しかし、これに留まらず、図書の案内や関連図書の推奨や案内ができる、ユーザにとって使い易いガイドシステムを本部門で構築し総合図書館の正式サービスとして配備するとともに、国内の他図書館でもこのガイドシステムが利用できるようにしました。これも、センターの活動として、教員の高い研究力に基づいた、評価の高いものではないかと思います。これについては、本年報でも紹介されています。

以上の3つは、センターの新しい活動の分かりやすい例ですが、全般的にもセンターの活動レベルがかなり上昇しております。この年報で、それを読み取っていただければ嬉しい限りです。編集の労をとってくださった小川剛史先生に敬意を表します。

平成20年6月20日

PART 1

概 況

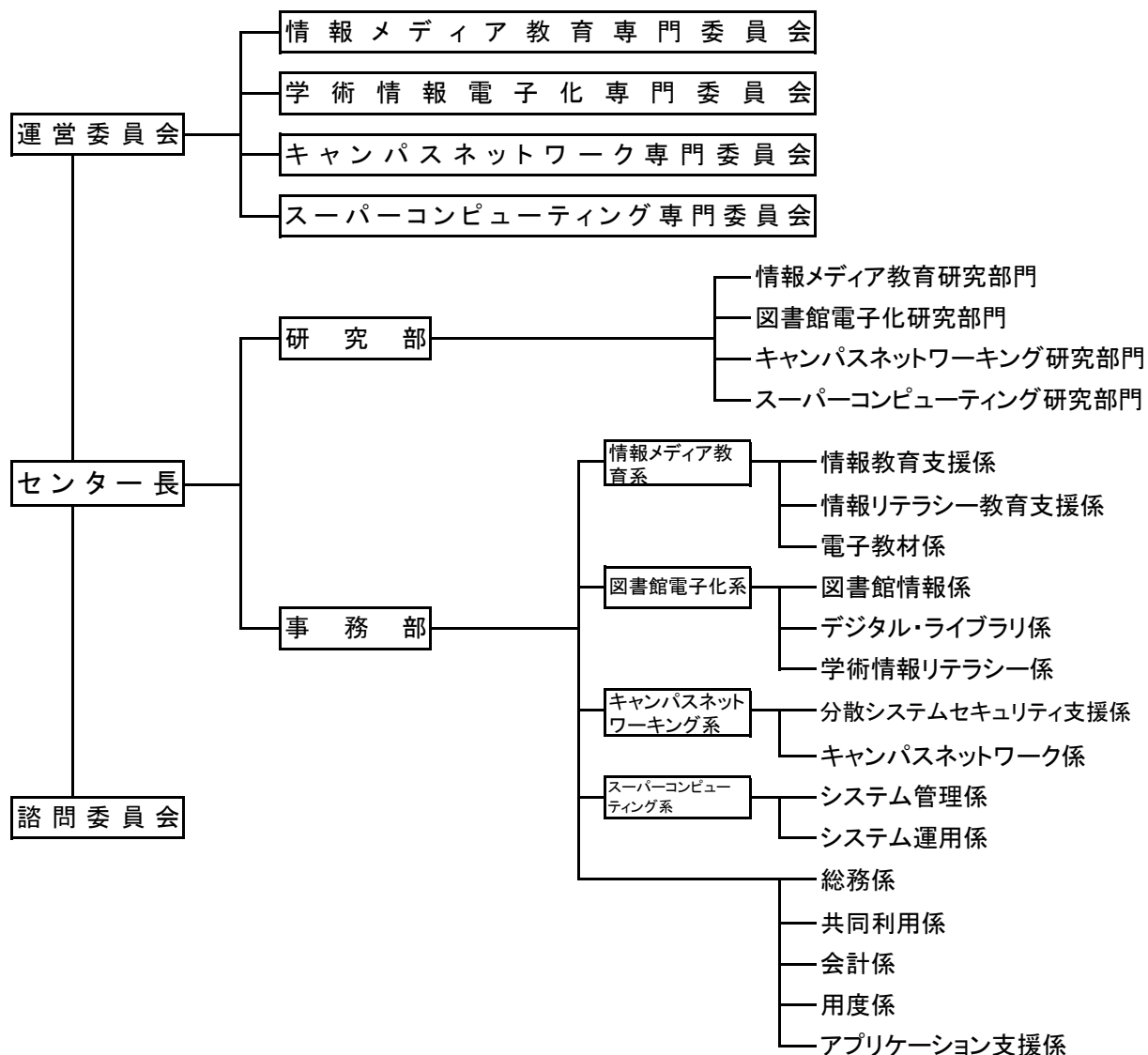
組 織

予 算

補助金等

組 織

組織図



現 員

2008. 3. 31

セ ン タ ー 長			兼 務
研究部	教 員	教 授	3
		准 教 授	5
		講 師	3
		助 教	9
	小 計		20
事務局	職 員	事務職員	24
		技術職員	24
	小 計		48
合 計			68

教職員

(1) 2008年3月31日現在

情報基盤センター長 米 澤 明 憲
(情報理工学系研究科教授)
秘書(事務補佐員) 渡 邊 かがり

研究部

情報メディア教育研究部門

教授(兼務)	柴 山 悦 哉
准教授	田 中 哲 朗
講 師	尾 上 能 之
助 教	関 谷 貴 之
助 教	丸 山 一 貴
教務補佐員	寺 脇 由 紀

図書館電子化研究部門

教 授	中 川 裕 志
講 師	二 宮 崇
助 教	吉 田 稔
助 教	清 田 陽 司
特任助教	清 水 伸 幸
リサーチフェロー	王 玉 馨
リサーチフェロー	鍋 島 彰 崇
事務補佐員	吉 富 美 樹
事務補佐員	赤 田 沙 織
教務補佐員	范 暁 蓉
教務補佐員	田 林 洋 一

キャンパスネットワークング研究部門

教 授	若 原 恭
准教授	中 山 雅 哉
准教授	加 藤 朗
講 師	小 川 剛 史
助 教	中 村 文 隆

助 教	中 村 誠
助 教	関 谷 勇 司
事務補佐員	川 崎 しのぶ
事務補佐員	吉 澤 文 代

スーパーコンピューティング研究部門

教 授	金 田 康 正
教 授 (兼務)	石 川 裕
准教授	佐 藤 周 行
准教授	佐 藤 文 俊
助 教	黒 田 久 泰
助 教	松 葉 浩 也
特任准教授	片 桐 孝 洋
特任准教授	吉 廣 保
事務補佐員	亀 田 文美代
事務補佐員	高 橋 弘 美

PKI プロジェクト

特任助教	西 村 健
事務補佐員	伊 東 雅 美

事務部

事務長	菊 池 力
副事務長	大日方 一 男
専門員	大 塚 浩 一
(企画調整担当)	
主査	佐 藤 安一郎
(情報業務担当)	
主査	早 野 裕 士
(情報基盤整備担当)	
主査	西 澤 明 生
(共同利用担当)	
主査	富 田 正 明
(予算執行担当)	
専門職員	櫻 田 芳 男
(スーパーコンピューティング担当)	

総務係

係 長	米 山 浩
主 任	灰 塚 毅 弘
技術職員	佐々木 一 孝
事務補佐員	麦 谷 真 弓

共同利用係

係 長	(共同利用担当主査 兼務)
主 任	木 崎 信 一
主 任	草 開 泰 之

会計係

係 長	野 呂 清 隆
主 任	小 林 拓 志

用度係

係 長	(予算執行担当主査 兼務)
主 任	岡 戸 彰 二
一般職員	三 上 匠

[事務電算化支援系] チーフ 佐 藤 安一郎

アプリケーション支援係

係 長	(情報業務担当主査 兼務)
主 任	入 江 健 司
一般職員	大 島 大 輔
技術職員	牛 丸 守

[情報メディア教育支援系] チーフ 平 野 光 敏

情報教育支援係

係 長	平 野 光 敏
主 任	山 本 和 男
技術専門職員	伊 藤 真 之
技術補佐員	小 藺 隆 弘
技術補佐員	宮 北 美 保
派遣職員	後 閑 直 子

情報リテラシー教育支援係

係長	増田均
主任	田川善教
技術職員	友西大
技術補佐員	安部達巳

電子教材係

係長	柏芳美
技術職員	坂井朱美
技術職員	佐藤孝明
技術補佐員	加藤康一
技術補佐員	永岡陽香

[図書館電子化支援系] チーフ 本多 玄

図書館情報係

係長	本多玄
一般職員	坂牧一博
一般職員	小松陽一

デジタル・ライブラリ係

係長	嶋邦宏
一般職員	西村昭子
一般職員	赤津愛美

学術情報リテラシー係

係長	吉野明美
一般職員	濱田智恵子
一般職員	相澤奈津子
派遣職員	渡部芳則

[キャンパスネットワークング支援系] チーフ 早野裕士

分散システムセキュリティ支援係

係長	椿山惣一郎
主任	佐野徹

組 織

キャンパスネットワーク係

係 長 中 村 昇 平

主 任 佐々木 馨

技術職員 松 本 浩 一

[スーパーコンピューティング支援系] チーフ 櫻 田 芳 男

専門職員 丹 下 藤 夫

システム管理係

係 長 井 爪 健 雄

主 任 石 崎 勉

システム運用係

係 長 有 賀 浩

主 任 佐 島 浩 之

平成19年度中の人事異動

(転入・昇任・配置換等)

2007年	4. 1	山口和紀	情報メディア教育研究部門教授(兼務)
2007年	4. 1	小川剛史	キャンパスネットワーク研究部門講師/大阪大学助手から
2007年	4. 1	片桐孝洋	スーパーコンピューティング研究部門特任准教授/新規 (任期 19.4.1~23.3.31)
2007年	4. 1	小出幸夫	情報リテラシー教育支援係/再雇用
2007年	7. 1	増田均	情報リテラシー教育支援係長/国文学研究資料館事業課システム管理係長から
2007年	7. 1	椿山惣一郎	分散システムセキュリティ支援係長/東京国立博物館情報課情報管理室専門職員から
2007年	7. 1	小松陽一	図書館情報係/国立情報学研究所基盤企画課連携システムチーム係長から
2007年	7. 1	大日方一男	副事務長/専門職員(スーパーコンピューティング担当)から
2007年	7. 1	早野裕士	主査(情報基盤整備担当)/分散システムセキュリティ支援係長から
2007年	7. 1	西澤明生	主査(共同利用担当)共同利用係長(兼務命)/共同利用係長から
2007年	7. 1	富田正明	主査(予算執行担当)用度係長(兼務命)/用度係長から
2007年	7. 1	櫻田芳男	専門職員(スーパーコンピューティング担当)/情報リテラシー教育支援係長から
2007年	7. 1	田川善教	情報リテラシー教育支援係主任/システム運用係主任から
2007年	7. 1	佐島浩之	システム運用係主任/キャンパスネットワーク係主任から
2007年	7. 1	坂井朱美	電子教材係/分散システムセキュリティ支援係から
2007年	8. 1	王玉馨	図書館電子化研究部門リサーチフェロー/図書館電子化研究部門教務補佐員から (任期 19.8.1~24.7.31)
2007年	9. 1	范曉蓉	図書館電子化研究部門教務補佐員/新規

組 織

2007年10.1	柴山悦哉	情報メディア教育研究部門教授（兼務）
2007年10.1	丸山一貴	情報メディア教育研究部門助教／電気通信大学助教から（任期 19.10.1～24.9.30）
2007年10.1	吉廣保	スーパーコンピューティング研究部門特任准教授／新規（任期 19.10.1～24.3.31）
2007年10.1	鍋島彰崇	図書館電子化研究部門リサーチフェロー／新規（任期 19.10.1～20.3.31）
2007年10.1	田林洋一	図書館電子化研究部門教務補佐員／新規
2007年10.1	草開泰之	共同利用係主任／用度係主任から
（転出・退職等）		
2007年4.1	山口和紀	情報メディア教育研究部門教授／総合文化研究科教授へ
2007年7.1	前田光教	情報リテラシー教育支援係主任／国文学研究資料館事業課システム管理係長へ
2007年7.1	小川大典	電子教材係主任／東京国立博物館情報課情報管理室専門職員へ
2007年7.1	小出幸夫	情報リテラシー教育支援係／生産技術研究所附属千葉実験所へ
2007年7.1	中村恭子	図書館情報係／国立情報学研究所基盤企画課総活・研修チーム係長へ
2007年9.30	山口和紀	情報メディア教育研究部門教授（兼務）／任期満了
2007年9.30	滝本友彦	システム管理係／辞職
2008年2.15	古木景子	学術情報リテラシー係／辞職
2008年3.31	加藤朗	キャンパスネットワークワーキング研究部門准教授／辞職
2008年3.31	尾上能之	情報メディア教育研究部門講師／辞職 文科省初等中等教育局教科書課調査官へ
2008年3.31	鍋島彰崇	図書館電子化研究部門リサーチフェロー／任期満了
2008年3.31	菊池力	事務長／定年退職
2008年3.31	牛丸守	アプリケーション支援係技術職員／定年退職
2008年3.31	櫻田芳男	専門職員（スーパーコンピューティング担当）／勸奨退職
2008年3.31	相澤奈津子	学術情報リテラシー係／任期満了

2008年 3. 31	寺 脇 由 紀	情報メディア教育研究部門教務補佐員／任期満了
2008年 3. 31	田 林 洋 一	図書館電子化研究部門教務補佐員／任期満了

東京大学情報基盤センター運営委員会委員名簿

任期：平成19年4月1日～平成21年3月31日

平成19年4月1日現在

氏 名	所 属 ・ 職 名	適 用
米 澤 明 憲	情報基盤センター長	規則第3条第1号
金 田 康 正	情報基盤センター・教授	規則第3条第2号
中 川 裕 志	情報基盤センター・教授	規則第3条第2号
若 原 恭	情報基盤センター・教授	規則第3条第2号
太 田 勝 造	大学院法学政治学研究科・教授	規則第3条第3号
平 尾 公 彦	大学院工学系研究科・教授	規則第3条第3号
堀 浩 一	大学院工学系研究科・教授	規則第3条第3号
佐 藤 隆 夫	大学院人文社会系研究科・教授	規則第3条第3号
青 木 秀 夫	大学院理学系研究科・教授	規則第3条第3号
矢 島 美 寛	大学院経済学研究科・教授	規則第3条第3号
山 口 和 紀	大学院総合文化研究科・教授	規則第3条第3号
山 口 泰	大学院総合文化研究科・教授	規則第3条第3号
相 田 仁	大学院新領域創成科学研究科・教授	規則第3条第3号
石 川 裕	大学院情報理工学系研究科・教授	規則第3条第3号
喜連川 優	生産技術研究所・教授	規則第3条第3号
鈴 木 宏 正	先端科学技術研究センター・教授	規則第3条第3号
西 郷 和 彦	附属図書館長	規則第3条第4号

オブザーバー 武川 正吾 総長補佐（人文社会系研究科教授）

予 算

収入・支出

2007 年度決算額

収入

区 分	決算額 (千円)
寄 附 金 収 入	3, 2 0 0
受 託 研 究 等 収 入	1 8 2, 0 6 0
自 己 収 入	2 5, 2 9 5
計	2 1 0, 5 5 5

支出

区 分	決算額 (千円)
人 件 費	5 6 3, 8 6 6
物 件 費	3, 4 5 6, 3 2 7
計	4, 0 2 0, 1 9 3

外部資金

1) 奨学寄附金受入(2007 年度)

内 訳	受入件数	受入額 (千円)
情報メディア教育研究部門	1	1, 2 0 0
図書館電子化研究部門		
キャンパスネットワーク研究部門		
スーパーコンピューティング研究部門	1	2, 0 0 0
計	2	3, 2 0 0

2) 受託研究(2007 年度)

内 訳	受入件数	受入額 (千円)
情報メディア教育研究部門		
図書館電子化研究部門	2	1 4, 1 4 8
キャンパスネットワーク研究部門	3	5 3, 2 7 0
スーパーコンピューティング研究部門	1	2 3, 5 4 2
計	6	9 0, 9 6 0

3) 共同研究(2007年度)

内 訳	受入件数	受入額(千円)
情報メディア教育研究部門		
図書館電子化研究部門		
キャンパスネットワーク研究部門		
スーパーコンピューティング研究部門	1	3,000
計	1	3,000

4) 政府系委託費(2007年度)

内 訳	受入件数	受入額(千円)
情報メディア教育研究部門		
図書館電子化研究部門	1	15,600
キャンパスネットワーク研究部門	1	5,000
スーパーコンピューティング研究部門	1	67,500
計	3	88,100

5) 科学研究費補助金(2007年度)

内 訳	受入件数	受入額(千円)
情報メディア教育研究部門		
図書館電子化研究部門	4	24,310
キャンパスネットワーク研究部門	1	1,500
スーパーコンピューティング研究部門	2	2,730
計	7	28,540

補助金等

平成19年度 科学研究費採択状況

研究代表者	教授 中川 裕志
研究種目	特定領域研究
研究期間	平成19～21年度
研究課題	多言語Webテキストからの知識マイニングに関する研究
研究費	5,100,000円(平成19年度)
研究代表者	教授 中川 裕志
研究種目	基盤研究(A)
研究期間	平成17～19年度
研究課題	Webテキストからの知識抽出支援システムに関する研究
研究費	13,910,000円(平成19年度)
研究代表者	准教授 佐藤 周行
研究種目	基盤研究(C)
研究期間	平成17～19年度
研究課題	最適化検証つきコンパイラの理論と実証のための展開
研究費	1,040,000円(平成19年度)
研究代表者	特任准教授 片桐 孝洋
研究種目	基盤研究(C)
研究期間	平成18～19年度
研究課題	数値計算と組み込みシステムのための自動チューニング方式
研究費	1,690,000円(平成19年度)
研究代表者	講師 二宮 崇
研究種目	若手研究(A)
研究期間	平成19～21年度
研究課題	主辞駆動句構造文法のための統計同期文法による機械翻訳
研究費	3,900,000円(平成19年度)

研究代表者 助教 清田 陽司
研究種目 若手研究 (B)
研究期間 平成 18～19 年度
研究課題 Web 上の情報資源と図書館情報資源を統合的に利用する
情報探索支援システム
研究費 1,400,000 円(平成 19 年度)

研究代表者 講師 小川 剛史
研究種目 若手研究 (B)
研究期間 平成 19～20 年度
研究課題 放送型データ配信を用いた自律移動型センサネットワーク
に関する研究
研究費 1,500,000 円(平成 19 年度)

平成 19 年度 受託研究費受入状況

研究代表者 教授 中川 裕志
相手機関名 (株)日本航空インターナショナル
研究期間 平成 19 年度
研究課題 『情報大航海プロジェクトー新総合安全運航支援システム』
に関わる基礎技術の研究
研究費 11,148,261 円(平成 19 年度)

研究代表者 教授 中川 裕志
相手機関名 (株)富士通研究所
研究期間 平成 19 年度
研究課題 高度なテキスト分析支援技術の開発
研究費 3,000,000 円(平成 19 年度)

研究代表者 教授 石川 裕
相手機関名 (独)科学技術振興機構
研究期間 平成 19 年度
研究課題 並列・分散型組込みシステムのためのデバングラフィック
システムイメージ OS
研究費 23,542,000 円(平成 19 年度)

研究代表者 教授 若原 恭
 相手機関名 情報・システム研究機構
 研究期間 平成 19 年度
 研究課題 最先端学術情報基盤構築事業
 研究費 42,800,000 円(平成 19 年度)

研究代表者 准教授 中山 雅哉
 相手機関名 (独) 科学技術振興機構
 研究期間 平成 19 年度
 研究課題 都市基盤広域センシングに関する手法・データ伝送と処理・
 統合リスク評価そしてフィールド適応に関する研究
 研究費 9,315,000 円(平成 19 年度)

研究代表者 准教授 加藤 朗
 相手機関名 (株) 日本レジストリサービス
 研究期間 平成 19 年度
 研究課題 BGP Anycast 環境下における DNS の運用と測定
 研究費 1,155,000 円(平成 19 年度)

平成 19 年度 共同研究費受入状況

研究代表者 教授 石川 裕
 相手機関名 (独) 国立環境学研究所
 研究期間 平成 19 年度
 研究課題 スーパーコンピュータを用いた GOSAT データ処理に関する
 研究
 研究費 3,000,000 円(平成 19 年度)

平成 19 年度 政府系委託費受入状況

研究代表者 教授 石川 裕
 相手機関名 文部科学省
 研究期間 平成 19 年度
 研究課題 先端の大規模計算シミュレーションプログラム利用サービ
 ス
 研究費 67,500,000 円(平成 19 年度)

研究代表者 教授 中川 裕志
相手機関名 文部科学省
研究期間 平成 19 年度
研究課題 日中・中日言語処理技術の開発研究
研究費 15,600,000 円(平成 19 年度)

研究代表者 准教授 中山 雅哉
相手機関名 文部科学省
研究期間 平成 19 年度
研究課題 大規模データ収集・発信のための高次ネットワークの基盤
の構築
研究費 5,000,000 円(平成 19 年度)

平成 19 年度 奨学寄附金受入状況

研究代表者 特任准教授 片桐 孝洋
相手機関名 (株)マイクロソフト
研究費 2,000,000 円(平成 19 年度)

研究代表者 助教 関谷 貴之
相手機関名 (株) トランスウェア
研究費 1,200,000 円(平成 19 年度)

PART 2

研究活動報告

情報メディア教育研究部門

図書館電子化研究部門

キャンパスネットワーキング研究部門

スーパーコンピューティング研究部門

PKI プロジェクト

情報メディア教育研究部門

山口 和紀

システムの管理コスト低減とデータ分析手法についての研究

柴山 悦哉

セキュアでデペンダブルなソフトウェアのためのテストの自動化に関する研究

田中 哲朗

ゲームプログラミングに関する研究

尾上 能之

情報システムを用いた教育研究支援

関谷 貴之

教育支援システムの開発と運用

丸山 一貴

web閲覧支援とデバッグ支援の研究

システムの管理コスト 低減と データ 分析手法についての研究

山口和紀

1 概要

教育用計算機システムについて実際的な研究を行うとともに、データ分析手法について原理的な研究を行った。

2 研究内容題目：教育用計算機システム

2.1 背景

教育用の計算機システムにおいて、管理コストが低く、安全に共用できる端末の提供方法が未確立だった。

2.2 内容

安全に共用できる端末の提供方法として、2004年にNetBootシステムを構築したが、導入から3年が経過したところで、その運用結果に基づいて評価を行った。

2.3 具体的成果

評価結果を公表した。([査読付 1])

3 研究内容題目：データ分析手法

3.1 背景

様々な分野でデータの量が増えてきており、多量のデータから本質的な要素を抽出することが重要になってきている。

3.2 内容

ゲーム、自然画像、ウェブ、シラバスに対してデータ分析手法を開発した。

3.3 具体的成果

ゲームにおいて棋譜データから勝率を推定し探索に利用する方法を考案した。([査読付 2] [査読付 3][査読付 4]) 大規模画像から edge detector を抽出する方法を考案し、実験的に効果を確認した。([査読付 5]) 非有基的集合論を用いてウェブから共通構造の抽出を行った。([査読付 6])概念の意味を用語集合で捉える方法を考案し、シラバスの分析に適用した。([発表 1])

4 成果要覧

査読付論文

[査読付1] 関谷貴之, 安東孝二, 尾上能之, 田中哲朗, 山口和紀: NetBootによる端末を用いた教育用計算機システムの開発と評価, 情報処理学会論文誌, Vol. 48, No. 4, pp.1651-1664, 2007年4月.

[査読付 2] 竹内聖悟, 林芳樹, 金子知適, 山口和紀, 川合慧: 勝率に基づく評価関数の評価と最適化, 情報処理学会論文誌, Vol. 48, No. 11, pp.3446-3454, 2007年11月.

[査読付 3] 竹内聖悟, 金子知適, 山口和紀: 情報量に基づく探索制御手法 --チェスにおける Singular Extension への応用--, 第12回ゲームプログラミングワークショップ, pp. 52-59, 2007年11月.

[査読付4] Shogo Takeuchi, Tomoyuki Kaneko, Kazunori Yamaguchi, Satoru Kawai : Visualization and Adjustment of Evaluation Functions Based on Evaluation Values and Win Probability, AAAI-07, pp. 858-863, 2007年7月.

[査読付5] Y. Matsuda and K. Yamaguchi : Extraction of Approximate Independent Components from Large Natural Scenes, ICONIP2007, Kitakyushu, LNCS 4984, Springer-Verlag, pp.635-642, 2007年11月.

[査読付6] I. Horie, K. Yamaguchi, and K. Kashiwabara : Pattern Detection from Web using AFA Set Theory, 9th ACM International Workshop on Web Information and Data Management, pp.113-120, 2007年11月.

その他の発表論文

[発表1] 関谷貴之 山口和紀: 講義同士の関係構造分析のためのシラバス処理方法の一検討, 情報処理学会第70回全国大会講演論文集 (分冊 4), つくば, pp.495-496, 2008年3月.

セキュアでデペンダブルなソフトウェアのための テストの自動化に関する研究

柴山悦哉

1 概要

ソフトウェアの社会基盤化が進み、重要インフラまでもがソフトウェアに依存するようになった今日、ソフトウェアの構築段階で紛れ込んだ些細なバグや運用段階のわずかな設定ミスが原因となり、大小さまざまな事故・事件が発生するようになった。抜本的な対策のためには、ソフトウェアの品質を向上させ、システムの安定運用を図る必要がある。しかし、これらの分野での長年にわたる研究努力にも関わらず、今日の段階で決定打となる方法は知られていない。ソフトウェアの品質向上は、人手に頼る部分が多いのが現状である。

一方、大量の計算資源を用いることで、定型的な処理を自動化し、ソフトウェアの品質を向上させる方向の研究や実践が近年拡大している。代表的な手法としては、コンパイルとリンクを頻繁に行うことでバグの早期発見を行う Continuous Build、可能な実行経路を網羅的に探索することで人手による発見が困難なバグを見つけるモデル検査、探索的な手法によりバグの存在箇所の絞り込みを自動的に行う Algorithmic Debugging、個々のコンピュータから少しずつ集めたモニタリング情報をもとに統計的にバグの推定を行う Statistical Debugging などが知られている。

2007年度には、これらのアイデアを一部取り入れながら、テストの自動化に関する研究として、以下のようなものを行った。

- 脆弱性テストに有効な入力データ生成と結果判定を自動化した Web アプリケーションのためのテスト手法
- ソースコードレベルのモデル検査を、ネットワーク型のアプリケーションのように外部システムを有するプログラムに適用する方式の提案
- Algorithmic Debugging の一種である Delta Debugging の手法を繰り返し適用することで regression bugs の発見を行う手法の提案

これらは、モデル検査、Delta Debugging, Taint 解析、Fuzz Testing などの手法を適宜組み合わせることで、現実的なアプリケーションに対するテストの自動化（テストケース生成の自動化を含む）を試みる研究である。

2 Web アプリケーションのための脆弱性テストの自動化に関する研究

2.1 背景

ソフトウェアのセキュリティホールに対する攻撃が日常的に発生するようになって久しい。従来は、C 言語で書かれた OS やアプリケーションプログラムに潜むメモリ関係のバグ（特にバッファオーバーフローバグ）に対する攻撃が多かったため、これが主要な問題と考えられていた。しかし、近年では、Web アプリケーションの脆弱性が大きな問題として浮上している。実際、報告されたソフトウェアの

脆弱性の中で、クロスサイトスクリプティングや SQL インジェクションの占める割合が増えており、今日ではバッファオーバーフローを凌ぐようになっている。

Web アプリケーションは、データベースサーバ、Web サーバ、Web ブラウザなどの既存インフラをうまく利用するため、同程度の機能のスタンドアロンアプリケーションと比べ、より容易に構築できる傾向がある。実際、数人で数ヶ月程度以下の開発案件も多く、技術者に求められる技能のレベルも、OS やインターネットの主要サーバと比べると平均的にはかなり低い。

この Web アプリケーション開発の敷居の低さは、単純に考えれば望ましいことである。しかし、必ずしも技術レベルの高くない開発者が多数参入して Web アプリケーションを大量に開発した結果、できあがったのはセキュリティホールが山であった。しかも、ネットワーク経由でアクセスしやすく、したがって攻撃しやすいものも少なくない。

Web アプリケーションの問題を考えるとときには、この量の多さと質の低さを念頭におく必要がある。既存の Web アプリケーションは、手作業で書き換えるには量が多すぎる。また、開発者に高度な技能を要求する対策技術は、その有効性が仮に高かったとしても、使える局面は限定的にならざるをえない。

2.2 内容

「背景」の節で述べたような状況設定のもとで、ソフトウェアの脆弱性とそれに対する攻撃の広がりに対応するために、セキュリティテストの自動化を促進する方式に関する研究を行った。

Web アプリケーションの脆弱性対策技術で重要なのは、高度な技能を必要とせず利用できることである。既に述べたように、Web アプリケーションの開発者の裾野は広く、その技術レベルは低いことを前提とせざるをえない。また、運用管理にすべてを託すのも危険である。

そこで、アプリケーションのインストールと若干の設定以外の全過程を自動化できる Web アプリケーションのセキュリティテスト方式を提案する。テストの自動化で本質的に重要なのは以下の 2 項目である。

- テスト用の適切な入力データの自動生成
- テスト実行の結果の自動評価

ここでの自動化は、JUnit などが実現している単体テストの自動化とは意味が異なる。JUnit 等で自動化されているのは、同じテストを何度も繰り返す部分であり、入力データも結果の妥当性判定基準も人間が手動で準備する。一方、ここで要求しているのは、入力データを自動的に生成し、テスト実行とその結果判定も自動的に行う方式である。

2.2.1 実行結果判定の自動化方式

セキュリティテストの実行結果判定は、何らかの異常が発生したかどうかを調べる作業である。たとえば、SQL インジェクションの場合、「想定外」の SQL クエリーが発行されたことを検知したい。簡単な例として、プログラム中で次の式により SQL クエリーを作る場合を考える。

```
"SELECT * FROM table WHERE user = ' " + name + "' "
```

人間がこの式を眺めると視覚的なパターンが見えてしまうが、それは錯覚であり、巧妙な入力データを与えるとこのパターンは破壊されてしまう。この例の場合、*name* の部分にシングルクオートを含む文字列が与えられると「想定外」となる。

人間には自然に見えてしまうこの種の視覚的パターンは、プログラムから抽出可能である。本研究では、プログラム中の視覚的パターンが、テスト実行時に保持されるかどうかを自動判定する。そのために、以下のような手法を用いた。

- 実行時に生成される各文字列オブジェクトに対し、1文字単位で taint 情報(その値が外部からの入力データに依存しているかどうかの情報)を追跡できるように処理系を改造する。
- SQL クエリーが発行される直前に構文解析を行い、tainted な文字が構文に影響を与えていないかどうかを判定する。具体的には、構文木の葉以外の場所に tainted な文字が出現すれば、影響を与えたものと見なす。

この方針に従って PHP の処理系を改造し、細粒度の taint 情報の管理機構を組み込んだ。現時点のプロトタイプでは、PHP 処理系が発行した SQL クエリーをログに記録し、それをオフラインで構文解析する方式でテストを行っている。

2.2.2 入力データ生成の自動化方式

セキュリティテストでは、開発者の想定を超えた入力データが必要になる。そこで、第一近似としては、既存の Fuzzer を用いて入力データの生成を行う。最近の Fuzzer は、SQL インジェクションなどの典型的なセキュリティバグに対し、問題が発生しやすいパターンの文字列を自動的に多数生成する機能を有している。しかし、Fuzzer は基本的にはランダムに入力データを生成するものであり、これだけではテストの網羅性は必ずしも高くない。そこで、実行のパスと各分岐の条件もログに記録できるように PHP の処理系を改造した。そして、以下のような手順でのテストを自動化するツールを作成した。

1. Fuzzer が自動生成した入力データを用いてテスト実行を行う。
2. 処理系のログからテスト実行時に通ったパスを調べ、カバレッジが低い場合には、さらにどのパスを通すべきかを求める。
3. Fuzzer が自動生成した入力データを少し変更し同様なテストを再度行う。
4. 単純な条件を持つブランチについては、入力データの変更が与えた影響を推定する。たとえば、`strcmp($s, "abc") < 0` という条件なら、`$s` と "abc" の差が大きくなる方向に変化したか小さくなる方向に変化したかを調べる。この情報に基づき、分岐したい方向へ制御を移すために必要な入力データをヒューリスティックに探索しつつテストを続ける。

2.3 具体的成果

PHP による Web アプリケーションの脆弱性は多数報告されている。これらの中から、最近報告された Web アプリケーションを 8 件選び、この手法を適用し、その有効性を確認した。比較のため、著名な Fuzz Testing ツールである Paros でも同様の実験を行った。この結果をまとめると次の表のようになる。

テスト対象	既知の脆弱性	Paros		提案手法	
		正検出	誤検出	正検出	誤検出
Neuronnews	3	0	0	7	0
shnews	2	0	0	9	0
en.faname	2	0	0	7	0
ajchat	1	0	0	1	0
taskfreak	1	1	1	3	0
tutorialCMS	2	0	2	4	0
PHPEchoCMS	1	1	0	1	0
GestDown	3	2	0	8	0
Total	15	4	3	40	0

Paros では、既知の脆弱性のうち一部 (脆弱性の件数で 4/15、アプリケーションの件数で 3/8) が検出されたに過ぎず、しかも誤検出率も高い (3/7)。一方、我々の方式では、すべてのアプリケーションで脆弱性の発見に成功しており、報告済みのものよりも多くの箇所で脆弱性を検出し、しかも誤検出は全く発生していない。

3 外部システムを有するプログラムのためのモデル検査手法

3.1 背景

一般に、ソフトウェアの品質を最終的に確認するためには、網羅的な検査が望まれる。これを近似的に実現するための手法として、モデル検査手法が近年注目を集めている。典型的には、ソフトウェアの内部状態を抽象化して有限状態機械と見なし、すべての可能な状態推移を網羅的に調べあげることで、検証すべき性質が常に成り立つ (あるいは成り立たないことがある) 等の性質を確認する手法である。

モデル検査の対象として抽象的なモデルを扱う方式のほかに、現実のプログラムを直接ツールに与えて検査を行う方式もある。前者の場合、現実のプログラムと抽象的なモデルの間に意味的ギャップが存在する可能性もあるが、後者の場合にはこのような心配をする必要がない。そのかわり、後者の方式の方が状態爆発の制御がより困難であり、優れたツールが必要になる。プログラムの静的解析技術を用いた最適化なども必要になることが多い。

どちらにしても、モデル検査のためには網羅的にすべての状態推移をたどる必要があり、通常、そのためにはバックトラック探索が行われる。ソースコードレベルですべての可能な状態推移をバックトラックしながら追う方式では、OS や外部プロセスの動的な状態や挙動などの扱いが問題となる。モデル検査の対象となるプロセス (群) の挙動はソースコードから得られるとしても、検査対象の外側に位置する OS やプロセスの挙動や状態の追跡は必ずしも容易ではない。

特に問題となるのが、検査対象の状態をバックトラックさせた時、外部システムもバックトラックさせる必要がある点である。外部システムのセマンティクスに関する情報がなかったり、外部システムまでモデル化すると状態数が増大したりするため、現実的には、外部システムをバックトラックさせるのは困難である。いかにして仮想的にバックトラックさせるかが課題となる。

3.2 内容

以上のような問題を解決するために、モデル検査の対象となるプロセスと外部システムの間、入出力のバックトラックをサポートするレイヤーを設ける手法を考案した。入出力のモデルとしてはストリーム通信を採用する。ネットワークやファイルを開くことでストリームが生成され、外界とのやりとりはすべてこのストリームを用いて行われる。

最初に入力値を読み込むときには、外部システムに対して入力リクエストを行い、必要な値を受け取る。同時に、この値をストリーム上の位置とともに中間レイヤーでキャッシュする。バックトラック後にもう一度同じ位置から読み出す場合には、中間レイヤーがリクエストを横取りし、キャッシュしていた値を返す。出力についても、2 回目以降は、中間レイヤーがリクエストを横取りし、外部システムから見て同じリクエストが複数回送られることを抑止する。この方式は、入出力の値が並行実行のタイミング等によって変化する場合には適用できないという限界がある。

3.3 具体的成果

Java 言語を対象に、このような中間レイヤーを含む実行系を構築した [査読付 1]。入出力性能についてはまだ改良の余地があり、1 回のライブラリ呼び出しで 100 バイトの入出力を行う場合には Sun の JVM と比較して 5 倍程度遅く、10,000 バイトの場合には 50% 程度遅くなっている。

4 成果要覧

査読付論文

- [査読付 1] C. Artho, B. Zweimller, A. Biere, E. Shibayama, and S. Honiden: Efficient Model Checking of Applications with Input/Output, Proceedings of Eurocast 2007, Lecture Notes in Computer Science, Vol. 4739, pp. 515–522, Springer-Verlag, December 2007.
- [査読付 2] C. Artho, E. Shibayama, and S. Honiden: Iterative Delta Debugging (poster), IFIP International Conference on Testing of Communicating Systems, June 2007.

その他の発表論文

- [発表 1] C. Artho, W. Leungwattanakit, M. Hagiya, Y. Tanabe, and E. Shibayama: Architecture-Aware Partial-Order Reduction to Accelerate Model Checking of Networked Programs, DIKU-IST Workshop, October 2007.
- [発表 2] C. Artho, W. Leungwattanakit, M. Hagiya, Y. Tanabe, and E. Shibayama: Hiding Backtracking Operations in Software Model Checking from the Environment, DIKU-IST Workshop, October 2007.

ゲームプログラミングに関する研究

田中哲朗

1 概要

ゲームプログラミング研究の一環として、2003年から他研究室と共同でオープンソース将棋プログラムの開発を行っている。さまざまな改良を加えた結果、世界でもトップレベルに達することができた。

また、パズルゲームの計算量に関する研究をおこなって、IQというパズルゲームのターン数判定問題がNP完全であることを証明した。

2 オープンソース将棋プログラムの作成

2.1 背景

1997年に、チェスプログラムが初めて人間の世界チャンピオンを破るという事件が起きたが、将棋もあと数年で同様のレベルに達することが期待され、コンピュータ将棋に関する研究は一番面白い時期を迎えようとしている。

しかし、コンピュータ将棋の研究を始める際に超えなくてはならない障壁は小さくない。まず、将棋は駒の種類がチェスより多く、独自の持ち駒ルールや、打ち歩詰め、二歩等のルールなどのため、ルール通り指させるだけでも大変である。

また、ある程度の強さを持つプログラム上で評価しないと、意味のある結果が得られないが、終盤の詰め将棋など十分な強さを持つものをバグなしで作るのは難しい。

そのため、コンピュータ将棋研究の全体的なレベルアップを狙って、最初から公開する前提で将棋プログラムの開発を2003年からおこなっている。将棋プログラム名はGPS将棋で、開発メンバーは田中研究室の学生の他に、総合文化研究科の教員、学生、会社員などが加わっている。

2.2 内容

この将棋プログラム自体は2003年度から開発を開始し公開もしているが、2007年度は

- 共有メモリ上の並列実行の効率化
- 統計的学習による探索延長パラメータの決定

などをテーマに改良をおこなった。

2.3 具体的成果

ライブラリはBSDライセンスで公開し、将棋プログラム全体はGPLで公開する形にした。これにより、純粋に研究目的の人も、あるいは商品売って儲けようという人もどちらも自由に使えるようにしながら、「コンピュータ将棋選手権何位」といった売り文句は使えないという形で公開することができた。

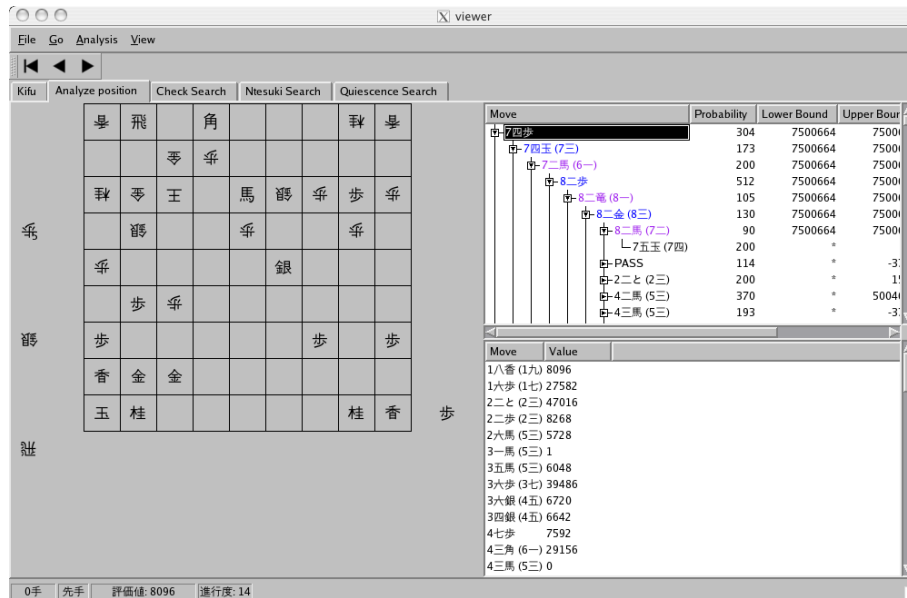


図 1: GPS 将棋の操作画面

作成したプログラム GPS 将棋(図 1)は、5 月の世界コンピュータ選手権で決勝には進めなかったものの全体で 10 位という好成績をおさめた。

3 パズルゲームに関する研究

3.1 背景

コンピュータを使ったパズルゲームは、従来のパズルと比較して、物理的な制約に縛られずに自由にルールを決められるという特徴がある。コンピュータが身近になった 30 年ほど前から、様々なパズルゲームが考案され、コンピュータゲームの市場の中でも一分野として認められ、一定のシェアを維持している。

人間にとって面白いパズルゲームのルール、ステージデータを作成するために、パズルゲームの問題を解くための計算量、人間が解く場合のルールのモデル化などの数理的な解析が今後重要になっていくと考えられる。

3.2 内容

I.Q というパズルゲームの計算量に関して解析をおこなった。I.Q は 10 年ほど前に発売されて大ヒットとなり、シリーズとして続編が発売され続けているパズルゲームである。

ある盤面(ステージ)が与えられた時に、決められたターン数でステージをクリアできるかという判定問題を定式化し、この問題のサイズ(盤面の大きさ)に対して必要な計算量を解析した。

3.3 具体的成果

NP 完全問題の代表的な問題の一つである 3-SAT 問題が I.Q のターン数判定問題に帰着可能であることを利用して、I.Q のターン数判定問題が NP 完全であることを証明することに成功した。

結果自体は予想されたことだが、I.Q を解くプログラムを作成したり、問題を作成するプログラムを作る際に有効な知見を得ることができた。

4 成果要覧

査読付論文

- [査読付 1] 田中哲朗: ボードゲーム「シンペイ」の完全解析, 情報処理学会論文誌, Vol. 48, No. 11, pp. 3470-3476, 2007 年 11 月.
- [査読付 2] 金子知適, 田中哲朗, 山口和紀, 川合慧: 駒の関係を利用した将棋の評価関数の学習, 情報処理学会論文誌, Vol. 48, No. 11, pp. 3438-3445, 2007 年 11 月.
- [査読付 3] 副田俊介, 美添一樹, 岸本章宏, 金子知適, 田中哲朗, マーティンミュラー: 証明数と反証数を用いた探索, 情報処理学会論文誌, Vol. 48, No. 11, pp. 3455-3462, 2007 年 11 月.
- [査読付 4] 関谷貴之, 安東孝二, 尾上能之, 田中哲朗, 山口和紀: NetBoot による端末を用いた教育用計算機システムの開発と評価, 情報処理学会論文誌, Vol. 48, No. 4, pp. 1651-1664, 2007 年 4 月.

その他の発表論文

- [発表 1] 水野秀一, 田中哲朗: I. Q Intelligent Qube の NP 完全性の証明, 情報処理学会ゲーム情報学研究会, 2007-GI-19, pp. 53-60, 東京, 2008 年 3 月.

公開ソフトウェア

- [公開 1] 金子知適, 田中哲朗, 副田俊介, 林芳樹, 竹内聖悟: osl-for-csa (バージョン 0.4), コンピュータ将棋協会のコンピュータ将棋選手権使用可能ライブラリ, 2008 年 1 月.

特記事項

- [特記 1] GPS 将棋: 世界コンピュータ将棋選手権第 10 位, コンピュータ将棋協会, 2007 年 5 月.

情報システムを用いた教育研究支援

尾上 能之

1 概要

教育用計算機システムや学習管理システムを運用しサービスとして提供することで、利用者がより少ない負担で教育研究活動を行なえるよう、支援を行なっている。またシステム開発の手間を軽減するための一策として、関数プログラムの最適化に関する研究を行なっている。

2 学習管理システム CFIVE

2.1 背景

当センターでは日本ユニシス株式会社と共同で開発した学習管理システム (LMS: Learning Management System) である CFIVE(Common Factory for Inspiration and Value in Educational) を、2004 年から運用し学内の利用者にサービスとして提供している。2007 年には夏・冬学期合わせて 100 前後の講義で利用されるようになり、以前より利用者や利用機会が増えたことにつれて、システムに内在する新たな問題が表面化することもできた。このような際は CFIVE の運用に携わっている教員 2 名と職員 1 名で対処にあたるが、システムの内部の修正については主に教員が行なっているのが現状である。

2.2 内容

今季 CFIVE で起きた主な障害や不具合と、行なった対応は以下の通りである。

- 課題の二重提出に関する調査と修正

CFIVE のような Web アプリケーションでは、ユーザは Web ブラウザ経由で様々な操作を行なうが、利用者の多い時間帯などではサーバからの反応が遅くなることもある。このとき待ち切れないユーザがダブルクリックを行なったため、課題ファイルのアップロードが複数回行なわれデータベースの不整合が起きることが判明した。そこで、このような場合も正しく処理を行なえるようデータベース上に該当エントリが一意になるような unique 制約を入れる処理を行ない修正した。

- 2 バイト文字の文字数制限についての調査

現在の CFIVE の実装では、日本語全角文字の扱いが手を加えた人によって 1 文字 2 バイトを想定して処理したり、3 バイトとして処理したりバラバラで汚ない状態である。一方最近のデータベースでは、VARCHAR(n) 型を用いると n バイトではなくマルチバイト文字で n 文字を保存できるため、文字の種類を意識することなく決められた長さの文字列を保存できる。そこで文字数制限が DB 上での溢れを予防する目的のみなら、現在の多バイトを想定した処理が修正可能であることを確認し、一部修正した。全体の修正は今後行なわれる予定である。

また CFIVE では利用者からのバグ報告や改善要求をバグトラッキングシステムで管理し、随時システムの改良を行なっている。今季行なった機能拡張、バグ修正のうち主なものを以下に挙げる。

- 学生自身によるコース登録機能
これまで管理者や教員が行っていた学生のコース登録を、教員が希望する場合には学生自身でコース登録できるよう内部処理や UI を変更した。
- グループ毎に学生がレポートを見せ合う機能
- 学生の自習を目的とした、同じテストを何回でも受けられる機能
- 掲示板への投稿を RSS で通知する機能
- 授業一覧において、各授業毎の履修者の出欠を Excel などの表形式ファイルで出力する機能

これ以外にも、今後利用者数がさらに増加したときの影響も考慮し、サーバ設備の増強や稼働しているソフトウェアをより効率的に動かすための工夫等も随時行なっている。

CFIVE はオープンソースのソフトウェアとして広く利用してもらえるよう一般に公開しているが、大学で運用することを念頭に設計された本格的なシステムであることもあって、利用者の中には構築の際につまづく事も多いようである。そこで当センターで提供しているサポート窓口に来た質問に対し、適宜アドバイスを与えることによって、より多くの場所で CFIVE を活用してもらえるよう働きかけている。中には CFIVE の運用を予定している機関の例の一つとして放送大学があり、インストールなどの環境構築や運用に関して協力を行なっている。

2.3 具体的成果

[招待 1] において、CFIVE も含めた東京大学における ICT 活用の現状について講演を行なった。

図 1 は CFIVE に申請のあった講義数 (各月毎と累積) であるが、この図からもわかるように利用講義数ならびに利用者数は増えてはいるものの、全学の規模からするとまだ僅かであると言えよう。今後も新たな機能を導入し利用者の利便性を改善しつつ、安定したサービスを提供できるよう努力していきたい。

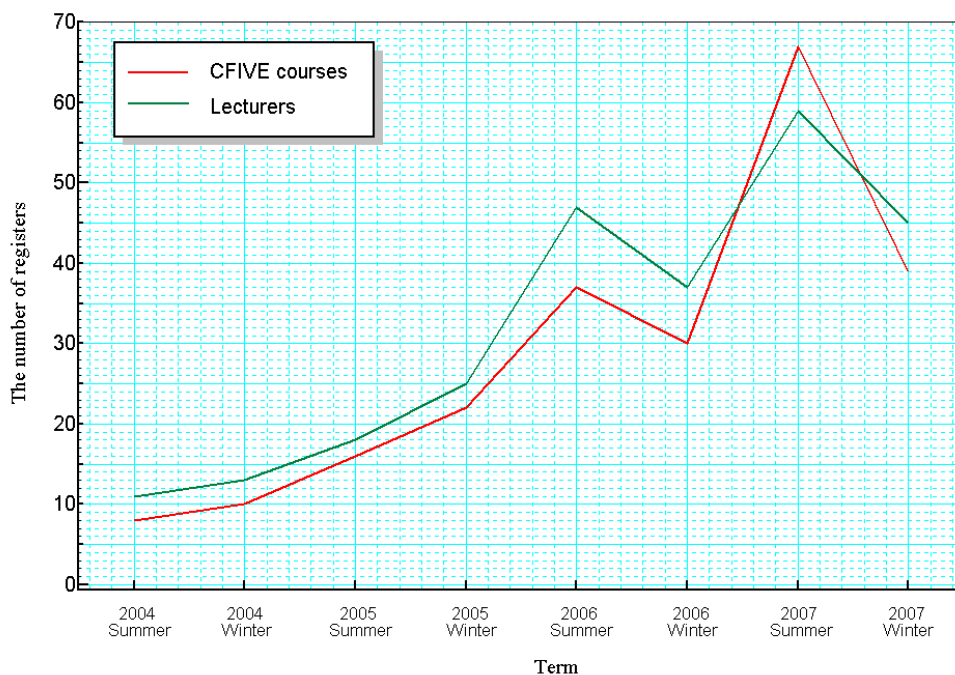


図 1: CFIVE 利用状況

3 教育用計算機システム

3.1 背景

当センターでは現在、2004年に導入した教育用計算機システム (Educational Campuswide Computing System, ECCS) の運営を行っており、教養学部における前期課程教育 (全科類で必修)、各学部 / 研究科における専門課程教育のための環境を提供しているだけでなく、電子メール等学内のすべての構成員の教育活動の支援をしている。これらのシステムを合わせると端末、サーバを合わせて 1,400 台を超える大規模分散システムとなっており、これを登録者数約 30,000 人、実利用者数約 20,000 人のユーザが利用している。また、近年のインターネット及びその上のサービスである電子メール、World Wide Web (以下 WWW と記す) 等の普及から、以前のように授業時間帯だけの利用ではなく、24 時間 365 日のサービスを提供している。

このような大規模なシステムを 10 名前後の技術職員だけで運用していくことは難しく、教員の協力が不可欠な状態である。

3.2 内容

今季行なったユーザサポートや障害への対処のうちのいくつかを以下に紹介する。

- DNS の誤設定によるメール送信障害への調査と対処
メールホスティングサービスのユーザから、メールの送信が行なえないとの報告があり対応した。我々のサービスではメールを送信する際、送信者のメールアドレスが実在するドメインであるか確認処理が行なわれるが、メールサーバの仕様として DNS を参照する際、通常行なわれる MX や A レコードではなく ANY レコードで参照していることが判明した。また該当ユーザは DNSSEC に関する設定をし TTL が不均一に設定されたドメインに所属していたため、ANY クエリでは目的の MX や A レコードが含まれないことがあることがわかった。そこで該当ドメインの DNS 管理者に設定変更依頼をするようユーザには回答することにした。
- 部局部屋への ECCS 有線 LAN 接続環境提供
センターでは、これまで部局で運営していた計算機室をセンターが用意する端末や有線・無線 LAN 接続環境に移行することで、ハードウェアや認証情報の管理など各部局の負担を減らすサービスを行なっている。この一環として、経済学部でこれまで管理していた院生部屋の DHCP 環境を、ECCS 有線接続環境に置き換える設定を行った。また計算機演習室の Windows 端末が ECCS のアカウントでログインできるように設定の協力を行った。通常 Windows では Active Directory という独自の認証を用いる必要があるが、CO-CONV というソフトを利用することで、センターで用意した LDAP サーバを参照する LDAP クライアントとすることで、これまで部局が独自に行なっていたアカウント情報の管理の手間を無くすことができた。

また現在のシステムは 2007 年度でレンタル契約が終了し、2008 年度からは新たなシステムを運用する必要があるため、次期システム調達への手続きを今年度は行なった。その中で担当した箇所を以下に挙げておく。

- 仕様策定委員会の一員として、導入説明書、仕様書・総合評価基準などの文書の策定
- 作業部会、技術審査委員会、標準環境検討 WG などの活動をサポート
- センター以外の各部局と連絡し、端末構成や必要なソフトウェアなどを協議

3.3 具体的成果

ECCS は利用者数や端末台数からみてもわかるように巨大なシステムで、このようなシステムを運用している大学は国内でも数少ない。こうした ECCS が大きなサービス停止もなく、利用者に様々なサービスを提供できていることが一番の成果と言えるであろう。

4 構成的手法を用いた関数プログラムの最適化

4.1 背景

生産的なソフトウェア開発を行なうには、

1. 信頼できるソフトウェアを開発すること、
2. 効率的なソフトウェアを開発すること、
3. すばやくソフトウェアを開発すること、

という3つの目的を考慮する必要があるであろう。しかしこれらの目的はトレードオフの関係にあり、両立させることは難しい。例えば (a) の信頼できるソフトウェアを開発するためには、ソフトウェア工学に対する組織されたアプローチ、すなわち高級言語や構成、抽象化などを使うことを必要とする。これらの技術はソフトウェアが可搬的であり拡張性がある必要がある場合にも本質的となる。しかし一般的な傾向として、高いレベルでの一般性や抽象化を用いると、そのプログラムを実行する際の効率が大きく損なわれ (b) の要件が満たされなくなることが多い。

プログラミング言語のパラダイムの一つである関数型言語は、(a) の高い信頼性と (c) のソフト作成時の手間を軽減させるための言語として、近年注目を集めている。関数型言語ではプログラム開発者に対して高いレベルでの柔軟性や信頼性を提供し、強い型付けシステムによりプログラマは正しいプログラムを書くのが容易になる。また、遅延評価、抽象データ型、型の多様性などの特徴が、効率的なソフトウェア開発の手助けとなる。しかしその反面、(b) のプログラム実行時の効率に関しては他の言語と比較して十分とはいえないのが実状である。

そこで、これらの望ましい特徴を保持したまま (b) の要件を満たすプログラムを作るには、コンパイラにおける最適化が十分に機能する必要がある。関数型プログラミングでは小さな部品となる関数を組み合わせて大きな作業を行なうことが多いが、その際に受け渡しされる本来不要な中間データ構造を、自動的に除去する融合変換のためのアルゴリズムを提案する。

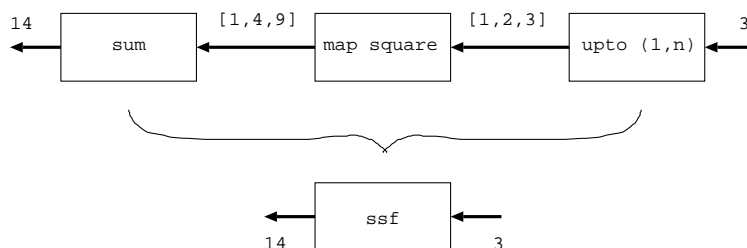


図 2: 融合変換の様子

4.2 内容

我々は融合変換(図 2)の手法を用いて、不要な中間データを生成しない効率の優れたプログラムを生成するためのアルゴリズムを提案し、それを実際に用いられている汎用の処理系上に実装した。融合変換の手法としていくつかのアプローチが提案されているが、我々は中間データとして用いられるこ

との多いリストや木などのデータ構造に着目し、このデータ構成子のある一定の条件のもとで成り立つ置換ルールに沿って他の演算子で置き換えることで融合変換を実現している。

簡単な例で言うと、関数プログラミングでは1から n までの整数の和を求める式を `sum [1..n]` と記述するが、これでは一旦長さ n のリストをデータの受け渡しだけのために生成してしまう。そこでコンパイル時に `[1..n] = 1:(2: .. (n:[]) ..)` の中に表われる `(:)` と `[]` を、`(+)` と `0` で置き換えることで中間リストを生成することなく結果を求めることができるようになる。

他の融合変換へのアプローチの中には `sum` や `map` といった各関数に対し、その関数に適用可能な書き換え規則をルールとして予め用意しておくことで、これと同様の結果を得られるものも存在する。一方我々の手法では、`sum` などの関数の定義を解析し、そこから `0` を元に各要素の間に加算演算が適用されることを自動的に導くことができる。したがって、ユーザ定義関数などの予め書き換え規則を用意していない関数間に受け渡しされる中間データも削除できるという利点がある。

4.3 具体的成果

[発表1]において発表を行ない、問題解決のためプログラムを作る上で関数型言語を用いる際の考え方や長所短所などを、実例を交えながら紹介した。

5 成果要覧

招待講演 / 招待論文

[招待1] 尾上能之: 東京大学における ICT 活用教育の現状について, ICT を活用した教育のための NIME セミナー, 東京, 2008 年 3 月.

<http://www.nime.ac.jp/seminar/kensyu.h19/080310/resume.html>

その他の発表論文

[発表1] 尾上能之: 関数「思考」プログラミングのケーススタディ, 情報処理学会 夏のプログラミング シンポジウム, 長野, 2007 年 8 月.

<http://www.atdot.net/s-prosym2007/>

教育支援システムの開発と運用

関谷 貴之

1 概要

広い意味で教育を支援するシステムとして、教育用計算機システムと学習管理システム CFIVE に関する開発と運用を行っている。またこれらの運用を支えるシステムとして知識ベースと呼ぶ一種の Knowledge Management System についても開発と運用を行っている。また、講義データベースと題して、講義シラバスを用いてカリキュラムを分析する研究を行っている。

2 教育用計算機システム

2.1 背景

高等教育機関において教育用の計算機環境を提供するに当たっては、多数の利用者に対して、各種のアプリケーションが動作する、一様なハードウェアならびにソフトウェア環境を、安価な導入費用や低い運用コストで構築する必要がある。本センターでは、2004年3月から2008年2月まで、NetBoot によって Mac OS X が動作する端末を中心とした教育用計算機環境システム (ECCS2004 という) を運用した。ECCS2004 の運用を通じて、教育用計算機環境のあり方について研究を行ってきた。

2.2 内容

2.2.1 ECCS2004 の設計

ECCS2004 の設計にあたっては、主に以下に述べる要求仕様を設定した。

多数の端末を提供

コンピュータやネットワークが情報の発信や伝達的手段として身近になった現在において、教育用の計算機環境は、教育や研究を支える道具として時間や場所を問わずに利用される。

そこで、講義での端末需要を満たすべく、数十から 200 台程度の端末を設置した教室を用意するだけでなく、図書館のような公共の場にも端末を設置する事とした。結果として、ECCS2004 全体では 1,000 台以上の端末を提供する事が要求仕様の一つとなった。

多数の利用者に基本的なインフラを提供

WWW や電子メールが教育や研究における情報発信・情報共有・情報交換の基本的な手段として普及し、教育用計算機システムがこれらの手段を提供している。そこで、ECCS2004 では端末上で WWW ブラウザやメーラが動作させる事が ECCS2004 の要求仕様の一つとなった。

各種のアプリケーションを実行可能

教養学部前期課程の情報リテラシー教育においては、OS には UNIX 系の OS を、プログラミング言語には Java を用いてきた。また、プログラミング作業は、Emacs などのエディタ上でファイルを作成して、ターミナル上でコンパイルするというものである。また、従来より教育用計算機システム

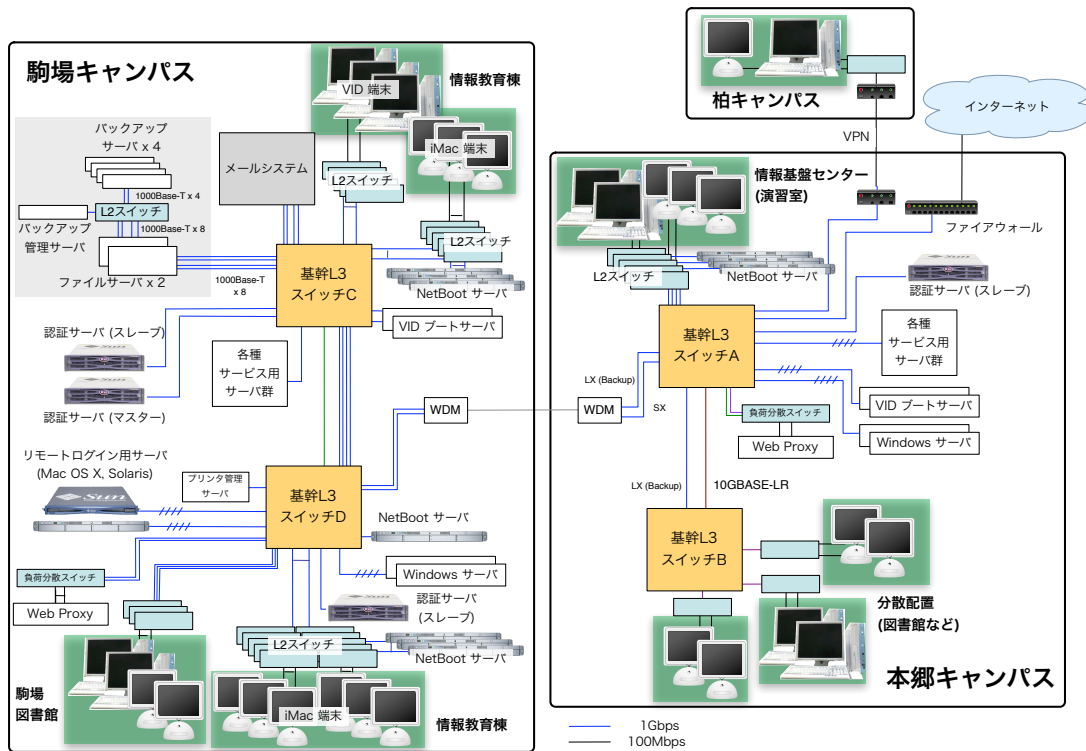


図 1: ECCS2004 のシステム構成

では、GNU ソフトウェアを利用してきた。従って、GNU ソフトウェアなどの各種オープンソースソフトウェアや Java VM が動作する事が要求仕様の一つとなった。

一方、情報リテラシー以外のいくつかの講義で利用するために、統計処理ソフトウェア SAS、表計算ソフトウェア Excel、そして CAD として AutoCAD が動作する事が求められた。

端末上のハードディスクを利用しない

1999 年から運用してきた教育用計算機システムでは、ディスクレスの端末を導入する事で、それ以前のシステムよりも端末の故障率を減らし、故障時の端末の交換作業も容易なものとなった。そこで ECCS2004 においても、ハードディスクを使わない端末を導入する事とした。

端末上で主に動作する OS は UNIX とする

本センターでは、UNIX 系の OS の管理の経験やノウハウが蓄積されており、経験の乏しい他の OS の場合に比べて、障害からの復旧も容易であると考えられる。そこで、ECCS2004 においても、UNIX 系の OS を端末の OS とする事を設計方針とした。

2.2.2 ECCS2004 の構築と運用

2.2.1 節で述べた要求仕様を踏まえて、ECCS2004 は NetBoot による端末を中心としたシステムを導入した図 1 は、ECCS2004 の大まかなシステム構成を示している。主な構成要素を以下に説明する。

端末

授業や教育・研究の情報基盤として中心的な役割を果たす iMac 端末 1,149 台と、Windows を必要とする講義用のディスクレス Windows 端末 227 台で構成される。iMac 端末は端末上のハードディスクのコネクタを取り外して動作しない状態にして、Xserve をブートサーバとする NetBoot によって起動する。

印刷

端末から印刷ジョブを送った後、プリンタのそばにある専用の端末を操作し、利用者自身がプリペ

イドカードで印刷料金を支払うシステムを導入した。更にこのシステムでは、プリンタの管理を外部業者に委託しており、プリペイドカードの料金収入に基づいて、トナーやプリンタ用紙の補給やメンテナンスを実施してもらっている。

認証

認証情報はLDAPサーバから各サーバや端末に対して提供する。マスターサーバ1台に対して、スレーブサーバが3台とWindows用のActive Directoryサーバが3台存在する。これによって、全ての端末や各種のサービスで認証情報を共有することを可能としている。

2.3 具体的成果

教育用の計算機環境は、世の中の計算機のハードウェアならびにソフトウェアの進歩に追いつくため、国内の多くの大学で4、5年を目安に定期的に更新している。そこで、本学のECCS2004の運用を通じて得られた知見を参考にして頂けるように、論文として発表した[査読付1]。

ECCS2004の最も大きな成果は、NetBootを用いたMac OS Xの端末を教育用計算機システムとして導入した事である。その後、ECCS2004と同様のNetBootシステムが他大学でも導入されるようになった。また、2008年3月から運用を開始した教育用計算機システム(ECCS2008という)においても、ECCS2004の運用ノウハウを活用して、引き続きNetBootを用いた端末を採用した。

なお、ECCS2004のiMac端末は、端末上のハードディスクを利用せずにNetBootで起動することとした。これによって、機械的に動作する部品を減らし、端末の故障率を下げる効果を期待した。その結果、iMac端末全1,149台のうち、システム導入時の初期不良で交換した端末を除くと、当初の2年間で故障のために交換したのは53台(約4.6%)であった。

一方、一般的なPCのハードウェアの故障率が近年低下しており、デスクトップPCでは最初の1年目の年平均故障率は5%とGartner社は報じている。調査対象の一般的なPCと教育用計算機システムの端末では動作環境も大きく異なるため、両者を単純に比較することはできないが、故障率を下げるという目的の実現に当たっては、ハードディスクを利用しなかった効果の有無は判断しにくい。

そこで、ECCS2008においては、端末上のハードディスクを一時的な保存領域として利用しつつ、NetBootシステムとしてOSのイメージはブートサーバから提供する形態を採用した。これによって、故障時の端末の交換など端末の運用コストを小さくしつつ、端末上のHDDをキャッシュとして活用することで、ECCS2004より端末の性能が向上することを期待している。

3 学習管理システム CFIVE

3.1 背景

情報基盤センターは、2004年に日本ユニシス・ソリューションと共同でオープンソースのLMS(Learning Management System)であるCFIVE(Common Factory for Inspiration and Value in Education)を開発し、同年4月より学内の教職員・学生向けのサービスとして、CFIVEの実行環境を提供してきた。また、サービスの利用者である教職員や学生からの要望や、本センターでの運用上の工夫などを取り込みながら、新しい機能の開発や不具合の修正を行い、現在も継続してCFIVEを改良している。

また、教育用計算機システムや教務システムとの認証情報の連携や、ICカード機能付学生証を用いた出欠管理システムなど、CFIVEを基盤としたシステムの構築も行っている。

3.2 内容

3.2.1 システムの改良

2004年4月にCFIVEをリリース後、半年から一年毎に改良を行っているが、2007年度の主な改良点は以下の通りである。

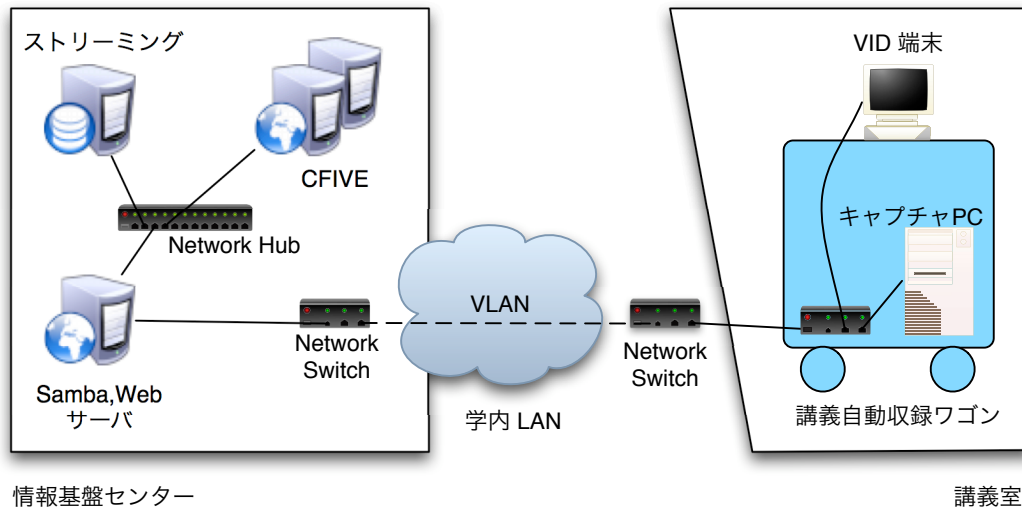


図 2: 講義自動収録システムのネットワークに関する仕組み

- 出席一覧のダウンロード
授業一覧の機能において、受講者毎に全ての授業の出欠を数えて、その結果を CSV/Excel/HTML の各形式のファイルとしてダウンロードする機能を実装した。
- 掲示板への投稿を RSS で通知
掲示板に投稿された記事を、RSS で通知する機能を実装した。
- 何回でもテストを受けられる機能
いわゆる小テストやドリルとしての利用を想定して、一定期間内に受講者が同じテストを何回でも受験できる機能を実装した。
- 学生自身によるコース登録
予め設定されたコースに、学生自身が受講登録を可能とする機能を実装した。

3.2.2 講義自動収録システムの構築 [発表 1]

講義の様態を撮影した映像と講義中に利用されたプレゼンテーション資料を合成した教材を作成し、この教材を CFIVE から視聴可能にする一連の処理を、自動的に行うシステムを構築した。なお、私はシステム全体の大まかな設計を担当した。

図 2 は講義自動収録システムの仕組みを大まかに示したものである。システムの利用者となる教員や学生は、プレゼンテーションを行うために用意された VID 端末の電源や、プレゼンテーションソフトウェアを操作するのみで、教材を作成することが出来る。VID 端末はシンクライアントであるため、通常の PC を用いる場合に比べて、セキュリティや保守性に優れている。その他、教材を作成するキャプチャ PC や各種機材を講義自動収録ワゴン教卓の中にコンパクトにまとめることで、ある程度の可搬性も実現した。

講義の終了後には、キャプチャ PC で作成された教材のファイルは、Samba, Web サーバやストリーミングサーバに転送され、教材のメタ情報が CFIVE に登録される。

なお、講義自動収録システムは、情報基盤センターのセミナーで試験的に運用することでその機能や使い易さを検証した。

3.3 具体的成果

CFIVE は 2004 年 4 月に学内向けにサービスを開始してから、これを利用する講義や学生数は着実に増加している。3.2.1 節で述べた 2007 年度に改良或いは実装した機能については、2008 年 4 月以降の講義で活用される予定である。

3.2.2 節で述べた講義自動収録システムについては、以下の点を達成し、セミナーや研究発表会の模様の収録には活用できることが明らかになった。

- 一連の作業を自動化を実現し、利用者は簡単な操作で収録が行える。
- 収録後の教材を CFIVE 経由ですぐに参照できる。

一方、黒板の板書や教卓から離れる教員を撮影するには、カメラの位置や撮影方法を検討する必要があることも分かった。

4 知識ベース [発表 2]

4.1 背景

情報基盤センター情報メディア教育部門では、本郷と駒場の二つのキャンパスに分かれて勤務する教職員が、協力して教育用の計算機システムやメールサーバ等の運用管理に当たっており、関係者の連絡手段には専ら電子メールを用いている。

電子メールはファイル交換やスケジュール管理等、単なる連絡手段以外の目的で広く利用されているとの調査があるが、本部門においても様々な目的で電子メールを用いている。例えば、システムの運用上の問題提議、それに対する意見、問題への対処作業の報告等は電子メールで行うことが多い。利用者がシステムの使い方を電子メールで問い合わせることも少なくない。またサーバのログやディスクの使用状況等は自動的に電子メールで送られるし、システムの状態を監視するツールは障害を検知するとそれを電子メールで通知する。

このように本部門の業務においては、絶えず届けられる電子メールに目を通しているが、過去に受け取ったメールを調べることもある。例えば「〇〇サーバへの通信ができない」障害を報せるメールを受け取った場合、過去のメールから同様の障害の発生の有無や対処方法を調べて参考にする事がある。

4.2 内容

そこで、2002 年頃より、電子メールの一連のやり取りを事例として記録し、類似の問題が起きた際には過去の事例を参照することで、関係者が知識の共有を行える、知識ベースと呼ぶ電子メールを基盤とした Web アプリケーションを構築して運用している。

4.2.1 知識ベースの設計

知識ベースの設計に当たっては、電子メール、ミーティングの議題、知識、3 種類の情報に関して以下に挙げる機能を実現する。なお、図 3 は、知識ベースにおける情報の流れの概念図である。

1. メーリングリストに送られた“Mail/電子メール”を Web でアクセス可能な形式で保存する。
2. 保存した Mail のスレッドをミーティングの議題の候補として、一個の独立したデータ“Topic/議題”として蓄積可能とする。
3. Topic をシステムの運用・管理等の各種業務に関する事例とすべく、適宜編集可能とする。
4. Topic には未処理・作業中・終了等といった状態を設定可能とする。
5. ミーティングの議題を Topic から選択することで作成可能とする。

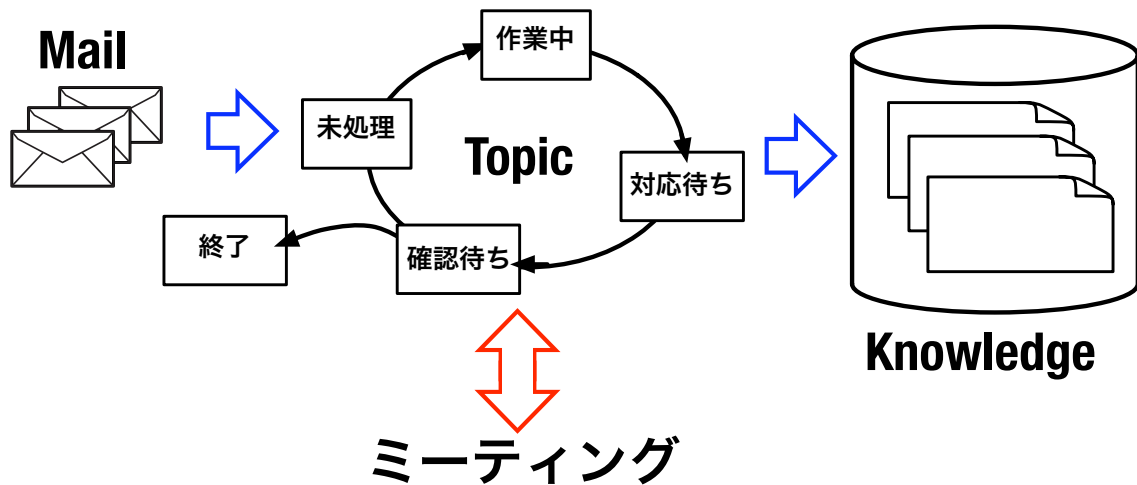


図 3: 情報の流れ

6. 知識や情報として部門内関係者で共有すべき Topic を、独立したデータ “Knowledge/知識” とし蓄積可能とする。

本部門におけるあらゆる情報は、電子メールで関係者間に事前に知られる或いは知らされることが多いことから、このように電子メールを基盤として、電子メールから Topic、そして Topic から Knowledge として変換できる機能を実現することとした。日常の業務においては、タスク管理を目的として Topic の状態を設定すると共に、その内容を適宜整理する。Topic を検索して参照することで、類似の問題が起きた際には過去の事例として参考にできる。

4.2.2 システム構成

知識ベースは、Apache が動作する Linux 上で、Perl による CGI として開発した。データベースとして PostgreSQL を、検索システムとして Hyper Estraier を用いた。知識ベース内部の主なデータである Mail, Topic, Knowledge とそれに関わる “Staff/教職員” と “Schedule/予定” の関係を図 4 に示す。Mail と Topic、Topic と Knowledge、Topic と Schedule、さらに Topic 同士は互いに多対多に関連づけることができる。これによって、ある Topic に関わる電子メールや Knowledge として登録された知識を容易に参照できる。

4.3 具体的成果

当部門には現在約 20 名の教職員が所属しているが、そのうち 12 名がほぼ毎日知識ベースを利用している。また、システムの設定方法や業務上の諸手続き等に関する Knowledge は毎月 7,8 件程度新たに登録されている。これらのことから本部門で知識ベースが積極的に活用されていると考えている。

5 講義データベース

5.1 背景

教材や論文のレポジトリ、オンラインシラバスという形で電子化されたコンテンツが大量に存在する現在、コンテンツの検索や類似度を算出する研究は広く行われている。これに対して、コンテンツが対象とする分野の範囲や、その分野の中でのコンテンツの位置づけを全体的に把握する研究は比較的少ないと考えられる。そこで、コンテンツの情報から、対象分野の用語や用語間の関係からなるドメイン・オントロジーを構築し、コンテンツの検索・俯瞰に活用するための研究を行っている。

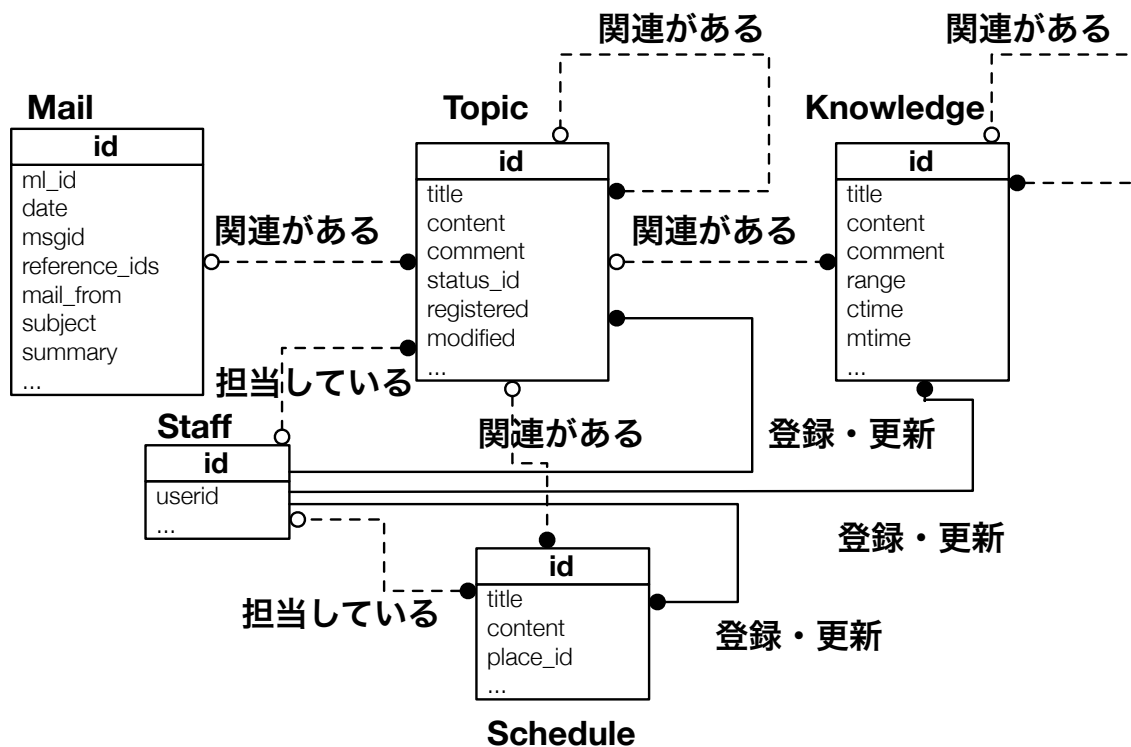


図 4: データ構造 (一部)

5.2 内容

2007 年度は、シラバスの文の単位で用語の集合を取り出すことを試みた。

5.2.1 用語集合の抽出

シラバスの中には、講義の内容に関わる記述だけでなく、授業の形態や成績の付け方等の内容には関わらない記述も存在することから、それらを区別無く取り扱うのではなく、用語集合を形成する互いに関連の有る単語は同一の文に現われると仮定した。

まず、文書(或いは教材) d_j の任意の一つの文に含まれる異なる単語 w_1, w_2, \dots, w_n を、任意に組み合わせた集合 $\{w_a, w_b, \dots\}$ を用語集合 \mathfrak{k} と呼ぶ。文書 d は、当該文書に含まれる幾つかの代表的な用語集合 \mathfrak{k} によって特徴付けられるものとする。

代表的な用語集合を抽出するにあたり、まず用語集合間の上位・下位関係を定義する。用語集合 $\mathfrak{k}_x = \{w_a, w_b, \dots\}$, $\mathfrak{k}_y = \{w_p, w_q, \dots\}$ について、 $\mathfrak{k}_x \supset \mathfrak{k}_y \wedge |\mathfrak{k}_x| = |\mathfrak{k}_y| + 1$ が成り立つ時、 $\mathfrak{k}_x \supset \mathfrak{k}_y$ と表記し、 \mathfrak{k}_y は \mathfrak{k}_x の上位の用語集合と呼ぶ。

次に用語集合 \mathfrak{k} における tf-idf [?] に相当する値 tsf-idf を式 1 で定義する。ここで、 $\mathfrak{k}fn(\mathfrak{k}, d)$ は文書 d 中で用語集合 \mathfrak{k} を含む文の数、 $N_S(d)$ は文書 d 中の文の数、 $N_D(\mathfrak{k})$ は用語集合 \mathfrak{k} を含む文書の数、 N_D は文書の数を表す。

$$\begin{aligned}
 \mathfrak{k}sf(\mathfrak{k}, d) &= \frac{\mathfrak{k}fn(\mathfrak{k}, d)}{N_S(d)} \\
 idf(\mathfrak{k}) &= \log \frac{N_D}{N_D(\mathfrak{k}) + 1} \\
 tsf-idf(\mathfrak{k}, d) &= \mathfrak{k}sf(\mathfrak{k}, d) \cdot idf(\mathfrak{k}) \tag{1}
 \end{aligned}$$

代表的な用語集合としては、式 1 の $tsf-idf(\mathfrak{k}, d)$ の値が大きな用語集合を選択する。しかし、ある文書において同じ単語を含む複数の用語集合を比較した場合、 $\mathfrak{k}sf(\mathfrak{k}, d)$ にはあまり違いが無いが、単

語数が多い用語集合の方が $N_D(\mathfrak{t})$ が小さくなって $\text{tsf-idf}(\mathfrak{t}, d)$ の値が大きくなる傾向がある。つまり、 $\text{tsf-idf}(\mathfrak{t}, d)$ の値のみに基づいて代表的な用語集合を選択すると、単語数が多くかつ似た単語が含まれる用語集合が複数現われる。

そこで代表的な用語集合 \mathfrak{t}_x は式2を満たすものとする。前述のように \mathfrak{t}_x の tsf-idf は、 \mathfrak{t}_x と同一の単語を一部含む用語集合 \mathfrak{t}_y の tsf-idf よりも大きくなることから、 \mathfrak{t}_y の tsf-idf の δ 倍である場合のみ、 \mathfrak{t}_x を代表的な用語集合とすることを式2は意図する。式 δ が大きいほど、より少ない単語で構成される用語集合が代表的な用語集合となる。実際のシラバスを用いた予備的な実験より、 $1 < \delta < 1.5$ の間で抽出される用語集合の構成が大きく変わることが分かっている。

$$\forall \mathfrak{t}_y (\mathfrak{t}_x \supset \mathfrak{t}_y), \frac{\text{tsf-idf}(\mathfrak{t}_x, d)}{\text{tsf-idf}(\mathfrak{t}_y, d)} > \delta \tag{2}$$

5.2.2 本学教養学部シラバスの分析

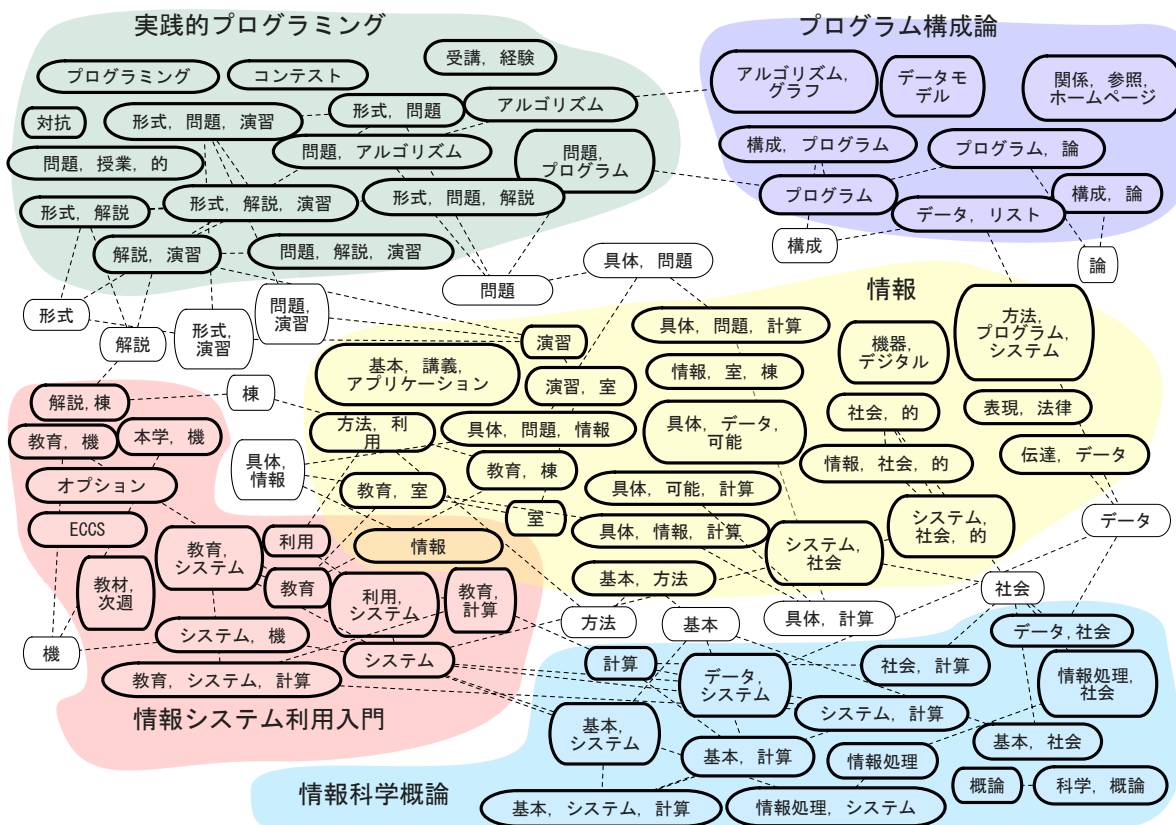


図5: シラバス構造 ([発表3] より)

上記の分析手法を、本学教養学部1,2年生向けの825の講義のシラバスに対して適用した。825個のシラバスから抽出された単語は約10,000個で、用語集合は約7,000,000個である。また、多くのシラバスに含まれる名詞の数は300個以下で、平均で130個程度の単語が抽出された。

単語の抽出には??節と同様にMeCabを用いた。また、個々の講義のシラバスから抽出する用語集合の数 N は、シラバスに含まれる単語の数 N_w に応じて10個前後となるように、 $N = \lfloor 6 \log(N_w/10 + 1) \rfloor$ として実験的に定めた。

図5は、表1に示す情報科学関連の5つの講義のシラバスから抽出した用語集合をノード、上位下位関係をエッジとしたグラフである。同じ講義のシラバスから抽出された用語集合を図形で囲み、講義同士の関係構造を表現している。講義「情報」と「情報システム利用入門」が用語集合 {情報} を

表 1: 情報科学に関する講義

タイトルと内容 (抜粋、一部修正)
実践的プログラミング: ACM-ICPC などに出題された問題を題材として、アルゴリズムを考えてプログラムを作る能力を実践的に養うことを目指す。
プログラム構成論: プログラミングの基礎となるデータモデルについて学ぶ。
情報: 多様な場と状況で情報処理システムと関わる全ての人々に必要となる、情報の本質とその人間のおよび社会的な側面を理解する能力を身に付ける。
情報システム利用入門: 本学の教育計算機システムを利用して、一般的な情報システムの利用方法を初歩から学ぶ。
情報科学概論: 現代社会の基盤構造の一つになりつつある情報処理システムおよびそれが扱うデータについての、系統的な枠組の理解を目的とする。

共有したり、「プログラム構成論」と「実践的プログラミング」で{アルゴリズム}の上位下位関係等で、相互に関連のある講義であることが読み取れる。

一方、「情報」と「プログラム構成論」が実際には強く関連があるにも関わらず、図5では両講義の用語集合を繋ぐエッジが比較的少なく見える。これは「情報」のシラバスが毎回の授業内容の詳細を記述しているのに対して、「プログラム構成論」のシラバスが全体概要を記述するのみで、両者の記述の粒度が異なっており、抽出された用語集合に違いがあったためである。「プログラム構成論」でも詳細な内容を記述するなど、シラバスを見直すべきであると解釈できる。

5.2.3 可視化ツール

図5は分析の結果得られたデータを、グラフ描画ツール graphviz の形式に変換した上で、市販の描画用のツールで色や配置を手作業で調整したものである。それに対して、カリキュラム設計者が比較的容易に分析結果を把握できるように可視化ツールを作成した。図6にツールのスクリーンハードコピーを示す。

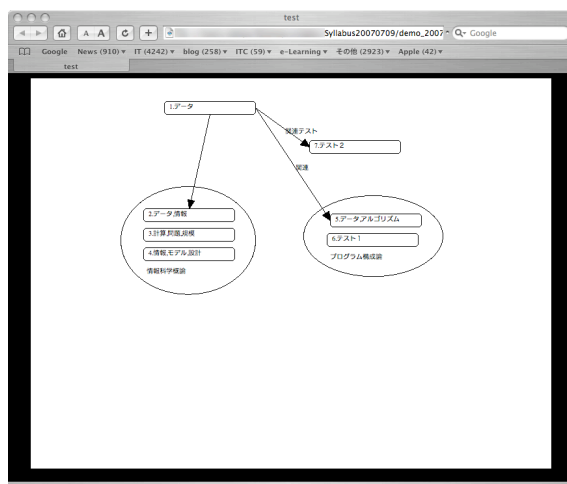


図 6: シラバス構造

5.3 具体的成果

対象としたシラバスの集合に対して、ある程度意味のありそうな専門用語の集合を抽出し、カリキュラムの構造を把握することができたと考えている。

6 成果要覧

査読付論文リスト

[査読付 1] 関谷 貴之, 安東 孝二, 尾上 能之, 田中 哲朗, 山口 和紀: NetBoot による端末を用いた教育用計算機システムの開発と評価, 情報処理学会論文誌 (ジャーナル), Vol.48, No.4, pp.1651-1664, 2007 年 4 月

その他の発表論文リスト

[発表 1] 佐藤 孝明, 前田 光教, 小川 大典, 関谷 貴之: 講義自動収録システムの構築, 平成 19 年度情報教育研究集会講演論文集, pp.300-303, 大阪大学, 2007 年 11 月

[発表 2] 関谷 貴之: 電子メールを基盤とした知識ベースによる業務支援, 平成 19 年度情報教育研究集会講演論文集, pp.340-343, 大阪大学, 2007 年 11 月

[発表 3] 関谷 貴之, 山口和紀: 講義同士の関係構造分析のためのシラバス処理方法の一検討, 情報処理学会第 70 回全国大会講演論文集 (分冊 4), pp.495-496, 筑波大学, 2008 年 3 月

web 閲覧支援とデバッグ支援の研究

丸山 一貴

1 web 閲覧支援の研究

1.1 背景

web の情報が爆発的に増加するに従い、情報を入手する方法に多様化が必要となった。従来は気に入ったポータルページを定期的に確認したり、調べたい事柄のキーワードを検索エンジンに入力するといった方法が一般的であった。これらは全てユーザの能動的なアクションに基づいており、見たい情報が予め決まっている場合の行動である。我々はこれを「目的的閲覧」と呼ぶ。

一方で、明確な目的を持たず、テレビのチャンネルをザッピングするように web を閲覧して、自分にとって興味のある情報を見つけるという行動がある。これを「発見的閲覧」と呼ぶ。ここで重要なことは、ユーザはそれを見るまで、それに興味があるかどうかは分からないということである。例えば、既に興味を持っている分野の情報であっても、そのユーザにとって既知の情報であれば価値がない。逆に興味がない分野の情報であっても、時事ニュースとの関連や意外性、あるいは潜在的な興味が喚起されることにより、価値のある情報となる。

こうした観点から、発見的閲覧を支援する仕組みとして、web 閲覧履歴に基づく情報推薦と、これに適したユーザインタフェースの研究を行っている（電気通信大学 電気通信学部 情報通信工学科 寺田研究室と共同研究）。

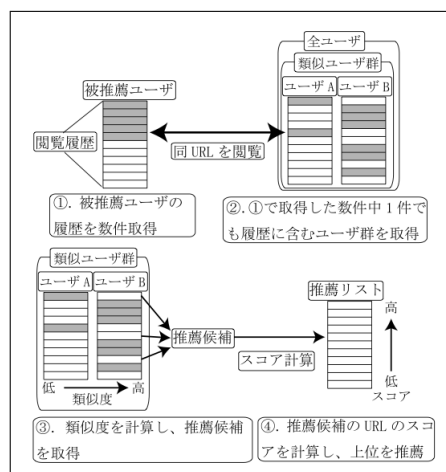


図 1: 協調フィルタリング

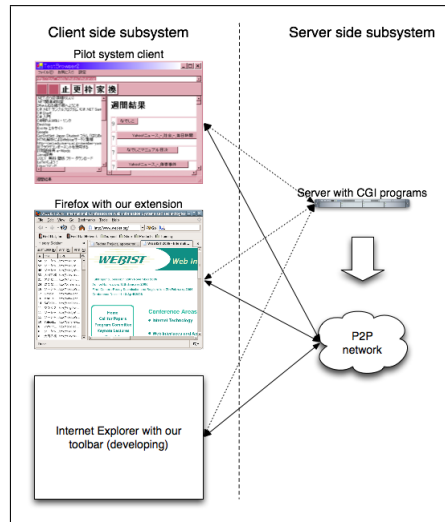


図 2: P2P 構成への移行

1.2 内容

1.2.1 web 閲覧履歴に基づく情報推薦

エンドユーザの web 閲覧履歴を収集し、協調フィルタリングを用いたスコアリングを行っている (図 1)。e コマースにおける顧客への商品推薦ではポピュラーな商品の推薦も意味を持ち、消耗品であればユーザにとっても有益な推薦たり得る。これに対して web 情報推薦においては、Google のトップページなど非常にポピュラーなものを推薦しても価値がなく、見苦しくなるため避けなければならない。そこで、多くのユーザが閲覧する URL についてはそのスコアが低くなるような評価関数を設計している。

閲覧情報の収集は、エンドユーザのウェブブラウザにアドオンを導入する方式で開発したが、多人数の閲覧情報を用いて推薦エンジンをチューニングするためにスニッファの情報を利用している。この場合はユーザが能動的にクリックした URL だけでなく、それに付随して取得される多数の URL も補足されてしまうため、スクリーニングに関する手法の研究を進めている。

現在は閲覧情報の収集を集中サーバで実施しているが、P2P ベースの分散システム化を検討している (図 2)。階層化したノード構成により、閲覧情報の流通とクラスタリングを両立する計画である。

1.2.2 発見的閲覧のためのユーザインタフェース

従来の web 閲覧は目的であり、ウェブブラウザもこれに適したものとして開発されてきたが、発見的閲覧にはより適切なブラウザが必要となる。発見的閲覧には (1) 受動的に閲覧可能で、(2) 興味を持った場合には容易に詳細な情報を見ることができ、(3) そうしたユーザのアクションが暗黙のレーティングとして推薦システムにフィードバックできるブラウザが適切である。こうした要件を満たすブラウザの開発を実施した。

1.3 具体的成果

スニッファデータのスクリーニングについて有効な手法を確立した (2008 年度に对外発表を予定)。発見的閲覧のためのユーザインタフェースのプロトタイプを作成した (2008 年度にユーザ評価を通じて改良し、对外発表を予定)。

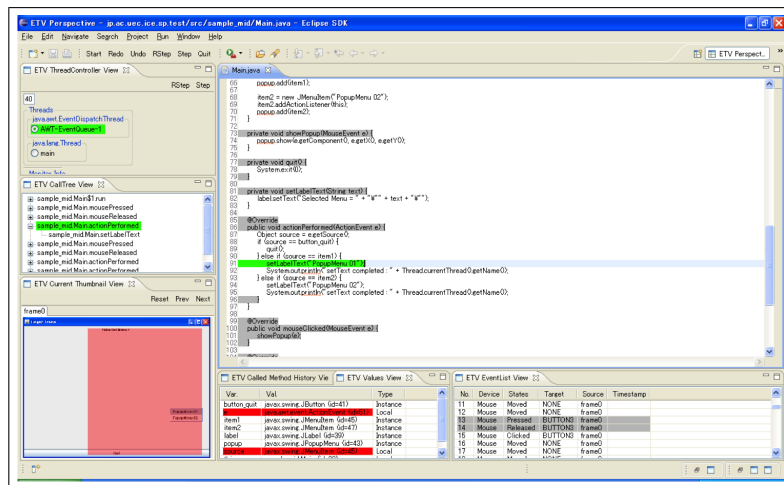


図 3: GUI 操作記録・再生型デバッグシステム

2 デバッグ支援の研究

2.1 背景

デバッグは職人芸の側面があり、また「間違いを直す」という作業の性質上、その内容が残らず場当たり的になりがちである。こうした問題を解決するため、instrumentation によってデバッガが解釈するためのコードを実行プログラムに挿入し、再現性のあるプログラムを対象とした擬似的な逆実行を C と Java で実現した。Java ではマルチスレッドの実行順序を記録・再生する手法を確立した。

一方、デバッガでプログラムの実行制御を行うのではなく、プログラムを予め実行して実行パスや変数の内容をトレースとして取得し、実行を再生する手法がある。しかし GUI を持つプログラムではトレース取得のオーバーヘッドが大きな問題となり、この手法の適用が難しかった。そこで、Java マルチスレッドの記録・再生手法を応用してこの問題に取り組んだ（電気通信大学 電気通信学部 情報通信工学科 寺田研究室と共同研究）。

2.2 内容

トレース型システムを GUI プログラムに適用できない本質的な問題点は、トレース取得時に GUI 操作をユーザが行うという前提と、トレース取得のオーバーヘッドが大きいためユーザがその場で GUI 操作するのが困難である、ということにある。そこで、(1) ユーザの GUI 操作を記録する実行パスと、(2) 記録した GUI イベントを元にトレースを取得するパスとを分離して、2パス実行方式で実現した。GUI の表示領域をキーとしてブレークポイント相当の動作を行う機能や、ステートメントが実行された際に変更が起こった表示領域を図示する機能も実装されている（図 3）。

2.3 具体的成果

共著で論文を執筆し、共著者が口頭発表を行った。

3 成果要覧

発表論文

- 柏村俊太郎, 丸山一貴, 寺田実: “Java プログラムを対象とする GUI 操作記録・再生型デバッグシステム”, 情報処理学会 2007 年度第 4 回プログラミング研究会, 2008 年 1 月。

図書館電子化研究部門

中川 裕志

機械学習とそのテキスト処理への応用

二宮 崇

自然言語構文解析とソーシャル辞書

吉田 稔

検索支援のためのテキストマイニング技術

— 同義語・接続語の高速検索とWeb 文書レイアウト解析 —

清田 陽司

図書館情報資源とWeb 情報資源の融合

～Web 2.0 時代の電子図書館の実現に向けて～

清水 伸幸

知識発見と機械翻訳の自然言語処理要素技術

— 同義語獲得、係り受け解析、機械学習テンプレートライブラリ —

王 玉馨

Web とテキスト情報処理についての研究

鍋島 彰崇

レポート文書の文のカテゴリー分け

機械学習とそのテキスト処理への応用

中川裕志

1 概要

2007 年度は、機械学習をテキスト処理に適用するいくつかのテーマについて研究を行った。Web 上で入手できるテキストが膨大な量になり、また組織内テキストの効率的利用も重視されつつある現在においては機械学習を利用した効率的なテキスト処理が必須であるという背景の下、このような方向のテーマで研究を行っている。また、ブログの普及は、商品などの評判分析への利用において著者の属性を利用したいという要求があるので、そのための方策についても研究した。またスパムブログの問題も顕在化しており、機械的な対処が必要である。具体的には、以下に示すテーマの研究を行った。

- 多重トピック文書の確率的生成モデル[査読付 1] [査読付 2] [査読付 5] [発表 1] [発表 2] [特記 1]
- ブログの著者属性の推定[査読付 6] [発表 3]
- スпамブログの分析[発表 8] [特記 3]
- Web のサーチエンジンで人名検索を行った結果を異なる人物毎にクラスタリングする名寄システム[発表 7]
- 特定分野コーパスからの同義語抽出アルゴリズムの開発[査読付 10] [発表 4] [発表 6] [特記 2]
- Suffix Array を利用した高速用例検索システム:[査読付 9] これは、吉田助教が UT-Kiwi として実装し、全学に試験的サービスを開始している。
- その他のテキスト処理アルゴリズムの開発[査読付 3] [査読付 4] [査読付 7] [査読付 8] [発表 5]

紙数の関係から、以下では、多重トピック文書の確率的生成モデルとブログの著者属性推定の研究について詳述する。

2 階層ベイズモデルによる多重トピック文書の確率的生成モデル

2.1 背景

近年の文書集合は、wikipedia や folksonomy に代表されるように、文書を、単一のトピックに分類するのではなく、重複を許して複数のトピックに分類する多重トピック分類がなされている。多重トピック文書の確率的生成モデルとは、多重トピック文書が生成される過程を確率モデルによりモデル化したものである。このモデル化によって、多重トピック文書の持つ固有の性質を抽出することができ、テキストマイニングへの幅広い応用が考えられる。例えば、文書の多重トピック分類や文書中の多重トピック性を持つキーワード抽出などが考えられる。多重トピック文書の確率的生成モデルは、潜在トピックを仮定するモデルと顕在トピックを仮定するモデルに分類される。潜在トピックモデルは教師なし

学習であり、難しいテーマである。クラスタ数を指定する LDA や、クラスタ数を指定しないノンパラメトリックな HDP(階層型ディリクレ過程)が提案されている。一方、ここで対象にするのは、トピックが予め与えられた顕在トピックモデルであり、上田によって提案された Parametric Mixture Model(PMM)[1]¹がある。PMM では、各トピックは単語を多項分布で生成するが、トピックは等確率で混合する多重トピック文書における単語の生成確率をモデル化する。この理由は、K 個のトピックから作られるすべての多重トピックの組み合わせは $K^2 - 1$ 通りあり、そのすべてに対してモデルパラメータを扱うのは現実的ではないためである。しかし、図 1 に示すように個々の多重トピック文書は、各トピックに対して偏った混合比率を持つことが一般的である。よって、我々は、このトピック毎の分布の偏りすなわち混合比を教師なし学習する方法を提案し、評価実験を行った。以下では、まず確率モデルと学習アルゴリズムについて説明し、次に評価実験結果について報告する。

2.2 内容

提案した確率モデルと機械学習アルゴリズムについて説明する。

各トピックは添え字 i で、各単語 w は添え字 v で区別する。単語 w_v トピック i の各単語の生成確率 θ_i トピックの混合比 π_i が与えられたもとで文書 w が成される確率は次式で与えられる。ただし、 α は事前分布であるディリクレ分布のパラメータである。

$$P(w | y, \alpha, \theta) = \int p(\pi | y, \alpha) \prod_{v=1}^V (\varphi(w_v, y, \theta, \pi))^{\pi_v} d\pi$$

$$P(\pi | y, \alpha) = \frac{\Gamma(\sum_i \alpha_i)}{\prod_i \Gamma(\alpha_i)} \prod_i \pi_i^{\alpha_i - 1} \text{ (ディリクレ分布) } \quad \varphi(w_v, y, \theta, \pi) = \sum_i \pi_i \theta_{iv}$$

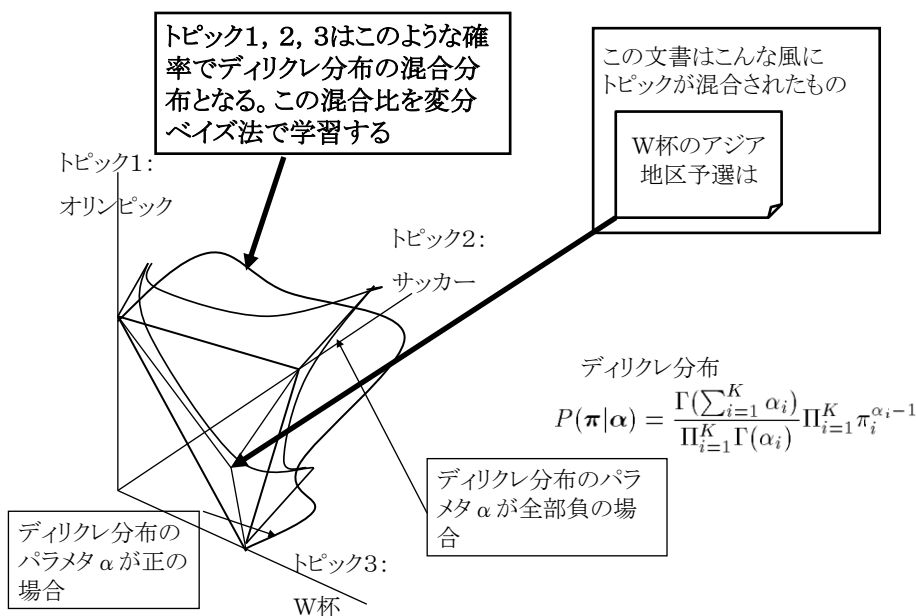


図 1 複数のトピック混合のモデル

この確率分布をモデルとして、コーパス教師データから混合比を推定する。そのためにまず上の式をベイズの定理で書き換え、教師データから得られる likelihood(多項分布)とディリクレ分布で

¹[1] Naonori Ueda and Kazumi Saito :Single-shot Detection of Multiple Categories of Text using Parametric Mixture Models, ACM SIGKDD 2002、626-631、2002

ある事前分布に分解する。(注:ディリクレ分布は、多項分布の自然共役分布なので事前分布として採用する。)ここで求めるのは $P(\pi | w, y, \alpha, \theta)$ だが、直接計算するためには確率の積分計算は必要で現実的でない。そこで、変分ベイズ法を用いることにする。変分ベイズ法では、求める分布 $P(\pi | w, y, \alpha, \theta)$ に対して潜在変数 z を導入した近似分布 $Q(\pi | \gamma)Q(z | \phi)$ を想定し、これらの KL-divergence: $KL(Q, P)$ を最小化するような Q を求める更新式を導き、繰り返し計算によって収束したものを推定分布とする。ただし、 n を単語を表す添え字として、近似分布は次式である。

$$Q(\pi_i | \gamma_i) = \frac{\Gamma(\sum_i \gamma_i)}{\prod_i \Gamma(\gamma_i)} \prod_i \pi_i^{\gamma_i - 1} \text{ (ディリクレ分布)} \quad Q(z | \phi) = \prod_n \prod_i (\phi_{ni})^{\delta_{ni}} \text{ (多項分布)}$$

繰り返しの計算の更新式は次式のようにあり、計算は高速に行える。

$$\gamma_i^{new} = \alpha_i + \sum_n \phi_{ni}^{old}, \quad \phi_{ni}^{new} = \theta_{i,wn} \exp(\Psi(\gamma_i^{new}) - \Psi(\gamma_j^{new})) / C$$

ただし、 C は正規化定数、 j は i と同じ範囲を動く。また、 Ψ はダイガンマ関数である。

この計算結果の π_i がトピック i の混合比の推定値である。

2.3 具体的成果

ここでは、評価実験の結果について説明する。

評価実験には生物医学分野の 5000 本の英文アブストラクトからなる MEDLINE コーパスを用いた。また、多重トピックとしては、1 文書に複数個付加されている MeSHTerm を用いた。なお、トピックとして採用したのは MeSHTerm の語彙のうち出現頻度 100 以上 1000 未満の中頻度の語彙である。それ以上高頻度の語彙は一般的すぎて意味のある分類とはいえず、それ以下の頻度の語彙は一般的に通用する有意義なトピックとは考えられないからである。その結果、トピック数は 88、語彙数は 46、075 語。全文書に対する多重トピック文書の割合は、0.4478 であった。

上田らの PMM[1] とここで提案する PDMM を F 値で比較した結果を図 2 に示す。この実験では、横軸に評価した文書集合中における多重トピック文書の割合を変化させて F 値の様子を調べた。多重トピック文書の割合が大きいとほど良い F 値であり、いずれの割合でも PMM よりも高い F 値であった。

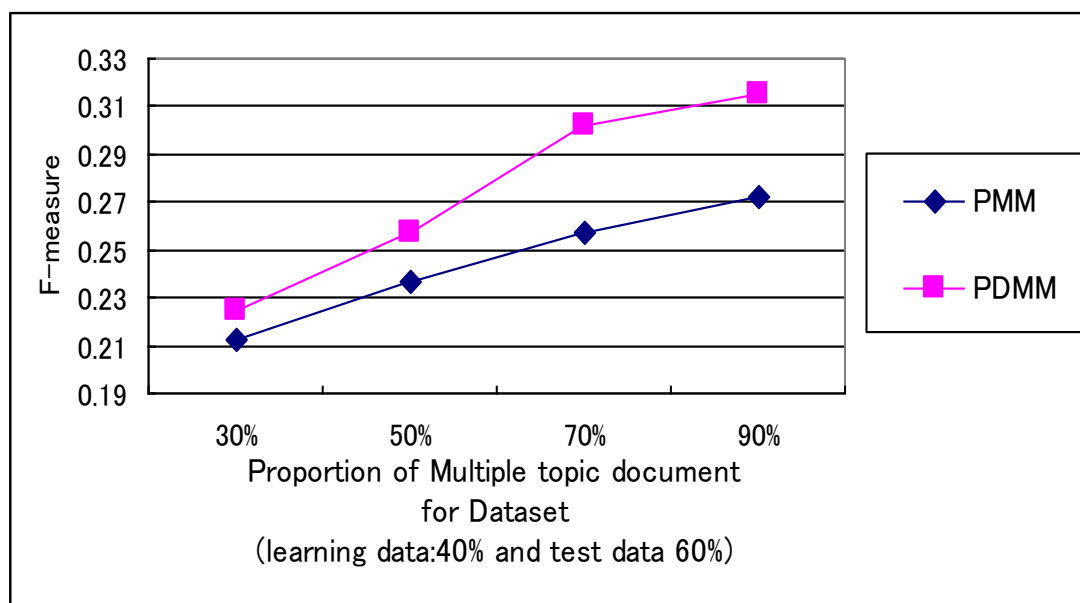


図 2 PDMM の評価 (MEDLIBNE での F 値)

次に、実験結果として得られた多重トピックの混合比と、そのトピック分配比になる文書クラスタにおいてエントロピーの高い重要語の例を示す。ここに示すクラスタのトピック配分比は以下の () 内の通りである。[Female(0.499)]、[Male(0.460)]、[Biological Markers(0.041)]

表 1 多重トピッククラスタにおける重要語 () 内は TF-IDF で並べた場合の順位

Ranking	Top10	Ranking	Bottom10
1(37)	biomarkers	67(69)	indicate
2(19)	Fusarium	68(57)	problem
3(20)	non-Gaussian	69(45)	use
4(21)	Stachybotrys	70(75)	%
5(7)	chrysogenum	71(59)	correlate
6(22)	Cladosporium	72(17)	population
7(3)	mould	73(15)	healthy
8(35)	Aspergillus	74(33)	response
9(23)	dampness	75(56)	man
10(24)	1SD	76(64)	woman

3 ブログ著者属性推定

3.1 背景

ブログは、1997 年末に誕生してからの 10 年で、手軽な情報発信の方法として急速に広まった。特に日本では、元々オンライン日記を好む国民性と相まって急速に普及し、最近では新しく生まれる記事の言語は日本語が最も多い。現在、日本語によるブログは 200 万サイトから 400 万サイト程度あるといわれ、日々 50 万記事以上が新しく投稿されている。ブログに投稿される記事はユーザーの率直な意見を反映しており、また些細な事柄も含まれているといわれている。アンケートを採るなどの人手で行う流行調査と比べ、ブログのデータの分析による調査は低コストかつ短期間で実施できるため、ブログは流行調査を行うための情報元として有望視されている。ブログの記事は電子的な形で提供されているため、これらの調査は完全に機械化することができそうに思える。しかし、現状ではまだ機械では困難として、調査したい製品名を含む記事をリストアップし人手で読み解く方法が主である。マーケティング等で調査をする際、その調査対象者の属性が重要となる。具体的には、性別、年齢、行動圏、等々である。もしこれらの属性を機械で適切に推定できれば、ブログを用いた自動調査の有用性は飛躍的に向上することが期待される。このための研究は既にいくつか行われているが、個々の属性の特徴に依存しない汎用的なものはまだない。特定の属性にとらわれない一般的なアルゴリズムを提案できれば、調査ごとに必要となる属性を機械的に推定することが可能になり、ブログを用いた調査の適用可能範囲を大きく広げられると考えられる。このため、本研究ではブログの記事を用いその情報源であるブログ著者の属性を推定するための汎用的なアルゴリズムを提案し、実際に性別、年齢層、居住地域の 3 種類の異なった著者属性の推定が実用的な精度で行えることを確認した。

3.2 内容

提案する手法は、ブログデータをマーケティング調査等に使用する際に、調査者が任意の属性を切り口にした調査を行うことを可能にする手法の構築を目指して構成した。実用性を考えれば、属性ごとに異なったアプローチをとることは難しい。また、現実的な教師データの集め方としてはサービスプロバイダの 1 つが内部で持っている情報を利用するといったことが考えられるが、この場合に利用できる教師データは最低でも数万サイト分の著者属性を利用できると思われる。これらのことを考慮し、下記の 4 点を要件とした。

- (1) 数万サイト規模の教師データによる学習が可能なこと
- (2) 多クラスの属性にも適用できること

- (3) 推定が高速であること
- (4) 各属性固有の特性を使用せず、ブログデータと教師ラベルのみから属性を学習できること

まず、各ブログサイトの記事中に出現する単語を bag-of-words モデルを用いてモデル化する。その後、高速な多クラス文書分類アルゴリズムである CNB(Complementary Naïve Bayse 法) を用いて属性値を推定する。ナイーブベイズ法と異なり、CNBでは、各クラスに「含まれない」確率を学習する方法 Complement Naive Bayes(CNB) を提案している。多くの場合、それぞれのカテゴリーについてその属性値をもつサイト数のばらつきは、その属性値を「もたない」サイト数のばらつきよりも比率が小さい。このため、補集合を使うことで教師データ数の偏りを抑えられることが多いとされている。

分類器の学習は、ラベル付きデータを用いる。また、出現する全ての単語を扱うと学習時・推定時共に時間がかかってしまうので、著者属性の推定に有用な語を χ^2 に基づいて選択し素性として利用することとした。

3.3 具体的成果

まず、実験に用いたデータについて紹介する。ブログの URL としては、2006 年 4 月頃から 2007 年 4 月頃までの約 1 年間にクロールした記事の中からブログサイトの URL を抽出したものをを用いた。これらのブログのうち、手書きのルールによってプロフィール欄の文字列を取得できたサイトが 13 万 9944 サイトあった(これらのサイトに含まれる記事は 1365 万 4635 件であった)。得られたプロフィール文字列の中には「性別:非公開」といった、意味を成さないものが多く含まれていた。各属性についてこれらの意味の成さないデータを除いたものを学習に使用した。

実験結果を示す。

まず、性別、居住域、年齢の推定精度と素性数の関係を表 2 に示す。

表 2 教師データ(1024 サイト)を用いた場合の推定精度

素性数	性別		居住域		年齢層	
	サイト単位	記事単位	サイト単位	記事単位	サイト単位	記事単位
16	.6588	.7085	.3860	.5213	.1625	.2938
128	.8038	.8519	.5843	.7219	.4335	.5536
1024	.8680	.8830	.6985	.8148	.5093	.5564
8192	.8515	.8656	.6868	.7561	.5003	.5418

次に素性数を 1024 語にした場合、教師データ数を変化させた場合の推定精度を表 3 に示す。

表 3 固定した素性データ(1024 語)を用いた場合の推定精度

教師データ数	性別		居住域		年齢層	
	サイト単位	記事単位	サイト単位	記事単位	サイト単位	記事単位
16	.7030	.6989	.4485	.5109	.3618	.3642
128	.8433	.8521	.6288	.7440	.4058	.4530
1024	.8680	.8871	.6985	.8148	.5093	.5564
8192	.8580	.8771	.7120	.8289	NA	NA

これらの結果、どの属性も教師データが多ければ多いほど精度が高くなる傾向がある。しかし、精度の伸びは鈍化してきている。これより、属性値の種類が数個であれば、教師データは数千件程度あればよいことがわかる。ただ性別と比べて居住域と年齢層では精度が安定するのに必要なデータ量が多くなっており、属性値のとりえる値の数が多ければその分多くの教師データが必要となることを

示唆している。

スケーラビリティを確認するため、大量のデータを用いた実験を行った。比較のため、CNB と代表的な高精度文書分類器である SVM で性別推定を行い、学習に要した時間と精度を比較した。これについても精度は教師データと異なる 4000 サイトのデータで測定した。結果を表 4 に示す。この結果より、大量のデータがある際には CNB は SVM よりも圧倒的に高速になることが分かる。

表4 CNB とSVM の学習速度と精度の変化(素性は1024 語)

	性別					
	CNB			SVM		
教師データ数	サイト単位	記事単位	時間 (秒)	サイト単位	記事単位	時間 (秒)
1024	.8680	.8830	.10	.8280	.8764	.1.38
8192	.8580	.8771	.19	.8173	.9094	38.57
65536	.8613	.8735	.88	.8303	.9202	419.15

4 成果要覧

受賞関連

[受賞 1] 優秀プレゼンテーション賞 DEWS2007, 佐藤一誠, 中川裕志: 階層ベイズモデルによる多重トピック文書の確率的生成モデルの構築, 電子情報通信学会第 18 回データ工学ワークショップ/第 5 回 DBSJ 年次大会 M5-3, 2007 年 3 月.

[受賞 2] MPS プレゼンテーション賞, 佐藤一誠, 中川裕志: Dirichlet Process Unigram Mixture Model に対する Collapsed Variational Bayes inference の適用, 情報処理学会 第 64 回 数理モデルと問題解決研究会 07-64-08, 2007 年 5 月.

査読付き論文リスト

[査読付 1] Issei Sato, Hiroshi Nakagawa: Knowledge Discovery of Multiple-topic Document using Parametric Mixture Model with Dirichlet Prior, Thirteenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp.590-598, SanJose, USA, Aug 15-18, 2007.

[査読付 2] 佐藤一誠, 中川裕志: Dirichlet Process Unigram Mixture Model に対する Collapsed 変分ベイズ法の適用, 情報処理学会論文誌, Vol.48 TOM19, pp.107-116, 2007.

[査読付 3] Issei Sato, Hiroshi Nakagawa: Semi-structure Mining Method for Text Mining with a Chunk-based Dependency Structure, Advances in Knowledge Discovery and Data Mining (The 11th Pacific-Asia Conference on Knowledge Discovery and Data Mining:PAKDD2007), pp. 777-784, Springer LNAI 4426, 2007.

[査読付 4] Nobuyuki Shimizu, Hiroshi Nakagawa: Structural Correspondence Learning for Dependency Parsing, EMNLP-CoNLL-ST 2007, Prague, Czech Republic, June 2007, pp.1166-1169.

[査読付 5] Issei Sato, Hiroshi Nakagawa: Bayesian Document Generative Model with Explicit Multiple Topics, EMNLP-CoNLL2007, Prague, Czech Republic, June 25-28, 2007, pp.421-429.

[査読付 6] Tomohiro Fukuhara, Takehito Utsuro, Hiroshi Nakagawa: Cross-Lingual Concerns Analysis from Multilingual Weblog Articles, The 6th International Workshop on Social Intelligence Design (SID 2007), pp.55-64, Trent, Italy, July 2-4 2007.

- [査読付 7] Minoru Yoshida, Hiroshi Nakagawa: Web Document Parsing: A New Approach to Modeling Layout-Language Relations, ICDAR2007 (The 9th International Conference on Document Analysis and Recognition), pp.203-207, Curitiba, Brazil, September 2007.
- [査読付 8] Ayako Hoshino, Hiroshi Nakagawa: A Cloze Test Authoring System and its Automation, ICWL2007-- The 6th International Conference on Web-based Learning, pp.174-181, Edinburgh, Scotland, August 15-17, 2007.
- [査読付 9] Minoru Yoshida, Akira Terada, Hiroshi Nakagawa: Gram-Free Synonym Extraction via Suffix Arrays, AIRS2008(Asia Information Retrieval Symposium 2008), pp.282-291, Harbin, China, January 2008.
- [査読付 10] 寺田昭, 吉田稔, 中川裕志: 同義語辞書作成支援ツール, 自然言語処理, Vol.15, 2008 (採録決定).

公開ソフトウェア

- [公開 1] 吉田稔, 中川裕志: UT-Kiwi, 学内 web ページからクローリングして収集したページに対して用例検索および同義語抽出を行うシステム <http://kiwi.r.dl.itc.u-tokyo.ac.jp/ut-kiwi/>

その他の発表論文リスト

- [発表 1] 佐藤一誠, 中川裕志: 階層ベイズモデルによる多重トピック文書の確率的生成モデルの構築, 電子情報通信学会第 18 回データ工学ワークショップ & 第 5 回日本データベース学会年次大会, 2007.
- [発表 2] 佐藤一誠, 中川裕志: Dirichlet Process Unigram Mixture Model に対する Collapsed Variational Bayes Inference の適用, 第 64 回数理解モデル化と問題解決(MPS)研究会, 情報処理学会, 2007.
- [発表 3] 大倉 務, 清水 伸幸, 中川 裕志: スケーラブルで汎用的なブログ著者属性推定手法, 情報処理学会・自然言語処理研究会 NL181-1, pp.1-6, 2007.
- [発表 4] 清水伸幸, 中川裕志: 同義語獲得のための距離学習, 第 2 回「機械翻訳技術のイノベーション」シンポジウム, 2008.
- [発表 5] 村上明子, 那須川哲哉, 中川裕志: オンラインディスカッションにおける有益発言の抽出, 言語処理学会第 14 回年次大会, C2-4, 2008.
- [発表 6] Yuxin WANG, Nobuyuki SHIMIZU, Minoru YOSHIDA, Hiroshi NAKAGAWA: Contextual Information Based Technical Synonym Extraction, 言語処理学会第 14 回年次大会, B4-4, 2008.
- [発表 7] 小野真吾, 佐藤一誠, 吉田稔, 中川裕志: 重要語抽出を用いた Web 文書上の同姓同名の曖昧さ解消, 電子情報通信学会 第 19 回データ工学ワークショップ 第 6 回日本データベース学会年次大会, A7-3, 2008.
- [発表 8] 佐藤有記, 宇津呂武仁, 福原知宏, 河田容英, 村上嘉陽, 中川裕志, 神門典子: キーワードの時系列特性を利用したスパムブログの収集・類型化・データセット作成, 第 19 回データ工学ワークショップ第 6 回日本データベース学会年次大会, A10-2, 2008.

特記事項

- [特記 1] 科学研究費補助金特定領域研究「情報爆発IT基盤」報告書(代表者) A01-05, 2008.

[特記 2] 経済産業省『情報大航海プロジェクト—新総合安全運航支援システム』に関わる基礎技術の研究報告書(分担), 2008.

[特記 3] 情報学研究所:公募型共同研究【研究テーマ提案型】「世界ニュースのための多言語ニュース記事及びウェブログ記事収集解析システムの構築と評価用データの構築」報告書, 2008.

自然言語構文解析とソーシャル辞書

二宮 崇

1 概要

本年度は、英語、日本語などの自然言語の構文解析の高精度化の研究を主に行った。構文解析は、文を入力とし、入力文の構文構造や単語間の依存構造を出力することを目的としたタスクであり、例えば、英語の主語や目的語を同定したり、名詞句や動詞句、関係節といった言語学的に規定される句構造を同定することを目的としている。構文解析により、世の中に大量に存するウェブテキストや新聞、論文、図書から種々の知識が構文構造または述語-項関係の形式で獲得できるため、より高度な情報検索・情報抽出システムの開発および自動図書リファレンスサービス等を含む自動応答システムの開発のための重要な基礎技術として期待されている。本研究では、主辞駆動句構造文法 (HPSG) と呼ばれる言語学的に厳密に定義された文法を用いた構文解析の高精度化の研究を行った。また、ウェブ上で用いられる辞書のソーシャル化の研究についても簡単に報告する。具体的には、下記の研究内容について報告する。

スーパータグ参照分布による HPSG 構文解析 [査読付 1] HPSG 構文解析のための新しい確率モデルの研究を行った。提案するモデルは、単語と品詞の特徴のみを用いた辞書項目の選択確率の積で定義される確率分布 (スーパータグ確率分布) を最大エントロピー法における参照分布とした確率的 HPSG である。我々が昨年度提案した確率モデルはスーパータグ確率分布と従来の確率的 HPSG を独立に学習させた後、構文解析時は単純にそれらの積を全体の確率モデルとしたため、最尤推定にはなっていなかった。今年度は、全体の確率モデルを最尤化するため、スーパータグ確率分布を最大エントロピー法における参照分布とし、確率的 HPSG を学習した。これにより、構文解析精度である F スコアが 0.8% ポイント上昇した。

ソーシャル辞書 [発表 1] ソーシャル化とは、個人に固有な情報を収集して統合することにより、単なる固有情報の集合以上のより有益な情報資源を作り出すことを指す。本研究では大多数の語学学習者に利用されることを想定したオンライン辞書を提供することにより、語学学習者が学習すべき単語を推薦するシステムを提案する。

また、社会科学研究所図書職員の前田 朗氏と協力して、図書館情報係が事務局となって開催された「図書系職員のためのアプリケーション開発講習会」に講師として参加した。詳細は、教育・サービス活動の章の図書館電子化部門、図書館関係システム運用・管理の講習会・研究会開催報告を参照されたい。

2 スーパータグ参照分布による HPSG 構文解析

2.1 背景

十数年以上前ではおもちゃ文法 (toy grammar) と非現実的な人工的文章に対してのみ構文解析の研究が行われていたが、近年、現実存在する文章に対する高精度かつ高被覆¹な構文解析の研究が多くなされるようになってきた。このような現実の文章に対して構文解析を行う際の大きな問題は、構文構造の曖昧性とその解消であり、例えば、最先端の構文解析技術では、一文あたり数千から数百億の構文構造候補が出力されるが、正しい構文構造はその中に一つしかない。曖昧性解消の際には、構文構造候補の句構造の形が曖昧性解消を行う大きな手がかりになると考えられるため、多くの構文解析の研究では句構造に対して、確率事象を定義することが多かった。例えば、確率的 CFG (PCFG) では文法規則に対して確率が定義され、主辞駆動句構造文法 (HPSG) や組み合わせ範疇文法 (CCG) では解析結果の句構造に対して確率が定義されてきた。今までに様々な構文解析の研究が行われてきたが、基本的なアイデアは句構造や文法規則の特徴を捉えることであり、一方、句構造の一部である語彙情報を利用することは構文解析の精度を大きくあげることが理解されてはいるものの補助的に用いられるだけであった。

一方、構文解析の曖昧性解消問題に対する非常に興味深い手法として、スーパータグ付けというアイデアが語彙化木接合文法 (LTAG) の文法枠組の中で Bangalore らによって 1999 年に提案された。スーパータグ付けは、品詞タグが付与された文に対し、スーパータグを付与するタスクであり、ここで、スーパータグというのは、LTAG では基本木のことであり、HPSG や CCG では辞書項目に相当する。スーパータグ付けは、軽いプロセスであるため、重い構文解析前に辞書項目候補を枝狩りするためによく使われる。Bangalore らは、「もし正しくスーパータグを付与することができれば、構文解析はほぼ自明である」と主張している。この意味するところは、スーパータグには下位範疇化フレームなど構文構造の情報も含まれているため、もし、スーパータグを正しく付与することができるのなら、それにより構文構造がほとんど決定する、ということである。2004 年 Nasr らは、正しいスーパータグが付与された文に対しては非常に高い精度で構文解析が行えることを実証し、同 2004 年 Clark らは CCG 構文解析で構文解析速度を非常に大きく向上させることに成功した。スーパータグ付けのアイデアは非常にシンプルであり、構文解析の前処理モデルとして非常に面白く、その効果も実証されている。しかしながらこれら今まで行われてきたスーパータグ付けの手法は、スーパータグ付け器を外部タグ付け器として用い、辞書項目の候補を枝狩りしメインの構文解析を補助する目的でのみ用いられてきた。

本研究では、昨年度に HPSG 構文解析の極端なケースとして、スーパータグ付けの確率 (スーパータグ確率分布) のみ、つまり、辞書項目を選択する確率のみで HPSG 構文解析を行うモデル (極語彙化モデル)、および、従来の確率的 HPSG モデルとの混合モデルについて提案した。混合モデルを用いた場合は 3~4 倍程度の高速化を達成し、また、およそ 2% ポイントの F スコアの向上を確認した。しかし、混合モデルは、スーパータグ確率分布と従来の確率的 HPSG モデルを独立に学習し、構文解析時は単純にそれらの積を与えただけであった。本研究では、スーパータグ確率分布を参照分布とする確率的 HPSG モデルを提案する。スーパータグ確率分布を参照分布として確率的 HPSG モデル全体が最適化されるように学習されるため、従来の独立した学習よりも良い確率モデルが得られる。

2.2 内容

本研究では、辞書項目の選択確率の積により与えられる確率分布 (スーパータグ確率分布) を最大エントロピー法の参照分布として与える確率的 HPSG モデルを提案する。実験では、従来の確率的 HPSG、

¹ 構文解析可能な文の割合が高いこと。特定のテキストに対して構文解析結果を出力する文の割合や依存構造の割合のことを被覆率という。

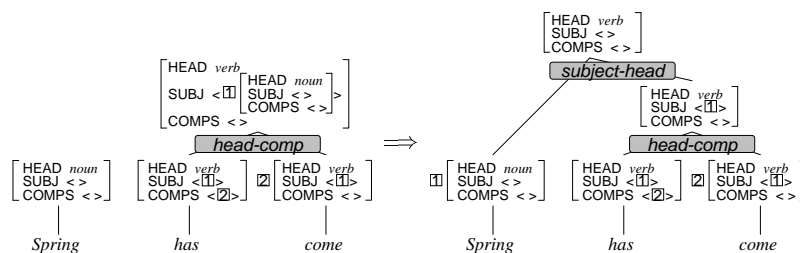


図 1: HPSG 構文解析の例

極語彙化モデル、それらの積による混合モデル、および提案するスーパータグ参照分布による確率的 HPSG との比較について報告する。

スーパータグ確率分布は、単語品詞 n グラムによる単純な辞書項目の選択確率の積で定義される。つまり、この確率分布は句構造の特徴を一切用いない。この辞書項目の選択確率の計算は、品詞タグ付けと同様の技術で実現可能であり、実験では、単語 3 グラム、品詞 5 グラムを用い、機械学習における選別手法として有名なエントロピー最大化法により計算された。このスーパータグ確率分布だけを用いて HPSG 構文木の確率を与えるモデルを極語彙化モデルと呼ぶ。このモデルでは、辞書項目の選択確率しか用いていないが、単に確率の高い辞書項目だけが選択されるのではなく、文法により適格な辞書項目のうち最も確率の高い辞書項目列が選択される。辞書項目の選択確率しか用いていないが、従来の複雑な句構造ベースの確率モデルと同程度の高い精度を達成できることが確認されている。

まず、HPSG 構文解析と従来の確率モデルについて説明を行い、その後、極語彙化モデル、混合モデルおよび本研究で提案するスーパータグ確率分布を参照分布とした確率的 HPSG について説明する。

2.2.1 HPSG 構文解析と従来の確率モデル

HPSG は制約に基づく語彙化文法と呼ばれる統語理論であり、少数の非常に一般的な生成規則と単語特有の性質を記述した非常に多くの辞書項目からなる。生成規則と辞書項目は素性構造と呼ばれるグラフ構造により記述され、文の構造は、それら辞書項目と生成規則の組み合わせにより説明される。文の構造の間の制約は素性構造の単一化により実現される。図 1 は “*Spring has come*” に対する HPSG 構文解析の例を示している。まず、“*has*” と “*come*” に対する語彙項目が “*Head-Complement Schema*” と呼ばれる生成規則の子ノードに相当する素性構造と単一化され、親ノードが生成される。より大きな句構造は同様に生成規則を繰り返し適用することにより得られ、最終的に文全体に対応する句構造が構文解析結果として出力される。

素性構造で定義される文法枠組に対して、確率モデルを定義することは単純ではないが、Abney、Johnson ら、Riezler ら、Malouf と van Noord、Kaplan ら、Miyao らによって、HPSG を含む単一化ベースの文法に対する確率モデルが、対数線形モデルまたはエントロピー最大化モデルと呼ばれる確率モデルとして提案された。文 $w = \langle w_1, \dots, w_n \rangle$ が与えられたとき、構文解析結果 T に対する確率は次のように与えられる。

$$p_{hpsg}(T|w) = \frac{1}{Z_w} \exp \left(\sum_u \lambda_u f_u(T) \right), Z_w = \sum_{T'} \exp \left(\sum_u \lambda_u f_u(T') \right)$$

ただし、 λ_u はモデルパラメータであり、 f_u は構文構造 T の特徴を表す素性関数であり、 Z_w は与えられた文に対する全ての可能な構文構造に対する確率の和である。つまり、構文構造 T にあられる特徴 f_u に対応する正規化された重み $\exp \lambda_u$ の積として定義される。モデルパラメータ λ_u は訓練データに対し、対数尤度を最大化するよう推定される。

また、宮尾らはこれに “予備的確率モデル” と呼ぶ確率 $p_0(T|w)$ を参照分布として導入している。予備的確率モデルは構文解析を必要とせず全体の確率を近似する確率モデルを想定しており、パラ

表 1: 素性

$f_{binary} = \left\langle \begin{array}{l} r, d, c, \\ sp_l, sy_l, hw_l, hp_l, hl_l, \\ sp_r, sy_r, hw_r, hp_r, hl_r \end{array} \right\rangle$ $f_{unary} = \langle r, sy, hw, hp, hl \rangle$ $f_{root} = \langle sy, hw, hp, hl \rangle$ $f_{lex} = \langle w_i, p_i, l_i \rangle$	<p>combinations of feature templates for f_{binary}</p> <hr/> $\langle r, d, c, hw, hp, hl \rangle, \langle r, d, c, hw, hp \rangle, \langle r, d, c, hw, hl \rangle,$ $\langle r, d, c, sy, hw \rangle, \langle r, c, sp, hw, hp, hl \rangle, \langle r, c, sp, hw, hp \rangle,$ $\langle r, c, sp, hw, hl \rangle, \langle r, c, sp, sy, hw \rangle, \langle r, d, c, hp, hl \rangle,$ $\langle r, d, c, hp \rangle, \langle r, d, c, hl \rangle, \langle r, d, c, sy \rangle, \langle r, c, sp, hp, hl \rangle,$ $\langle r, c, sp, hp \rangle, \langle r, c, sp, hl \rangle, \langle r, c, sp, sy \rangle$ <hr/> <p>combinations of feature templates for f_{unary}</p> <hr/> $\langle r, hw, hp, hl \rangle, \langle r, hw, hp \rangle, \langle r, hw, hl \rangle, \langle r, sy, hw \rangle,$ $\langle r, hp, hl \rangle, \langle r, hp \rangle, \langle r, hl \rangle, \langle r, sy \rangle$ <hr/> <p>combinations of feature templates for f_{root}</p> <hr/> $\langle hw, hp, hl \rangle, \langle hw, hp \rangle, \langle hw, hl \rangle,$ $\langle sy, hw \rangle, \langle hp, hl \rangle, \langle hp \rangle, \langle hl \rangle, \langle sy \rangle$ <hr/> <p>combinations of feature templates for f_{lex}</p> <hr/> $\langle w_i, p_i, l_i \rangle, \langle p_i, l_i \rangle$ <hr/>
--	--

r name of the applied schema
 d distance between the head words of the daughters
 c whether a comma exists between daughters and/or inside daughter phrases
 sp number of words dominated by the phrase
 sy symbol of the phrasal category
 hw surface form of the head word
 hp part-of-speech of the head word
 hl lexical entry assigned to the head word
 w_i i -th word
 p_i part-of-speech for w_i
 l_i lexical entry for w_i

メータ推定を容易にするために導入されている。この確率が低い構文構造の計算はパラメータ推定の段階で省略されるため、パラメータ推定が高速化される。この予備的確率モデルを導入した構文構造全体に対する確率モデルは次のように定義される。

(従来の確率的 HPSG)

$$p_{hpsg'}(T|\mathbf{w}) = p_0(T|\mathbf{w}) \frac{1}{Z_{\mathbf{w}}} \exp \left(\sum_u \lambda_u f_u(T) \right)$$

$$Z_{\mathbf{w}} = \sum_{T'} p_0(T'|\mathbf{w}) \exp \left(\sum_u \lambda_u f_u(T') \right), p_0(T|\mathbf{w}) = \prod_{i=1}^n p(l_i|w_i)$$

ただし、 l_i は T 中出现する単語 w_i に対して割り当てられる辞書項目であり、 $p(l_i|w_i)$ は w_i に対して l_i を選択する確率である。

評価ではこのモデルと比較をする。このモデルで使われた素性 (=機械学習のための特徴ベクトルの要素) は表 1 のとおりである。素性テンプレート f_{binary} と f_{unary} は二分岐と単分岐に対応し、 f_{root} は構文構造のルートに対応し、 f_{lex} は予備的確率モデルの語彙項目に対応する。

2.2.2 スーパータグ参照分布の確率的 HPSG

本研究ではスーパータグ確率分布を最大エントロピー法の参照分布とした確率的 HPSG モデルを提案する。このモデルはスーパータグ付け器の確率を参照分布としたモデルともとらえられるし、予備的確率モデルの単語品詞ユニグラムを単語品詞 n グラムに拡張したモデルともとらえられる。まず、スーパータグ確率分布について説明する。

辞書項目集合 L と、文 $\mathbf{w} = \langle w_1, \dots, w_n \rangle$ 、辞書項目の選択確率 $p(l_i \in L|\mathbf{w}, i)$ が与えられたとき、スーパータグ確率分布は次のように定義される。

(スーパータグ確率分布)

$$p_{suptag}(T|\mathbf{w}) = \prod_{i=1}^n p(l_i|\mathbf{w}, i),$$

ただし、 l_i は構文構造 T 中出现する単語 w_i に割り当てられる辞書項目であり、 $p(l_i|\mathbf{w}, i)$ は w_i に対して l_i を選択する確率である。極語彙化モデルは文法的に許される構文木に対しスーパータグ確率分布を与えた確率モデルである。スーパータグ確率分布と従来の確率的 HPSG を独立に学習し、それらの確率の積を確率モデルとしたモデルを混合モデルと呼ぶ。

我々が提案する確率モデルは、スーパータグ確率分布を参照分布とした確率的 HPSG であり、従来の確率的 HPSG の予備的確率モデルがスーパータグ確率分布に置き換えられている。

表 2: 辞書項目の選択確率のための素性

$f_{exlex} = \left\langle \begin{array}{c} w_{i-1}, w_i, w_{i+1}, \\ p_{i-2}, p_{i-1}, p_i, p_{i+1}, p_{i+2} \end{array} \right\rangle$	combinations of feature templates
	$\langle w_{i-1} \rangle, \langle w_i \rangle, \langle w_{i+1} \rangle,$ $\langle p_{i-2} \rangle, \langle p_{i-1} \rangle, \langle p_i \rangle, \langle p_{i+1} \rangle, \langle p_{i+2} \rangle, \langle p_{i+3} \rangle,$ $\langle w_{i-1}, w_i \rangle, \langle w_i, w_{i+1} \rangle,$ $\langle p_{i-1}, w_i \rangle, \langle p_i, w_i \rangle, \langle p_{i+1}, w_i \rangle,$ $\langle p_i, p_{i+1}, p_{i+2}, p_{i+3} \rangle, \langle p_{i-2}, p_{i-1}, p_i \rangle,$ $\langle p_{i-1}, p_i, p_{i+1} \rangle, \langle p_i, p_{i+1}, p_{i+2} \rangle$ $\langle p_{i-2}, p_{i-1} \rangle, \langle p_{i-1}, p_i \rangle, \langle p_i, p_{i+1} \rangle, \langle p_{i+1}, p_{i+2} \rangle$

表 3: Section 23 に対する実験結果

	Section 23		
	適合率	再現率	F スコア
従来の確率的 HPSG	87.3%	86.5%	86.9%
極語彙化モデル	87.2%	86.5%	86.9%
混合モデル	89.5%	88.6%	89.0%
(本研究) スーパータグ参照分布	90.0%	89.6%	89.8%

(スーパータグ参照分布の確率的 HPSG)

$$p_{ref}(T|\mathbf{w}) = \frac{1}{Z_{ref}} p_{suptag}(T|\mathbf{w}) \exp\left(\sum_u \lambda_u f_u(T)\right)$$

$$Z_{ref} = \sum_{T'} p_{suptag}(T'|\mathbf{w}) \exp\left(\sum_u \lambda_u f_u(T')\right).$$

実験では、句構造に対するモデルパラメータは、スーパータグ参照分布が与えられた上で最尤化される。

辞書項目の選択確率 $p(l_i|\mathbf{w}, i)$ は次のように定義される。

(辞書項目の選択確率)

$$p(l_i|\mathbf{w}, i) = \frac{1}{Z_w} \exp\left(\sum_u \lambda_u f_u(l_i, \mathbf{w}, i)\right), Z_w = \sum_{l'} \exp\left(\sum_u \lambda_u f_u(l', \mathbf{w}, i)\right).$$

ただし、 Z_w は w_i に対する可能な全ての辞書項目候補に対する確率の和である。実験で用いられた素性テンプレートは表 2 に記載されており、単語 3 グラムと品詞 5 グラムからなる。

2.2.3 評価

従来の確率的 HPSG、極語彙化モデル、混合モデルおよびスーパータグ参照分布による構文解析の解析速度と精度を評価した。文法は、Enju 2.1 と呼んでいる Penn Treebank の Section 02~21 (39,832 文) を用いて辞書項目を獲得した HPSG 文法を用いた。この文法の辞書は、10,536 語に対する 3,797 辞書項目からなる。確率モデルも同様に Penn Treebank の Section 02~21 を用いて学習し、Section 22 を開発用テストセットとして用い、Section 23 を用いて性能評価を行った。精度の評価は、HPSG 構文解析の結果出力される述語項構造の依存構造に対する適合率、再現率およびそれらの調和平均である F スコアで評価した。

Section 23 に対する結果は、表 3 に示されている。本研究で提案するスーパータグ参照分布の確率的 HPSG モデルは、従来の確率的 HPSG と比較し 2.9%ポイント高い F スコアを達成し、昨年度我々が提案した混合モデルと比較し、0.8%ポイント高い F スコアを達成した。

2.3 具体的成果

本研究では、HPSG 構文解析の高精度化のため、スーパータグ確率分布を参照分布とした確率的 HPSG モデルを提案した。スーパータグ確率分布は辞書項目の選択確率の積で定義される非常にシンプルな確率分布でありながら、この分布を参照分布として用いた我々のモデルは、従来の確率的 HPSG より 2.9%ポイント高い F スコアを達成し、昨年度我々が提案した混合モデルより 0.8%ポイント高い F スコアを達成した。

スーパータガーを用いる研究はいくつか存在するが、句構造に対する確率モデルとの融合の方法はいずれも最尤推定とはなっておらず、例えば、辞書項目候補の枝刈りだけに用いられたり、単純にスーパータグ確率と句構造に対する確率との積が与えられるだけで、確率モデルとしては不十分であった。本研究は初めてスーパータグ確率と句構造に対する確率モデルの融合を最尤推定の枠組で与え、実験によりその性能を確認した。また、スーパータグ確率分布は、学習の際の枝刈りにも用いられるため、最尤推定の学習時間が大幅に短縮された。

3 ソーシャル辞書

3.1 背景

近年、語学学習の分野において、英日辞書を含む多種の辞書が電子化され、ウェブ上でオンライン辞書が無料で提供され、オンライン辞書を利用する学習者が増えている。学習者が辞書を用いて検索した単語の履歴である辞書引き履歴には、学習者が理解していない単語など有用な情報が多く含まれており、多くの電子化辞書が辞書引き履歴を記録し、その学習者が後で閲覧することが可能な機能を提供している。しかし、辞書引き履歴は学習者個人の情報としてみなされることが多く、多くの学習者の辞書引き履歴を収集することは、あまり一般的ではなかった。

近年では、このような個人に固有な情報を収集して統合することにより、単なる固有情報の集合以上のより有益な情報資源を作り出すことをソーシャル化と呼び、注目されている。たとえば、アマゾン²をはじめとする多くのショッピングサイトにおいては、商品の購入履歴をソーシャル化することによって、人気商品の推薦などの新しい機能が提供されるようになっている。

本稿では、前述のショッピングサイトなどの手法を参考に、学習者が学習すべき単語を推薦するシステムを提案する。ショッピングサイトにおいて、個人の商品の購入履歴を収集したデータを活用して、購入者に商品の推薦を行う手法は“協調フィルタリング”と呼ばれる。本研究では、この協調フィルタリングを辞書引き履歴に応用し、学習者に単語の推薦を行うシステムを提案する。

3.2 内容

本研究では、語学学習者がオンライン辞書を引くときに使用した検索キーワードである辞書引きキーワードを収集し、語学学習者に学習すべき新しい単語を提示するシステムを提案する。提案するシステムは、推薦システムと辞書システムから成る。辞書システムは、辞書引きキーワードに基づいて辞書を検索し、学習者に辞書項目を提示する。推薦システムは、語学学習者が辞書を引くときに使用した検索キーワードである辞書引きキーワードを収集し、協調フィルタリングを用い単語の推薦を行う。ここでは、辞書引きキーワードの履歴を辞書引き履歴と呼ぶこととする。まず、協調フィルタリング手法について説明し、その後本システムについて説明する。

3.2.1 協調フィルタリング

協調フィルタリングとは、次のようなアルゴリズムの総称である。まず、多数のユーザーのアイテムに関する嗜好情報であるプロファイルが与えられる。次に、アクティブユーザーが 1 人と与えられる。プロファイルの情報を参考にして、アクティブユーザーに推薦すべきアイテムを抽出する。プロファ

²<http://www.amazon.com/>

イルには、ユーザーを行、アイテムを列とし、ユーザーのアイテムに対する評価を要素とする行列がよく用いられる。ユーザーのアイテムに対する評価が直接得られない購入履歴を対象とする場合は、購入回数を正規化することによって、アイテムに対する評価値とみなす。

本研究の実装では、協調フィルタリングアルゴリズムの一つであるスロープワン法を用いた。スロープワン法では、次のようにして、アクティブユーザーに推薦するアイテムを求める。まず、ユーザー i のアイテム j に対する評価値が、要素 $m_{i,j}$ に格納されているプロファイル行列 $M = (m_{i,j})$ を考える。評価値は区間 $(0, 1]$ の値とする。評価値は値が大きいほど評価が高いとする。ユーザーによる評価が未定義である場合のみ、 $m_{i,j} = 0$ とする。

この時、アイテム j_1 と j_2 を共に評価しているユーザーのインデックスの集合を、

$$I(j_1, j_2) = \{i | m_{i,j_1} \neq 0 \text{ かつ } m_{i,j_2} \neq 0\} \quad (1)$$

と定義する。そして、アイテム j_1 と j_2 の平均の評価の差 $dif(j_1, j_2)$ を次のように定義する。

$$dif(j_1, j_2) = \frac{1}{|I(j_1, j_2)|} \sum_{i \in I(j_1, j_2)} m_{i,j_1} - m_{i,j_2} \quad (2)$$

スロープワン法ではアクティブユーザー i の評価が未定義であるアイテム j に対して、次のように、予想評価値 $r_{exp}(i, j)$ を定める。

$$r_{exp}(i, j) = \frac{1}{\sum_{j' \neq j} |I(i, j')|} \sum_{j' \neq j} dif(j, j') + m_{i,j'} \quad (3)$$

したがって、アクティブユーザー i に推薦されるアイテムは、アクティブユーザーの評価が未定義であるアイテム j の中で、予想評価値 $r_{exp}(i, j)$ が $1 \sim N$ 番目に高いものとなる。

3.2.2 システム

学習者情報の登録 学習者は、推薦システムに対してユーザー名やパスワードなど、学習者を一意に特定するための“学習者 ID”を事前に登録しておく。

辞書引き情報の蓄積 辞書を用いて単語を検索する手順について説明する。学習者は、通常の辞書引きシステムと同様に、辞書システムと推薦システムそれぞれに対して辞書引きキーワードを送信する。辞書システムから学習者に辞書を引いた結果が返信される。通常のオンライン辞書を用いた辞書引きと異なるのは、学習者を特定するための学習者 ID を推薦システムに対して渡さなければならない点である。学習者 ID については、ブラウザに記録した認証情報を使って認証を行えば、学習者が毎回入力する必要はなくなる。学習者は、既存のオンライン辞書とほぼ同様の手順で、本稿のシステムを用いて辞書を引くことが可能である。本研究の実装では、辞書引きシステムとして、フリーの英英辞典である Wiktionary³を用いた。

単語推薦 学習者が自分に関する単語を推薦するように推薦システムに要求すると、推薦システムが蓄積された辞書引き履歴から推薦アルゴリズムを用いて学習者に単語の推薦を行う。本研究の実装では協調フィルタリングに Taste⁴というライブラリを用いた。単語推薦アルゴリズムは、スロープワン法を用いた。スロープワン法は、学習者(ユーザー)の単語(アイテム)に対する評価値を用いる手法であるので、学習者 i が単語 j を検索した回数 $N(i, j)$ を、区間 $[0, 1)$ の評価値 $r(i, j)$ に変換する必要がある。本研究では、シグモイド関数を参考に $N(i, j) \geq 0$ に対して値域が $[0, 1)$ である次の関数を用いた。

$$r(i, j) = \left(\frac{1}{1 + e^{-N(i,j)}} - \frac{1}{2} \right) * 2 \quad (4)$$

³<http://en.wiktionary.org/wiki/>

⁴<http://taste.sourceforge.net/>

この関数では、学習者 i が単語 j を辞書引きしていない、すなわち、 $N(i, j) = 0$ である場合には、 $r(i, j) = 0$ となる。また、 $N(i, j) = 1$ の場合には 0.462 となり、以後、 $N(i, j)$ の増加に伴い 1 に漸近する。

3.3 具体的成果

本研究は、学習者の辞書引き履歴を蓄積し、協調フィルタリングを用いて単語を推薦するシステムを提案、実装した。このシステムによって、学習者は単語を効率的に学習することが可能になり、学習者の語学学習を支援することができると考えられる。

また、本システムを多数の学習者が利用することにより、辞書引き履歴が自動的にかつ大量に蓄積されるが、このようにして得られる辞書引き履歴を語学学習支援研究のためのコーパスとして利用することが考えられる。語学学習支援研究のためのコーパスとしては、学習者が作文(書く・話す)した文を収集した学習者コーパスに注目が集まっており、最近、NICT JLE Corpus などの大規模なコーパスが整備されてきている。学習者コーパスは、多くの情報が取れるという意味で質が高く非常に有用であるが、大量に用意するのが難しい。本研究で取り上げた辞書引き履歴は、逆に、質は低いけど自動的にかつ大量に蓄積することが可能だと考えられ、この辞書引き履歴を学習者コーパスの一種として利用することにより、より良い語学学習支援研究が可能になるものと期待される。

本研究のシステムの詳細については脚注のサイト⁵を参照されたい。現在、システムは図書館電子化研究部門内のみに限定的に利用可能であり試験運用中であるが、来年度である 2008 年度に一般から利用可能になるよう公開の予定である。

4 成果要覧

査読付論文

[査読付 1] Takashi Ninomiya, Takuya Matsuzaki, Yusuke Miyao and Jun'ichi Tsujii: A log-linear model with an n-gram reference distribution for accurate HPSG parsing, In the Proc. of IWPT-2007, pp.60-68, 2007.

その他の発表論文

[発表 1] 江原遥, 二宮崇, 中川裕志: 辞書引き履歴のソーシャル化による学習者支援, 第 14 回言語処理学会年次大会併設ワークショップ「教育・学習を支援する言語処理」, 東京, 2008 年 3 月.

⁵<http://caelum.r.dl.itc.u-tokyo.ac.jp/~chara/?socialdictionary>

検索支援のためのテキストマイニング技術

— 同義語・接続語の高速検索と Web 文書レイアウト解析 —

吉田 稔

1 概要

近年のコンピュータ及びネットワークの普及に伴い、Web ページや電子メールといった基本的な電子文書のみならず、電子ジャーナルや組織内レポート等、従来は紙媒体で流通していた多種多様な文書が電子的に蓄積される状況が一般的となりつつある。検索技術はこのような電子文書の活用において必要不可欠な技術であり、多量な電子文書を最大限に活用するためには、テキスト検索の高度化が欠かせない。本年度は、より高度なテキスト検索の実現に向けて、以下のテーマについて研究を行った。本稿では、このうち最初の 2 つについて詳細に解説する。

- 同義語・接続語の高速提示[査読付 1, 発表 1, 発表 3, 発表 7]

入力クエリに対し、On-Demand かつ高速に同義語抽出や接続語抽出といったテキストマイニングを行うための手法を研究した。また、マイニング結果を検索支援のために提示するシステム UT-Kiwi を開発し、公開を行った。

- Web 文書構造の解析[査読付 2]

Web 文書のレイアウトを解析し、タイトルや見出しといった要素を抽出することで、検索結果の構造的表示、章立て等の文書構造を利用したより精度の高い検索の実現、等が可能になる。本年度は、レイアウト解析に、言語処理に用いられる構文解析技術を応用する試みを行った。

- 数値による検索システムの開発[発表 4, 発表 5, 発表 9, 発表 10]

テキスト中の数値やその属性を抽出することで、特定の数値(売上高等)の大きさの順にテキストを並べかえる等、数値を利用したテキスト検索が可能となる。本年度は、機械学習手法による属性抽出の精度向上について主に研究を行った。

2 UT-Kiwi: 検索支援としてのテキストマイニングシステム

2.1 背景

近年、WWW 上や組織内に蓄積される電子的文書の量は増大の一途を辿り、それらの文書を人間が把握することが困難となっている。WWW 全体における文書量の増大のみならず、その中でトピックの限定された部分集合(特定サイト内の Web 文書集合、Wikipedia、さらには企業内文書集合等)のサイズも増大し、WWW 全体と同様に、把握が困難となりつつある。

検索エンジンは、文書集合を効率的に利用する有力な手段の一つであるが、その機能は基本的には「入力された文字列(クエリ)を含む文書を返す」という範囲に留まっており、「どのようなクエリを入力するべきか」という問題に対しては、研究の余地が多く残されている。

本研究では、このような「クエリ入力支援」に対する新しい提案として、テキストマイニングによる検索支援システム UT-Kiwi を提案する。

テキストマイニングとは、与えられたテキスト集合の中での、「言葉の使われ方」(主に、言葉に関す

る統計的情報)について分析するタスクである。UT-Kiwi では、入力されたクエリに対し、「用例抽出」「同義語抽出」という二種類のテキストマイニングをリアルタイムに行い、マイニング結果を提示する。

UT-Kiwi は、前述の「トピックの限定された文書集合」を主に対象としており、それらのテキストに対する検索を支援する。基本アイデアは、「文書集合をオンメモリで配置し、Suffix Array による高速検索を実装する。さらに、Suffix Array による検索を応用することで、高速なテキストマイニングを行う」というものである。これにより、テキストマイニングを検索支援として用いることが可能となる。

2.2 内容

UT-Kiwi(<http://kiwi.r.dl.itc.u-tokyo.ac.jp/ut-kiwi/>)は、現在東京大学の Web サイト内の HTML 文書を文書集合として、サービスを行っている。図1に、UT-Kiwi の画面を示す。図中、A.に示されているのが検索窓であり、ユーザがここにクエリを入力すると、システムは入力内容に応じて用例や同義語を提示する。ここで用例とは、クエリ前方、後方にどのような文字列が頻繁に接続しているかを示したものであり(図中「B.後方接続」「C.前方接続」)、この場合、「シンポジウム」の前方に「記念」や「国際」、後方に「日本」や「講演」といった言葉が続き易いことがわかる。同様に、シンポジウムの同義語として、「講演会」や「ワークショップ」という文字列が用いられやすいということも提示する(図中、「D.類義語・同義語」)。ユーザは、提示された用例等をクリックすることで、クエリの補完を行える。補完された結果は、そのまま Google へのクエリとして、検索に用いることができる。

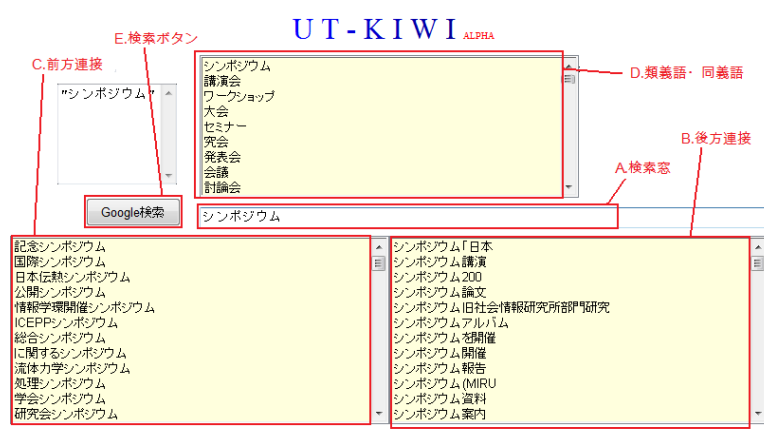


図1 : UT-Kiwi

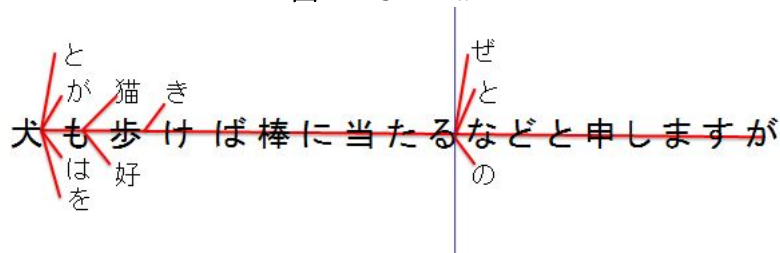


図2 : Kiwi アルゴリズム概要

2.2.1 用例検索

用例検索は、Kiwi というアルゴリズムを用いて行う。図2に、Kiwi アルゴリズムの概要を示す。Kiwi アルゴリズムでは、ある文字列 s が与えられたとき、「 s に接続する文字の種類数」(以下、「分岐数」)を計測する。例えば、図2では、「犬も」の後続文字として、「猫」「歩」「好」の3種類の文字があり、分

岐数は3となる。文字列 s にどのような文字列が接続するかを、木構造の探索により列挙していくが、そのさい、分岐数が増加している場合のみ、用例候補として考慮する。(図2の例では、例えば「犬も歩けば棒に当たる」が、分岐数1から4へ変化する点として、候補になる。)これは、「分岐数の増加は、意味的な切れ目とよく合致する」という経験則に基づいたルールである。また、UT-Kiwi では、用例候補が見つかったノードに対しては、それ以上深く探索を行わないという制限を加えることにより、探索の高速化を行っている。

最終的に、このようにして得られた用例候補を、スコア関数

$$\text{score}(s) = \text{freq}(s) \cdot \log(1+\text{length}(s))$$

により順位付けし、提示する。

2.2.2 同義語抽出

同義語抽出において問題となるのは、ユーザーによって入力される文字列の多様性である。同義語抽出や言い換えに関しては従来より多くの研究が行われているが、その大部分は、予め規定された候補語(例えば、「名詞」や「動詞+目的語」)どうしの類似度を測るものである。このため、本システムが想定する、「ユーザがあらゆる文字列を入力する状況」においては、文書のあらゆる部分文字列どうしについての類似度を計算する必要が生じ、必要な計算時間および記憶容量は膨大なものとなる。この問題に対し、本研究では、「ユーザからクエリが与えられた際、On-Demand にそれと類似する文字列を文書から取り出す」という問題設定を考える。このさいに要求されるのは、「どんな文字列にも対応できる」「高速な」アルゴリズムである。この実現に、Suffix Array による検索を活用する。

同義語抽出アルゴリズムでは、一般に「意味の類似する文字列は、同一の文脈で用いられることが多い」という仮説に基づき、文脈の類似度を計算することにより文字列の類似度を測定する。特に、「隣接する文字列」は、文脈として非常に有用であることが知られており、この「隣接する文字列」は、Suffix Array により容易に検索することができる。このため、本研究では、(1)「クエリ文字列に隣接する文字列」を検索し、(2)得られた隣接文字列から適切な「文脈」を選択し、(3)「文脈に隣接する文字列」を検索するという3段階の処理により同義語を抽出するアルゴリズムを開発した。以下、上記(1)(2)の処理をSTEP-1とし、(3)をSTEP-2として説明を行う。

高速化のため、文字列の検索を探索木として表現(図3・図4)し、探索対象の文字列のスコア付けを探索と同時に進行アルゴリズムを採用した。これにより、枝狩りによる探索空間の削減が可能になった。

以下、簡単のため、右側接続文脈の探索についてのみ述べる。STEP-1 では、クエリ q に接続する文脈 x のスコアを

$$\text{score}(x) = \text{freq}(q.x) \log |S| / \text{freq}(x)$$

で定義し(freq は頻度、 $q.x$ は文字列 q と x の連結、 $|S|$ はコーパス全体の文字列長)、スコア上位 N 個の文脈を取得するが、その際に、

$$\text{score}'(x) = \text{freq}(q.x) \log |S|$$

なる上限を用いて枝狩りを行う。 $\text{score}'(x) \geq \text{score}(x.y)$ が如何なる y についても必ず成り立つため、 $\text{score}'(x)$ が現在 N 位の候補のスコアを下回れば、それ以上の探索を打ち切ることができる。

STEP-2 では、効率的な枝狩りを行うため、(a)枝狩りに有用なスコア関数により、候補語を絞り(STAGE-1)、(b)より精緻なスコア関数により、絞られた候補語のリランキングを行う(STAGE-2)、という2段階の処理を行う。

STAGE-1 では、「同義語候補 x の左側(右側)に接続する文脈の種類」を「左(右)スコア」と定義し、まず左スコア上位 N 個の同義語候補、続いて右スコア上位 N 個の同義語候補を探索する。左スコア、右スコア共に、探索の深さに対する非増加な関数となるため、現在のスコアが N 位のスコアを下回った場合、そこで探索を打ち切ることができる。こうして得られた $2N$ 個の候補を、最終的に、スコア関数

$$\text{score}(x) = \sum \log(\text{freq}(c+x) / \text{freq}'(c+x))$$

により順位づけする。ここで、 \sum は、すべての文脈に関する和、 $c+x$ は、文脈 c と文字列 x を適切な

3 Web 文書構造解析への構文解析アプローチに関する研究

3.1 背景

本研究の目標は、Web 文書のレイアウトを解析するための汎用性の高いアルゴリズムの開発である。Web 文書においては、文章のみならず、レイアウト、すなわち、「文字の2次元的な配置」が、意味を表現する手段として頻繁に用いられる。図 5 は、単語間の意味的關係が、文章ではなく、配置により表現されている例である。例えば、「名前」と「山田太郎」が属性・属性値関係にあることが、文字列の2次元的な配置と、それを修飾する色や記号等によって表現されている。このような性質を持つ Web 文書を解析するためには、従来の言語処理とは異なる手法が必要となる。

本研究では、Web 文書に、階層的な文書構造の存在を仮定し、ヘッダツリーと名付ける。ヘッダとは、「タイトル」「見出し」等の概念を一般化したものであり、「文書中の他の部分を修飾する部分文書」と

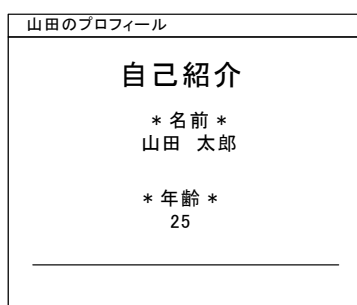


図 5 : Web 文書の例

して定義される。例えば、図 6 のような文書においては、「自己紹介」(文書全体を修飾)「名前」(文字列「山田太郎」を修飾)「年齢」(文字列「25歳」を修飾)等の文字列がヘッダと見なされる。ヘッダツリーとは、ヘッダを内部ノードとして構成される木構造であり、例えば図 5 のページのヘッダツリー

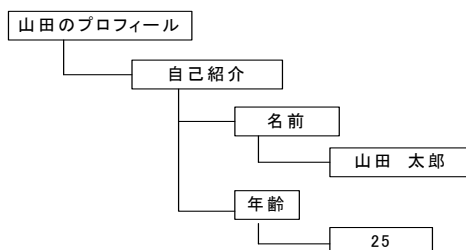


図 6 : ヘッダツリーの例

は、図 6 のようになる。

3.2 内容

本研究では、このヘッダツリー抽出において、自然言語処理等で用いられる構文解析技術を応用

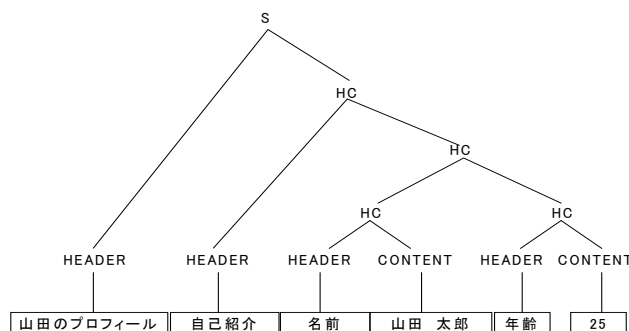


図 7 : 構文解析木の例

することで、汎用性の高いアルゴリズムの開発を目指した。文法枠組として確率的文脈自由文法 (PCFG) を用い、Web 文書中の文字列を構文解析することにより、ヘッダツリーの抽出を行う。図 7 に、図 1 の Web 文書に対する構文解析木を示す。このとき、PCFG の確率パラメータ推定の際に、色や位置等のレイアウト情報を制約として加えることにより、言語的な情報と、レイアウト情報を、自然な形で統合することが可能になる。

PCFG のルールとしては、例えば、”HC→HEADER CONTENT”(「ヘッダとその修飾先 (コンテンツ) が隣り合っている時、それらは非終端記号 HC から生成される」)、「HC→HC HC”(「非終端記号 HC が隣り合っているとき、それらは別の非終端記号 HC から再帰的に生成される」) 等が定義されている。

$$\arg \max_T P(T|\mathcal{B}, \mathcal{L}) = \arg \max_T P(T)P(\mathcal{B}|T)p(\mathcal{L}|T)$$

Web 文書に対する最適な構文解析木は、確率モデルに基づき、

に基づいて推定される。ここで、 T は構文解析木であり、隠れ変数となる。 $P(\mathcal{B}|T)$ は文字列生成確率であり、言語的情報を表すパラメータとなる。この $P(\mathcal{B}|T)$ は、EM アルゴリズムを用いることにより推定される。

$P(\mathcal{L}|T)$ はレイアウト生成確率であり、「並列な構造は、類似の見た目を持ちやすい」「ヘッダとその修飾先の間は、空白が少ない」等、レイアウトに関するヒューリスティクスに基づいて値を決定する。

3.3 具体的成果

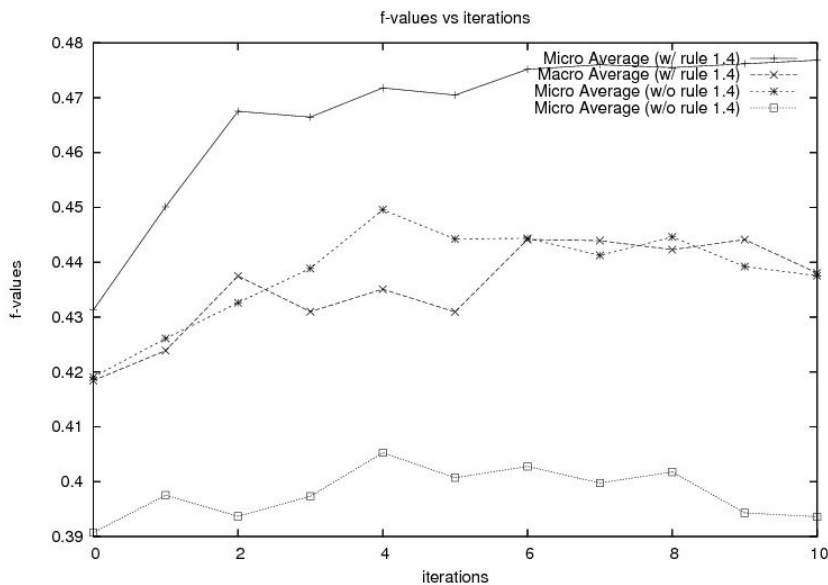


図 8 : パラメータ推定の繰り返し回数による、精度の変化。

評価実験には、WWW からランダムに収集した Web 文書のうち、1,000~10,000 バイトの比較的小さなサイズの文書を 50 個選び、そのヘッダツリーを人手でタグ付けした正解データを用いた。評価指標としてはマイクロ平均精度、マクロ平均精度を用い、それぞれ 0.477, 0.438 の精度を得た。EM アルゴリズムによるパラメータ推定を行わない場合の精度はそれぞれ 0.431, 0.418 であり、パラメータ推定の有効性が確認された (図 8)。

4 成果要覧

査読付論文リスト

- [査読付 1] Minoru Yoshida, Hiroshi Nakagawa, and Akira Terada: Gram-Free Synonym Extraction via Suffix Arrays, Proceedings of the Fourth Asia Information Retrieval Symposium (AIRS 2008), pp.282-291, 2008, LNCS 4993
- [査読付 2] Minoru Yoshida and Hiroshi Nakagawa: Web Document Parsing: A New Approach to Modeling Layout-Language Relations, Proceedings of the 9th International Conference on Document Analysis and Recognition (ICDAR2007), pp.203-207, 2007

その他の発表論文リスト

- [発表 1]吉田 稔, 中川裕志: UT-Kiwi: 検索支援としてのテキストマイニングシステム, 情報処理学会第70回全国大会, 4J-4, 筑波, 2008年3月
- [発表 2] 清田 陽司, 阿辺川 武, 吉田 稔, 田村 悟之, 坂井 哲, 増田 英孝: 汎用シソーラス探索ライブラリの開発, 言語処理学会 第14回年次大会 発表論文集, PD1-3, pp. 257-260, 東京, 2008年3月
- [発表 3] Yuxin WANG, Nobuyuki SHIMIZU, Minoru YOSHIDA, Hiroshi NAKAGAWA: Contextual Information Based Technical Synonym Extraction, B4-4, pp.761-764, 東京, 2008年3月
- [発表 4] 杉浦 隆博, 吉田 稔, 山田 剛一, 増田 英孝, 中川 裕志: 機械学習による新聞記事からの数値とその属性名の抽出, MuST(動向情報の要約と可視化に関するワークショップ) 2007年度成果進捗報告会, 東京, 2008年3月
- [発表 5] 廣川 敬真, 吉田 稔, 山田 剛一, 増田 英孝, 中川 裕志: 新聞記事のテキスト情報と株価動向の関係の解析, MuST(動向情報の要約と可視化に関するワークショップ) 2007年度成果進捗報告会, 東京, 2008年3月
- [発表 6] 小野真吾, 佐藤一誠, 吉田稔, 中川裕志: 重要語抽出を用いた Web 文書上の同姓同名の曖昧さ解消, 電子情報通信学会 第19回データ工学ワークショップ(DEWS 2008), A7, 宮崎, 2008年3月
- [発表 7] 寺田昭, 吉田稔, 中川裕志: 計算機支援による同義語辞書作成ツール, 第45回飛行機シンポジウム, 3D7, 小倉, 2007年10月
- [発表 8] 前島一弥, 横川智浩, 山田剛一, 絹川博之, 吉田 稔, 中川裕志: Web 上の表情報を対象とした例示検索 ～表の構造的特徴の利用～, 第6回情報科学技術フォーラム(FIT2007), D-046, 愛知, 2007年9月
- [発表 9] 杉浦隆博, 吉田 稔, 山田剛一, 増田英孝, 中川裕志: 数値による新聞記事テキストマイニングシステムの提案, 第6回情報科学技術フォーラム(FIT2007), E-013, 愛知, 2007年9月
- [発表 10] 杉浦隆博, 吉田 稔, 山田剛一, 増田英孝, 中川裕志: 新聞記事の数値による情報検索システムの提案と実装, 第21回人工知能学会全国大会, 2H5-9, 宮崎, 2007年6月

公開ソフトウェア

- [公開ソフト1] 吉田稔, 中川裕志: UT-Kiwi, <http://kiwi.r.dl.itc.u-tokyo.ac.jp/ut-kiwi/>

図書館情報資源と Web 情報資源の融合

～Web 2.0 時代の電子図書館の実現に向けて～

清田 陽司

1 概要

人が情報検索を行うプロセスは、人の持っている情報要求が徐々に具体化するプロセスとしてとらえることができる。このプロセスを Taylor は以下の 4 階層に分類している¹。

(Q1)直感的要求(visceral need):漠然と知識が欠けていることを認識しているが、どんな知識が欠けているかはわからない

(Q2)意識された要求(conscious need):どんな知識が欠けているかはほぼわかるが、どんな情報が必要かは明確に説明できない

(Q3)形式化された要求(formalized need):欠けている知識が明確にわかり論理的な質問ができる

(Q4)調整済みの要求(compromised need):検索を行う対象(質問を投げかける相手や検索システム)の知識構造や理解力を想定し、相手の制約を認識して検索を行うことができる

一方、情報検索を工学的アプローチによって研究する立場からは、さまざまな情報検索システムが提案されてきているが、ほとんどの研究は Taylor の階層における Q4 を前提としており、Q1～Q3 は対象とされていない。テキスト検索システムに関する評価型ワークショップである TREC などでは共通の検索課題を設定するが、評価の客観性を確保する必要性から、課題として設定される検索質問は必然的に Q4 のレベルのものとなる。また、ユーザインタフェースの研究分野では Q3 から Q4 へのプロセスが扱われている。例えば、利用者の入力した質問キーワードに関連するキーワードの提示についての研究などがある。

しかし、Q1～Q3 のプロセスをコンピュータシステムで扱う研究はほとんど存在しない。その大きな理由としては、情報要求が具体化するプロセスを形式化するのが困難だということ、プロセスを客観的に評価する方法論が確立していないことがあげられる。しかし、人が必要な情報を得られないのはそもそも Q1～Q3 の階層から Q4 に具体化できないのが原因となっていることが多い。

では、必要な情報を効率的に見つけられる人はどのようにして「情報要求の具体化」を行っているのであろうか？ この命題を考える際に、文献探索の専門家である図書館員が行っている情報探索支援行動が参考になる。一般に、図書館のレファレンスサービスにおける質問回答の際、図書館員は利用者の調べたい主題や探索事項が不明確な場合にさまざまな角度からインタビューを行い、徐々に利用者の情報要求を明確にしていく。さらに主題や探索事項に関連するさまざまな情報資源に関する豊富な知識を活用することによって、利用者に適切な資料を提示することができる。

コンピュータシステムによって図書館員の情報探索支援行動を模倣するには、多様な情報資源を活用するための仕組みを持たせる必要がある。世の中に存在する情報資源のなかでも重要な役割を果たしているのは、図書館情報資源と Web 情報資源である。しかし、図書館情報資源には情報の新鮮度・カバレッジに問題があり、Web 情報資源には情報の信頼性・専門性に問題がある。この間

¹ Robert S. Taylor. Question-Negotiation and Information Seeking in Libraries. *College and Research Libraries*, 29(3):178-194, 1968.

題を解決するには両者を統合して組織化・利用することによってお互いの欠点を補完することが必要である。本研究では、図書館において専門家によって厳密に構成されたトップダウン的な分類体系、Webにおいて多数の参加者によって構築されたボトムアップ的な分類体系をうまく組み合わせることで、この問題を解決する。

今年度は、利用者の「情報要求の具体化」をシステムとのインタラクションによって実現するため、以下に挙げる3つのトピックに取り組んだ。

- **Wikipedia を活用した図書館件名標目の自動導出** [発表 1][受賞 1][報道 1][報道 2][報道 3]
 図書館において行われているレファレンスサービスを補完し、利用者自身が多様な情報資源(図書館や Web)を使いこなすことを支援するシステムの実現を目指している。Web の分類体系と図書館の分類体系を統合するにあたって Wikipedia を活用する。今年度は、Web 上の分類体系と図書館分類体系を統合的に活用する情報探索システムを実際に図書館の現場にて運用し、ログを収集した。
- **テーマグラフ可視化インタフェースの実装** [発表 2]
 図書館員や利用者からのフィードバックを踏まえて、テーマグラフ可視化インタフェースの実装を行った。
- **汎用シソーラス探索ライブラリの開発** [発表 3]
 図書館や Web に存在する多様な分類体系を効率的に活用するには、統一的にそれらの分類体系にアクセスできるライブラリがあることが望ましい。そこで、よく利用される分類体系に共通する部分を抽出し、頻繁に活用される処理をパッケージ化したライブラリを開発した。

2 Wikipedia を活用した図書館件名標目の自動導出

2.1 背景

効率的な情報探索を行うには、膨大な情報資源を整理・組織化している各種の情報探索ツールを活用することが不可欠である。インターネットが一般に普及する以前、情報探索を行うにあたって中心的な役割を果たしていたのは図書館であった。図書館には、膨大な資料を整理・組織化し、利用者の情報探索に役立てるためのツールとしての分類体系が存在している。一方近年では、Web 技術を活用することで膨大な知識を多人数の共同作業によって組織化し、効率的な情報探索に役立てようという取り組みが盛んに行われるようになってきている。その代表例としては、Wikipedia²や、各種の Folksonomy サービス(Flickr³など)があげられる。その結果として、Web 上に一種の情報探索用オントロジーとよべるものが存在するという状況が生まれつつある。

情報探索ツールとしての図書館分類体系と Wikipedia の特徴を整理すると、下記ようになる。

- 図書館分類体系
 専門家による慎重な作業を経て**トップダウン的**に構築されていることから、安定的に利用することができる、深い組織化がなされているという長所がある。また、特に BSH、LCSH(米国連邦議会図書館件名標目表)、NDLSH(国立国会図書館件名標目表)などの各種件名標目表は、情報探索に拡がりをもたせるためにきわめて有用なツールである。反面で、改定の周期が長いと新しくできた概念には対応できない、一般的に上位概念は 1 つしかもてないため概念の多様な側面を反映しづらいという欠点をもっている。
- Wikipedia

² <http://ja.wikipedia.org/>

³ <http://www.flickr.com/>

Web 上で Wiki システムによって共同編集されているオンライン百科事典であり、誰でもどこからでも自由に編集することができる。現在、200 以上の言語で作成されており、日本語版では 30 万項目以上の記事が存在する。多数の人が編集に参加していることから、世界中のほとんどの概念を網羅していると思われる。また、ハイパーリンクを容易に張ることができるという Wiki の特徴を生かし、項目の組織化の取り組みも行われている。膨大な項目から効率的に一覧を生成するために「Wikipedia カテゴリ」という**ボトムアップ的**な組織化の仕組みが導入されており、項目を多様な観点から分類することが可能となっている。例えば、「価格」という項目には「マーケティング」「経済学」「市場」といった複数のカテゴリを付与することができる。また、カテゴリ自体にも上位概念(親)のカテゴリを付与することが可能である(図 1)。複数の親をもたせることができるのは、従来の分類体系にはない大きな特徴である。反面で、頻繁な編集が行われるため体系が安定していない。信頼性に欠ける情報も少なくないといった欠点がある。

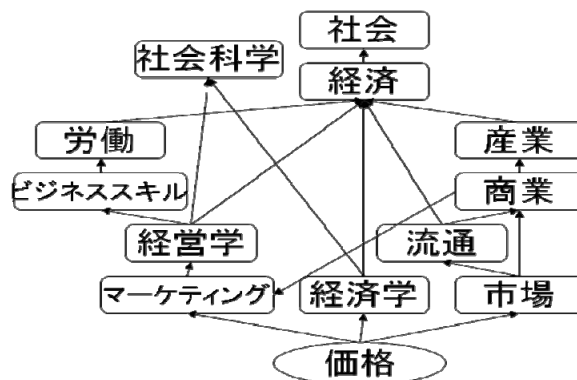


図 1: Wikipedia のカテゴリ体系

2.2 内容

今年度は、2006 年度に開発した Reference Navigator の成果を踏まえ、Wikipedia カテゴリ体系と図書館分類体系を高精度で対応づけ、さまざまなユーザ質問に対して一般的な概念を効率的に提示する手法を研究した。その結果、Wikipedia に含まれる 40 万以上の項目に対して、高精度で件名標目・図書分類コードを導出可能であることが確認された。Wikipedia の項目「阪神・淡路大震災」に対して本手法を適用した様子を図 2 に示す。

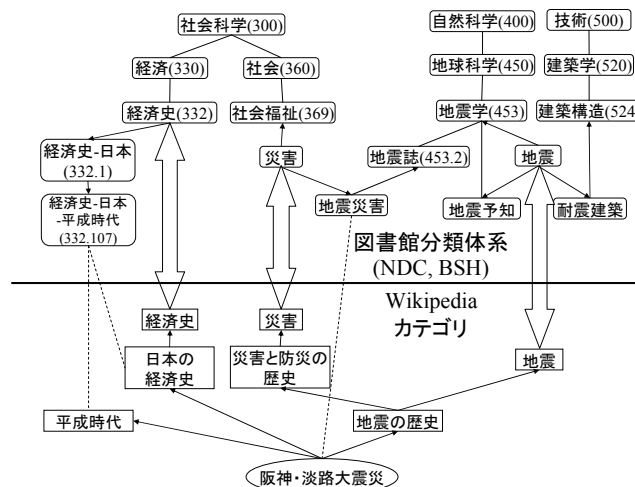


図 2: Wikipedia を利用した自動的な件名標目・図書分類コードの導出

さらに、本手法を図書館の現場において情報探索に利用できるように、対話型 Web 検索システム「Littel Navigator」を実装した。Littel Navigator のユーザインタフェースを図 3 に示す。利用者は、調べたいことがらに関連していることばをキーワード入力欄に入力することで情報探索を開始できる。例えば、「阪神・淡路大震災」を入力すると、システムは「地震」「経済」「地震学」「災害」などのテーマを自動的に導出して表示する。さらに、それぞれのテーマが導出されたパス(経路)や、それぞれのテーマに関連する書架分類コード(NDC)、有用な情報リソース、関連キーワードなどが表示される。たとえばテーマ「地震」に関連する書籍を探したければ、NDC コード「453」をクリックすることで、蔵書検索システム(OPAC)の検索結果を閲覧することができる。

Littel Navigator は「図書館と Web の長所を活かせる検索システム」を開発コンセプトとし、曖昧な質問(キーワード)から具体的な事項へ誘導し、自動レファレンス応答の実現を図ることを目標として開発している。また、Littel Navigator は「情報探索のヒント」検索システムという側面も持たせており、「パスファインダー自動生成システム」と言い換えることもできる。パスファインダーとは、特定の主題やトピックに関する資料・情報を探す際に、図書館が提供できる関連資料をリストとして提供するサービスのことであり、各図書館が Web サイトなどを用いて広く公開しているのが特徴である。この特徴によって、さまざまな情報リソースの統合的利用や、曖昧な質問から具体的な探し方へのナビゲーションが可能となっている。

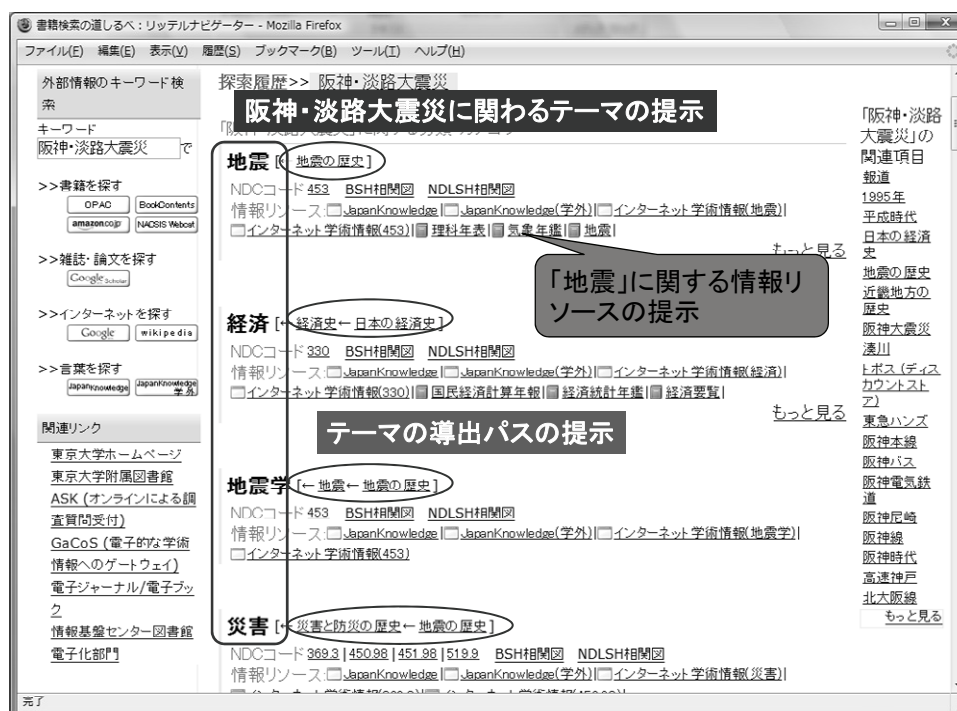


図 3: Littel Navigator のユーザインタフェース

2.3 具体的成果

Littel Navigator は、2007 年 5 月より東京大学附属図書館などにおいて公開運用を開始している。東京大学キャンパス内では、<http://u-tokyo.navi.littel.jp/> にアクセスすることによって利用可能である。また、Littel Navigator の技術をベンチャー企業にライセンスし、国内外の大学図書館や公共図書館にサービスを提供していくという取り組みを行っている。この件に関しては日本経済新聞[報道 1]、東京大学新聞[報道 2]、日経産業新聞[報道 3]などで報道された。また、Littel Navigator のコンセプトについて NLP 若手の会第 2 回シンポジウムにおいて発表し[発表 1]、奨励賞を受賞した[受賞 1]。

3.3 具体的成果

本研究は、Littel Navigator の図書館の現場での運用過程において、複数の図書館員から「テーマ同士のつながりを調べたい」という要望が寄せられたことがきっかけとなっている。研究成果については、情報処理学会第 70 回全国大会にて発表した[発表 2]。今後、テーマグラフ可視化インタフェースを改良し、Littel Navigator のメニューとして組み込むことを計画している。

4 汎用的な分類体系探索ライブラリの開発

4.1 背景

グラフ構造を持った言語資源は自然言語処理研究のあらゆる場面で活用されている。代表的な例としては、NTT 日本語語彙大系などの木構造をもつシソーラスを用いた語の類似度計算などがある。さらに最近では Wikipedia など、ページ間の構造が巨大なグラフ構造をなして一種のシソーラスとして扱える言語資源が登場し、さまざまなアプリケーションで活用されるようになってきている。その結果、自然言語処理研究においてはさまざまな種類のシソーラスを使いこなすことが必要とされる状況が生まれている。

シソーラスを扱うためのコードは、自然言語処理の研究者が各自書いていることが多い。言い換えると、個々の研究者が「車輪の再発明」をしている状況である。もちろん、グラフ構造の扱い自体を研究対象とする場合などは一からコードを書くことが必要とされる。しかし、さまざまな自然言語処理アプリケーション(e.g. 語義曖昧性解消、情報検索、機械翻訳)においてシソーラスを利用する場合にまでコードを一から書く必要に迫られることは、自然言語処理アプリケーションの研究を始めるハードルを過度に高くしている可能性がある。このハードルを下げるためには、典型的なシソーラス処理をパッケージ化した共通基盤としてのライブラリが存在することが理想である。しかし、現時点では広く普及しているライブラリは存在しない。これまでシソーラスを扱うライブラリが普及しなかった原因は以下のように整理できる。

- プログラミング言語の壁

研究に利用するプログラミング言語は、用途や個人の好み、研究組織ごとに蓄積されているライブラリなどによって左右されるため、標準言語を定めることは難しい。また、プログラミング言語の壁を超えて汎用的に利用されるグラフ探索アルゴリズム実装は存在しなかった。

- データ構造の標準階層モデルが存在しない

汎用的なデータ構造を実現するには、機能に応じて階層的に分割された標準的なモデルがあることが望ましいが、シソーラスのデータ構造についてそのようなモデルは存在しなかった。

- シソーラスデータ流通に伴う問題

シソーラスの構築には多大なコストがかかるため、有料で入手する必要があるものも少なくない。また、シソーラスデータにアクセス用ライブラリが添付されていることはほとんどない。それらのシソーラスへのアクセス用ライブラリを利用者がボランティアで開発・公開しているケースもあるが、様々なシソーラスを網羅できるようなライブラリを一利用者が開発することは、検証に必要なデータの入手にかかるコストを考慮すると著しく困難である。

4.2 内容

自然言語処理研究のハードルを低くし、研究分野の裾野を広げるため、頻繁に利用されるシソーラス探索処理をパッケージ化したライブラリの可能性を検討し、以下の戦略によってライブラリを実装した。

- 研究に利用するプログラミング言語は、用途や個人の好み、研究組織ごとに蓄積されているライブラリなどによって左右されるため、標準言語を定めることは難しい。また、プログラミング言語の壁を超えて汎用的に利用されるグラフ探索アルゴリズム実装は存在しなかった。基本機能は C++言語で実装し、他のプログラミング言語へのラッパーとして SWIG (Simplified Wrapper and Interface Generator)を採用する。これによって、自然言語処理分野の研究者に利用されている Perl, Ruby, Java, Python などの諸言語に対応する。また、汎用的なグラフ探索アルゴリズム実装である The Boost Graph Library を採用する。これによって、必要なグラフ探索アルゴリズムを利用できるようにする。グラフ探索にはエッジに付与された重みを考慮する。
- シソーラスのデータ構造を、redirect, leaf, node の 3 階層(図 6)からなるグラフとして表現する。これによって、多様なシソーラスに対応するための柔軟性を確保する。利用者は、シソーラスのグラフ構造を表現するために、2 つのテキストファイル「edge ファイル」(図 7)と「node ファイル」(図 8)を用意するだけでよい。主なシソーラスに対する適用例を表 1 に示す。
- 大部分のシソーラスは知的財産として扱われるためライブラリと同時に配布することはできないが、主要なシソーラスについては変換スクリプトを用意しておくことで、シソーラスのデータファイルさえ入手すれば容易にライブラリを利用できるように配慮する。

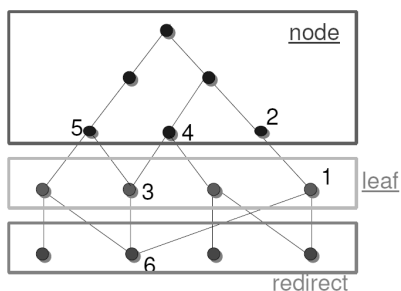


図 6: シソーラスの 3 階層モデル

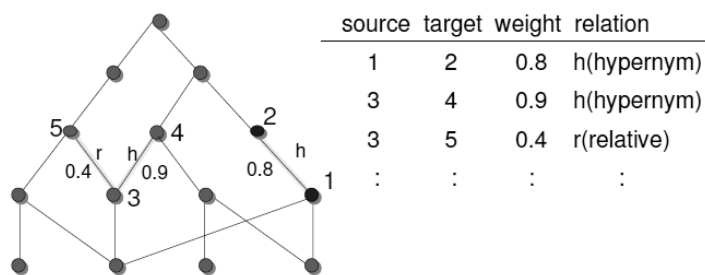


図 7: edge ファイル

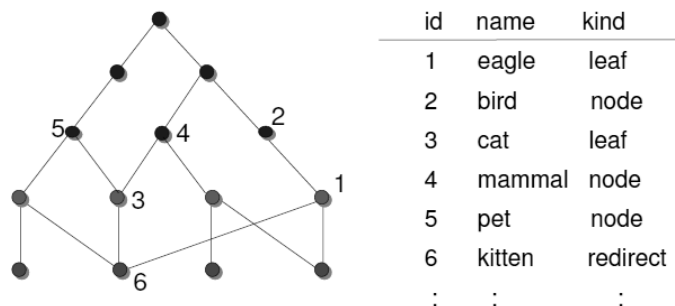


図 8: node ファイル

表 1: 各種シソーラスにおけるマッピング

シソーラス	edge ファイル	node ファイル
NTT 日本語彙大系, EDR, 分類語彙表	概念間関係, 単語 → 概念	単語 (leaf), 概念 (node)
WordNet	synset 間関係 (同意, 反義, ...), 単語 → synset	単語 (leaf), synset (node)
Wikipedia	カテゴリ間関係, 項目 → カテゴリ, リダイレクト・曖昧性回避 → 項目	リダイレクト, 曖昧性回避 (redirect), 項目 (leaf), カテゴリ (node)
図書件名標目表 (LCSH, BSH), MeSH	件名間の関係 (上位下位, 関連語など)	件名 (node)
2 言語間辞書	source 言語の単語 → target 言語の単語	source 言語の単語 (redirect), target 言語の単語 (leaf)
表記揺れ辞書	表記揺れを含む単語 → 代表単語	表記揺れを含む単語 (redirect), 代表単語 (leaf)

グラフ探索系 API

木構造グラフ用

get_parent (ノードの親ノード)
 get_child (ノードの子ノード)
 get_depth (ルートノードからの深さ)
 calc_similarity (2 ノード間の類似度)
 print_tree (ノードツリーの表示)

汎用

get_shortest_path (2 ノード間の最短パス)
 (breadth first, dijkstra)
 get_distance (2 ノード間の距離)

ノード参照系 API

ノードの情報取得

get_name (ノードの名称)
 get_kind (ノードの種類 redirect/leaf/node)
 is_redirect, is_leaf, is_node
 (ノードの種類を true/false で判別)

ノード名の検索

exact_match (完全一致)
 search_forward (前方一致)
 search_backward (後方一致)

図 9: 汎用シソーラス探索ライブラリの主な API

ライブラリの関数群(API, Application Programming Interface)は、グラフ探索系 API とノード参照系 API に大きく分けられる。実装済みの主な API を図 9 に示す。これらの API は、The Boost Graph Library を利用して実装されている。API を活用することにより、広範な種類のシソーラスを活用したアプリケーションを容易に構築することが可能になる。

4.3 具体的成果

上記の方針にもとづいて汎用シソーラス探索ライブラリの試作を行い、現在 WordNet、Wikipedia、件名標目表などのグラフ探索ができるようになってきている。2008 年 3 月の言語処理学会第 14 回年次大会にて、汎用シソーラス探索ライブラリに関するプレゼンテーションを行った[発表 3]。現在、2008 年度上半期の一般公開に向けて準備を進めている。

5 成果要覧

受賞関連

[受賞 1] 清田 陽司, 田村 悟之, 坂井 哲: 奨励賞, NLP 若手の会第 2 回シンポジウム, 2007 年 9 月

その他の発表論文

[発表 1] 清田 陽司, 田村 悟之, 坂井 哲: なんでもテーマ検索: Wikipedia の項目から情報探索ソースへの誘導, NLP 若手の会 第 2 回シンポジウム, 東京大学本郷キャンパス, 2007 年 9 月

[発表 2] 坂井 哲, 清田 陽司, 増田 英孝, 中川 裕志: 図書館と Web の分類体系を統合的に活用したテーマグラフ可視化インタフェース, 情報処理学会第 70 回全国大会講演論文集, 4ZK-9, 筑波大学, 2008 年 3 月

[発表 3] 清田 陽司, 阿辺川 武, 吉田 稔, 田村 悟之, 坂井 哲, 増田 英孝: 汎用シソーラス探索ライブラリの開発, 言語処理学会 第 14 回年次大会 発表論文集(PD1-3), pp. 257-260, 東京大学駒場キャンパス, 2008 年 3 月

報道関連

[報道 1] 日本経済新聞 2007年(平成 19年) 4月 18日朝刊 ベンチャー面: 図書館で本探し キーワード検索ーリッテルがシステム

[報道 2] 東京大学新聞 2007年 10月 16日号: 特集 進化する書籍情報の電子化

[報道 3] 日経産業新聞 2007年(平成 19年) 11月 6日 20面: 点検 大学発 VB(東大)リッテル 図書館の本 分類表示 ～言語処理技術を活用～

知識発見と機械翻訳の自然言語処理要素技術

—同義語獲得、係り受け解析、

機械学習テンプレートライブラリ—

清水伸幸

1 概要

本研究の最終的な目的は、英語、日本語など、自然言語で書かれたテキスト情報から、有用な知識を効率的に発見するための自然言語処理の要素技術を、最新の機械学習の機械学習技術を用いて発展させることにある。具体的には、文書からユーザが必要とする意味関係の抽出(テキストマイニング)、機械翻訳に有用な二つの基本的な要素技術と、これらの成果を一般に公開するためのテンプレートライブラリに関する研究を行っている。

- 同義語獲得

大量の文書からユーザが必要とする事例の統計を取るには、まず同義語を認知することが重要である。ここでの事例とは、テキストマイニングにおいては記事に記されたイベントにかかわる事例、翻訳においては翻訳事例である。機械学習による同義語類似指標の学習、類似単語ネットワークを通じた同義語獲得を目指す。

- 係り受け解析

同義語獲得と同様、テキストマイニング、機械翻訳においても、文章の係り受け関係の解析は重要な基礎技術である。上記の同義語獲得においても、最も性能が高い同義語獲得手法は係り受け関係を素性として利用する。そこで、近年発明された、非交差性を持たない言語における係り受け関係の構造付き分類学習器を、違うドメインに適応する手法と組み合わせて精度の向上を目指す。

- 機械学習テンプレートライブラリ

上記の成果を社会にフィードバックするため、他の機械学習手法も含めた構造付き学習器のテンプレートライブラリを開発、一般に公開することを目指す。

2 同義語獲得

2.1 背景

現在、Wordnetなどのシソーラスは、テキスト処理の効率と性能を向上させるための最も重要なリソースの一つであり、さまざまなタスクで利用されている。その重要性は、特定のドメインにおける高品質のシソーラスがなければ、例えば事故報告コーパス中の重大事件を過大、あるいは過小に評価する可能性があることから分かる。ただし、対象にしているドメインにおいて、自動的に同義語を取得するのは簡単ではない。以下に一般的な同義語獲得の手法を述べる。通常、同義語獲得は、同義

語であれば似た文脈情報を持つという分布仮説(Harris 1985)に基づいて行われる。まず、ターゲット単語それぞれについて、コーパスから抽出された重要度の高い単語の文脈特徴（通常は係り受け関係）の統計情報を抽出し、文脈特徴のベクトルによって表す。次に、cosine 類似度などの類似性量度を選び、それをクエリ単語と同義語候補の単語対に適用して、類似度を計算する。類似度の降順に各クエリ単語の同義語候補リストを作る。最終的には、同義語リストからトップ候補を選んで、クエリ単語の同義語と認定することで同義語獲得が終了する。

2.2 内容

本研究テーマでは、単語対について類似度を計算したのち、単語をノードとし、同義語候補対を類似度で重みづけしたエッジを持つ単語類似度ネットワークを解析して、同義語抽出に役立てる手法を提案した。このネットワークにはスケールフリー性があり、ハブとなる単語をクエリとしてすると、そうでない場合と比べ同義語獲得が難しいということが判明した。この関係を図 1 に示す。図を見るとわかるように、どちらの単語をクエリとしてみるかによって、お互いの同義語候補としてのランクが異なる。そういった場合には、ハブである単語をクエリとして、ハブでない単語へのランクより、逆からみたランクのほうが信頼性が高い。

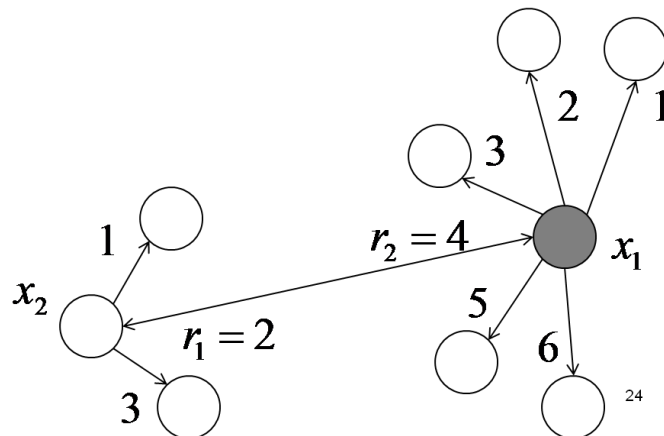


図 1: ハブとそうでない単語対の相互から見たランク（ハブは灰色）

このため、ターゲットとしてクエリ単語が与えられた際、候補語のランクをそのまま使うのではなく、候補語をクエリとしてターゲット語のランクを利用し、元の候補語を再ランクするスコア手法を提案した。これにより、より同義語としてよい単語を獲得することに成功した。

2.3 具体的成果

手法の評価実験として、JAL-FDM レポート コーパス (Flight Data Monitoring Report of Japan Airline) を用いて、専門用語の同義語の獲得を行い、すべての素性のケースにおいて、提案手法が優れた同義語候補を抽出することを人手による正解データに照らし合わせて確認し、国内学会 [発表 2] と国内の機械翻訳関係者が集まるセミナー [特記 1] において発表した。

3 係り受け解析

3.1 背景

他言語への翻訳や意味解析をおこなうにあたり、文章の係り受け関係を正しく解析することは重要である。このため、英語においては、Penn Treebank とよばれる、比較的大規模な係り受け関係を人手でつけた正解データが存在する。しかし、このコーパスは Wall Street Journal の新聞記事に対してつけられたものであり、このコーパスで学習した係り受け解析器を、生物の論文誌など、他の分野に適応すると性能が劣化することが知られている。この性能の劣化を避けるため、近年、いくつかのドメイン適応の手法が考案されている。本研究では、特に、係り受け解析のドメイン適応問題について最も効果的とされる手法を試みた。

3.2 内容

背景となる係り受け解析器のドメイン適応問題に対して、本研究テーマでは、Structural Correspondence Learning (SCL)と呼ばれる手法を用いて、大規模な教師なしテキストを利用したドメイン適応の実験を行い、係り受け解析での性能を競う CoNLL 2007 Shared Task に参加した。SCL とは、以下のような手順を経て教師なしテキストからドメイン適応に役立ちそうな素性を作りだし、本来のタスクにおいて使う手法である。

- (A) 教師なしテキスト、教師ありテキストの両方から、与えられた教師ラベルを使わずに、ドメイン適応に役立ちそうなサブタスクとなる分類問題をいくつか作り出す
- (B) 線形分類器でサブタスクを学習する
- (C) 学習したクラス毎の重みベクトルを一つの行列にまとめ、Singular Value Decomposition (SVD) をかける
- (D) SVD の結果を本来の教師あり学習の素性として付け加える

SCL は、現時点で最も効果的なドメイン適応手法の一つとして知られている。それにもかかわらず、残念ながら、実験の結果、過学習が認められ、SCL はそれほどの効果を発揮していないことが判明した。しかしながら、この Shared Task においては、一つのグループを除いて、他のすべてのグループがドメイン適応に失敗するという結果に終わった。主な原因としては、適応先の正解データのアンノテーション・ガイドラインが、適応元のアンノテーション・スキームと若干異なることにあることが明らかになり、ドメイン適応の研究のタスク設定の難しさが示される結果となった。

3.3 具体的成果

本研究は Shared Task の評価の一部として [査読付 1] において掲載された。

4 機械学習テンプレートライブラリ

4.1 背景

近年の自然言語処理では、多くの場面で機械学習を用いることが必要となってきた。一方で、言語の構造をより適切にモデル化した機械学習手法がさかんに研究されている。これらの手法は様々な自然言語処理タスクにおいてその有用性が示されているが、より複雑なモデルになるにつれて実装は煩雑になり、また学習手法間の理論的・経験的關係も把握しづらくなってきている。例えば、品詞タグ付けなどの系列ラベル付け問題に対しては、以下のような機械学習手法が提案されている。

- Conditional Random Fields (CRFs)
- Max-margin Markov Network (M3N)
- Structured Perceptron

これらの手法は、共通のデータ構造(マルコフ連鎖)を対象としている。一方、損失関数はそれぞれ異なり、そのため最適化アルゴリズムも異なっている。このように、共通のデータ構造に対して様々な学習手法が提案されているという状況は、より複雑なデータ構造が必要な構文解析においても同様である。そこで、自然言語処理へ機械学習を応用する研究や、より自然言語処理に適した機械学習手法の開発を容易にするため、機械学習ライブラリ的设计・開発を行った。

4.2 内容

自然言語処理における機械学習の研究では、ほぼ同じデータ構造で、損失関数や最適化アルゴリズムが異なる手法が多く提案されている。本研究では、以下のような式で識別関数を定義する機械学習モデルを対象とする。

$$\eta(y|x) = f(w \cdot \Phi(x, y))$$

ここで、 x, y は履歴とターゲット、 Φ は素性関数、 w はパラメタ(重み)ベクトル、 f は任意の関数である。例えば、文書分類では x は文書で y はラベル、品詞タグ付けでは x は文で y は品詞列、係り受け解析では x は文で y は係り受け木となる。現在自然言語処理で利用されている多くの機械学習手法は、上記のように定式化することができる。以上のような状況を鑑み、機械学習テンプレートライブラリは以下のような目的を持って設計された。

- 同じデータ構造について、様々な損失関数や最適化アルゴリズムを容易に比較できる。
- 木構造や連鎖構造など、言語の構造を表現するのに適したデータ構造を様々な機械学習アルゴリズムで利用できる。
- データ構造に依存する計算(デコード、期待値計算など)と、学習アルゴリズム(損失関数、最適化アルゴリズム)を切り離し、どちらを改良しても残りの部分を利用可能にする。
- 新たなデータ構造(または学習器)を実装すれば、様々な学習器(またはデータ構造)との組み合わせを容易に試すことができる。
- 既存の機械学習ソフトウェアと遜色ない時間・メモリ効率性を目指す。

これらの要求を満たすため、データ構造と学習最適化をテンプレートにより抽象化し、間にインターフェイスを定義することで、メモリ、効率性を損なわずに様々な学習器の組み合わせを扱えるようにした。具体的には、学習器は、データ構造クラスに定義されているいくつかのメソッドを利用して必要な統計量を計算し、パラメタを更新する。図2に、主な学習器と必要なメソッドの組み合わせを示す。

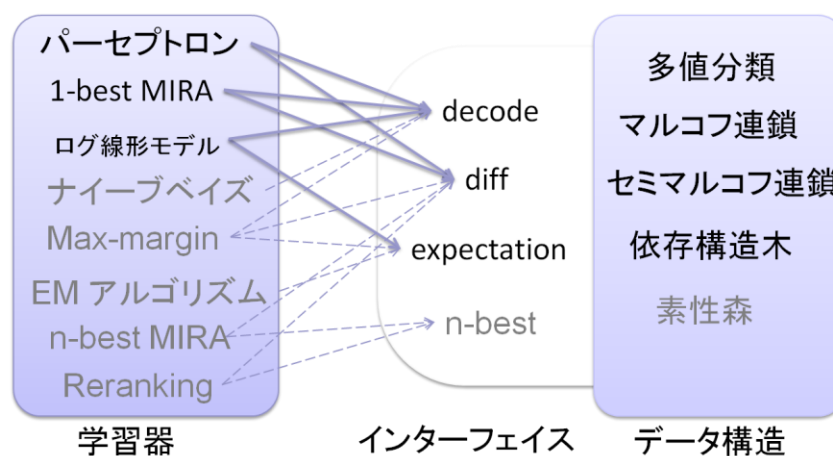


図2: 学習器とデータ構造間のインターフェイスの相関図

学習器が利用する主なメソッドは以下のとおりである。

- デコード (decode)

$\eta(y|x)$ を最大にする y を求める。パーセプトロンや MIRA の学習などで用いられる。ビタビアルゴリズムなどを用いる。
- 差分 (diff)

二つのターゲット y_1, y_2 の差分を計算する。パーセプトロンや MIRA の学習で用いられる。
- 期待値 (expectation)

$\Phi(x, y)$ の $\eta(y|x)$ による重みつき平均(期待値)を計算する。ログ線形モデルや max-margin モデルの学習で用いられる。前向き・後向きアルゴリズムなどを用いる。

以上のような方法で、データ構造およびそれに付随するアルゴリズムと、機械学習の損失関数および最適化アルゴリズムを切り離すことで、様々な構造化機械学習手法を見通しよく実装した。

4.3 具体的成果

機械学習テンプレートライブラリは、ソースコードと、典型的なデータ構造・学習器についてバイナリパッケージを公開する予定である。いくつかの構造については実装が進んでおり、依存構造木では CoNLL2007 Shared Task の英語コーパスにおいて、パーセプトロン学習 (10 iterations) が4時間ほどで終了することを実証し、ライブラリの有用性を確認した。本ライブラリは、国内学会 [発表 2] において発表した。また、[査読付 1] において使用された依存構造木の係り受け解析器はこのライブラリを使用している。

5 成果要覧

査読付論文

[査読付 1] Nobuyuki SHIMIZU and Hiroshi NAKAGAWA: Structural Correspondence Learning for Dependency Parsing, *Proceedings of the CoNLL Shared Task Session of EMNLP-CoNLL 2007*, pp. 1166-1169, 2007.

その他の発表論文リスト

[発表 1] 清水伸幸, 宮尾祐介: 機械学習テンプレートライブラリ, 言語処理学会 第 14 回年次大会 発表論文集, pp.245-248, 2008.

[発表 2] Yuxin WANG, Nobuyuki SHIMIZU, Minoru YOSHIDA and Hiroshi NAKAGAWA: Contextual Information Based Technical Synonym Extraction, 言語処理学会 第 14 回年次大会 発表論文集, pp.761-764, 2008.

特記事項

[特記 1] 清水伸幸, 中川裕志: 一同義語獲得のための距離学習一, 「機械学習技術のイノベーション」シンポジウム セミナー発表, 2008 年 3 月 21 日.

Web とテキスト情報処理についての研究

王 玉馨

1 概要

2007 年度は、web ページ分類と同義語抽出のテーマについて研究を行った。Web は高品質な学術的な情報サービスに価値を高める潜在的情報源としてますます重要になっている。効率的に品質保証(すなわち、高い適合率と高い精度)された Web ページを収集するため、関連ページにある情報を用いて、web ページ分類のテーマで研究を行った。また、シソーラスはテキスト処理たとえばテキスト分類における効率と性能を向上させるための最も重要なリソースの一つであるから、シソーラスの根幹をなす同義語の獲得は重要なタスクである。コーパスから同義語抽出の性能向上させるための方策についても研究した。具体的には、以下に示すテーマの研究を行った。

- ページグループ構造を考える高品質な Web ページ収集

ページがグループ構造であることを考えて効率的な選択と正確な分類二段階で高品質な Web ページを収集する。ローカルのページグループの考えを適用して、まず効率的な選択ではノイズページを除いてすべて可能的なホームページを包括的に集める。つぎに、正確な分類では適合率保証と精度保証二種類の分類器の組み合わせを用いて、正確に品質保証された Web ページを分類して、品質を保証するために人手で査定する必要なページ数を削減にする。

- 単語類似性ネットワークを通る自動同義語獲得

コーパスから同義語の対を抽出するための一般的な方法では二つ単語間の類似度(例えば、cosine 類似度)が必要である。類似度を使って、特定のクエリ単語に対しての類似語ランキングができるので、同義語候補リストから正しい同義語が同定できる。この研究では、それに加えて、単語類似性ネットワークを分析する新しい方法を提案する。単語類似性ネットワークは閾値以上の類似度をアークとして、単語をノードとして定義する。提案する自動同義語候補選択のためのランク閾値(Rank Threshold for synonym candidate Selection method, RTS)を超えるアークによって構成される単語類似性ネットワークはスケールフリーグラフである。この性質に基づいて、相互ランキング法(Mutual Re-ranking Method, MRM)という新しい同義語候補のリランキング手法を提案する。

2 ページグループ構造を考える高品質な Web ページ収集

2.1 背景

Web は高品質な学術的な情報サービスに価値を高める潜在的情報源としてますます重要になっているため、低いコストで品質保証(高い適合率と高い精度)の Web ページの収集は必要である。しかしながら、さまざまなページタイトルと Web ページ構造、および関連ページの希薄性のため、大量の手作業を必要とする。このテーマについて多くの研究者が Web ページの検索と分類を研究したが、これら大部分の研究は大体ベストエフォートであって、品質保証されていない。この問題を解決する

ために、この研究では高い適合率と高い精度両方を保証するホームページを効率的に収集する方法を求める。

以前の研究が、一般に様々な Web 関連の特徴を使用することによってホームページを集めるのが有効であることを示した。ホームページが論理的なページグループによって表されることを考慮すれば、エントリーページの特徴コンテンツに加えてページグループ構造(ローカルのリンク構造)の周囲ページの特徴情報を利用しなければならない。したがって、この研究では SVM でページグループ構造の特徴を用いて高品質なホームページを収集する方法を提案する。Web ページの数が巨大なので、提案した方法でページ特徴の抽出のコストは高いはずである。低コストにするため、この研究では全処理は二段階に分けた。効率的な選択ではノイズページを除いて高い適合率でホームページを集める。続いて、正確な分類では快速な選択段階から出力された候補ページを高い適合率と高い精度で分類する。二つ処理の段階でもページグループ構造を考えている。この Web ページの収集方法では、特に高い適合率の分類性能を目指す。

2.2 内容

提案した二段階処理は以下の図1で示している。

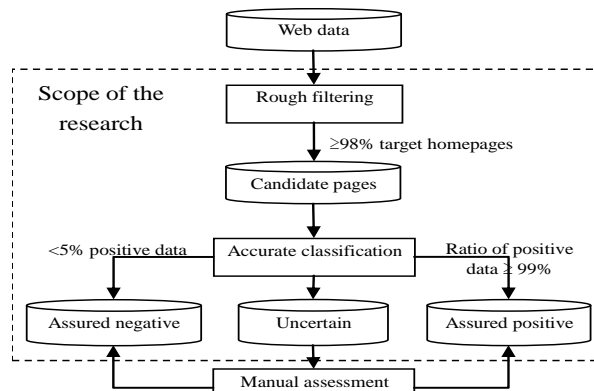


図 1. 二段階の処理.

高速な選択処理段階は図 2 の通りである。

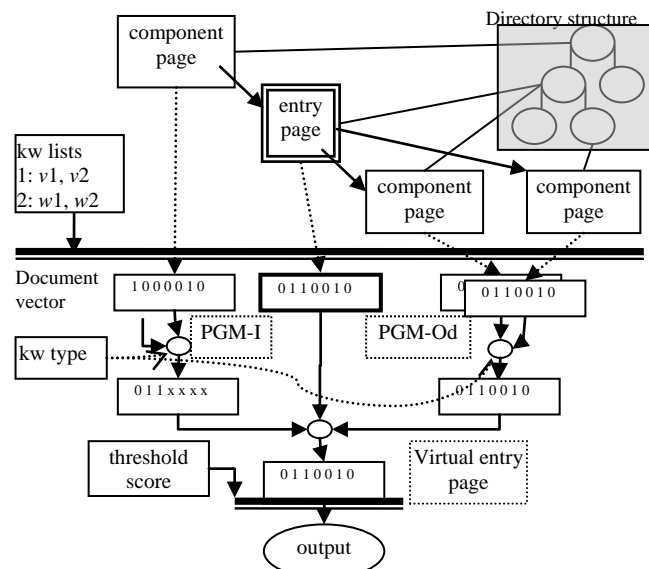


図 2. 高速な選択処理.

この段階の処理では同じカテゴリにおいてホームページに共通の基本情報要素を握るために、ページの基本特徴を持っているキーワードリストを導入する。例えば、研究者のホームページについて定義したキーワードは 12 種類である。それは研究テーマ、履歴、専門、発表など、共に 86 個である。エントリーページはローカルページグループにあるページからページグループモデル (Page Group Model, PGM) でキーワードをもらう。PGM は関連ページ種類によっていくつがある。例えば、Od (out-links downward)、Ou (out-links upward)、I (in-links) と U (directory entry page) など。もし、もらったキーワードと自分が持っているキーワードの共にあらかじめ決められたキーワード種類数以上だったら、このエントリーページは目標ページとして選択されて、続いて次の段階で処理される。

正確な分類処理段階は図 3 の通りである。

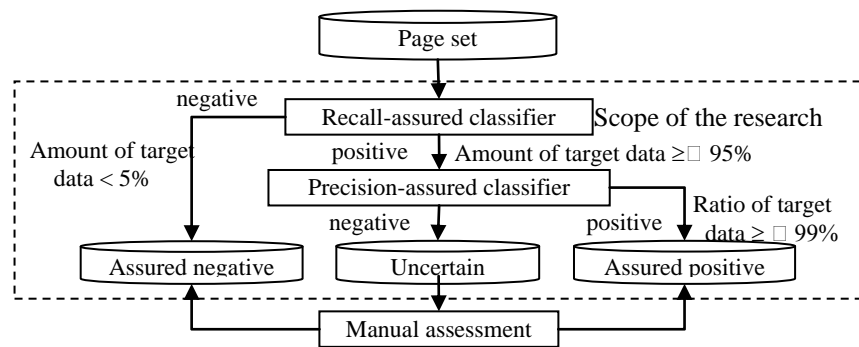


図 3. 正確な分類処理.

ここで、高適合率分類器と高精度分類器ともページグループ構造の特徴を適用している。高速な選択処理で選んだエントリーページを分類するため、相応なローカルグループの周辺ページは表 1 の通りにいくつのグループに分けている。分類器の特徴は表 1 で定義されたページグループの中にあるグループ組み合わせである。グループ特徴の組み合わせの数は 1 個から全部まで可能であるし、組み合わせの方法は八種類別別並び、 $P_{same}(r)$ と $P_{low}(r)$ と $P_{upper}(r)$ ことに並び、 $P_{ent}(r)$ と $P_{in}(r)$ と $P_{ent}(r)$ ことに並びなど色々ある。特徴値は各ページの普通のテキストとタグ付きテキストから得る。

表 1. 周辺ページのグループ.

	r	$P_{ent}(r)$	$P_{in}(r)$	$P_{ent}(r)$
$P_{same}(r)$	G_{cur}	$G_{out,same}$	$G_{in,same}$	$G_{ent,same}$
$P_{low}(r)$		$G_{out,low}$	$G_{in,low}$	
		$G_{out,upper}$	$G_{in,upper}$	$G_{ent,upper}$

2.3 具体的成果

本研究提案した方法を評価するため、100GB の日本語 Web コーパス (NW100G-01) で実験した。評価セットは手作りで、正解 426 個と負解 10,912 である。

高速な選択処理の実験結果によって、ページグループモデルを用いて、エントリーページは 4 種類以上のキーワードを持っている場合、正解の適合率は 99% になった。このとき、出力されたページ数は全部選択されたページ数の 22.9% である。

正確な分類処理の実験結果によって、高適合率と高精度分類器で分類されたページの最高性能 (F-measure) は 88.65% である。99% 精度の場合、最高適合率は 23.18% で、95% 適合率の場合、最高精度は 41.33% である。ページグループというアイデアを適用しないより、各評価指標分別 5.39%、5.2%、14.84% 向上させた。

ちなみに、提案した高品質なホームページ収集する方法を用いて、精度 99%と適合率 95%を保証するため、人手で査定する必要なページ数は前の 57.1%までに抑えた。

3 単語類似性ネットワークを通る自動同義語獲得

3.1 背景

シソーラスはテキスト処理たとえばテキスト分類における効率と性能を向上させるための最も重要なリソースの一つである。したがって、シソーラスの根幹をなす同義語の獲得は重要なタスクである。特定のドメインにおける高品質のシソーラスがなければ、例えば事故報告コーパス中の重大事件を過大か過小に評価してしまうかもしれない。ただし、対象にしているドメインにおいて、自動的に同義語を取得するのは簡単ではない。

同義語獲得するには様々な方法を提案されている。それらは、同義語名詞なら似ている文脈情報を持つという分布仮説(Harris 1985)に基づいている。同義語獲得では、一般的に以下のステップの通りにこの仮説は実行される。そのため、第 1 段階の処理では、コーパスから抽出された重要度の高い単語の文脈特徴における統計情報を抽出する。したがって、各単語はこれらの文脈特徴のベクトルによって表される。第 2 段階の処理では、*cosine* 類似度などの類似性量度を選び、それをクエリ単語と同義語候補の単語対に適用して、類似度を計算する。類似度の降順で各クエリ単語の同義語候補リストを作る。最後に同義語リストからトップ候補を選んで、クエリ単語の同義語と認定する。

本研究はコーパスから同義語抽出の性能向上させるために、新しい方策を提案する。分布仮説の第 2 段階処理の後に新たに第 3 段階の処理を加えることを提案する。第 2 段階の類似度によって形成されるネットワーク、すなわち単語をノード、類似度の順位が閾値以内の単語間にアークを持つとしてネットワークの構造を調べてみると、スケールフリーの性質を持っているのが分かった。したがって、クエリ単語の同義語である可能性が比較的高い単語だけを対象にする自動同義語候補選択のためのランク閾値を決める手法 RTS (Rank Threshold for synonym candidate Selection method) と単語類似度ネットワークの構造を活用する同義語候補の相互ランキング法 MRM (Mutual Ranking Method, MRM) を提案する。MRM でランキングする際、スケールフリーネットワークにある類似度の降順でランクされた同義語候補はハブ単語と非ハブ単語を分けて扱う。同義語関係は対称だが、本研究で提案した MRM は対称であるというわけではない。以前の研究では単語類似性ネットワークの構造的な特性を使用していない。

3.2 内容

提案した RTS と MRM 方法について説明する。

あらかじめ決められた数よりも多く出現した名詞を目標名詞と呼ぶ。目標名詞はノードとして、二つノード間で付随する類似度がアークとして、単語類似性ネットワークを描ける。今のタスクはあるクエリノードに隣接しているアークからの真の同義語アーク(クエリ単語に対しての正解同義語)を選ぶことである。この論文の中で、自動同義語獲得の対象はコーパスの名詞に限る。この節で調べる単語類似度ネットワークは Reuters-21578 コーパスから得たものである。

3.2.1 スケールフリーネットワークと RTS 手法

スケールフリーネットワークでは、ほとんどのノードは次数が少ないが、いくつかのノードは次数が大きなハブノードである。より具体的には、スケールフリーネットワークはノードからの出次数 k が次式で表わされるようにベキ乗則に従う分布 $P(k)$ である接続グラフである。

$$P(k) \propto k^{-\gamma} \quad (1)$$

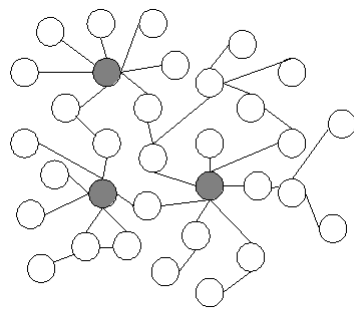


図 4. スケールフリーネットワーク.

図 4 はスケールフリーネットワークの例である。図 4 の中で強調している黒いノードはハブノードである。

実験対象としているコーパスから得られた単語が形成するグラフがスケールフリーネットワークであることを示すために、単語類似性ネットワークにおけるノード間の類似度の順位が閾値 RT 以内のアーチだけを取り出して、ノード間の関係を図 5 で表した。

クエリ単語と第 i 候補単語の類似度が s_i であるとする、クエリ単語に対する第 i 候補単語の出現確率をここでは以下の式で近似する。

$$p_i = \frac{e^{s_i}}{\sum_i e^{s_i}} \quad (2)$$

このように近似すると、パープレキシティ RT は以下のようにになる。

$$RT = 2^H \quad (3)$$

$$H = \sum p_i \log p_i \quad (4)$$

コーパスから得たグラフの定義によれば、 RT はあるクエリ単語に対して情報理論上の同義語の数と見なされるので、ランク閾値として使用するの妥当と考えられる。ここで、提案した式 (3) で定義された自動的にそれぞれのクエリ単語の同義語候補のランク閾値 RT を決める方法 RTS (Rank Threshold for synonym candidate Selection method) と呼ぶ。

単語類似性ネットワークにおける各クエリ単語から同義語候補まで二つ単語間の類似度が最上位の RT 個をとって、隣接ノード数とその頻度の関係は図 5 で示されている。予想通り、単語類似性ネットワークにおいて、度数と与えられた度数を持つノード数の関係はほぼべき乗則に従っている。

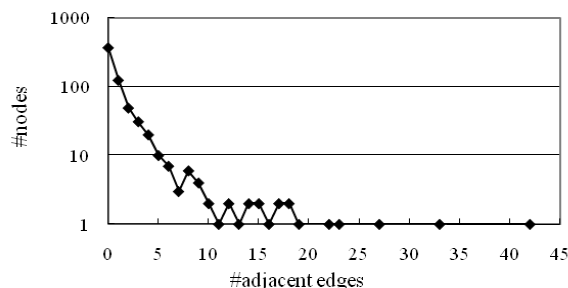


図 5. 隣接ノード数とその頻度の関係.

3.2.2 相互ランキング手法

クエリ単語 x は同義語候補 y があると仮定する。もし y が x に対して真の同義語だったら、逆にクエリ単語 y に対して x が真の同義語ということも成立である。この考えに基づいて、我々は相互ランキング手法(MRM)を提案する。

リランキングする前に、とりあえずクエリ単語 x と x の各同義語候補(クエリ単語以外の目標単語)間の類似度を計算する。そうして、各同義語候補は類似度の降順で同義語候補リストを作る。このリストはオリジナル同義語候補リストと呼ぶ。

続いて、MRM は以下の式(5)で計算されたランクスコア(RS)を用いて、このオリジナル同義語候補リストをリランクする。まず、オリジナル同義語候補リストから一つの同義語候補 y を選んで、 y がクエリ x のオリジナル同義語候補リストの順位は $rank(x, y)$ と表す。反対に、 y はクエリ単語として、 y の同義語候補リストも得られる。もし、 x はクエリ y の同義語候補リストで現れたら、 x がクエリ y のオリジナル同義語候補リストの順位は $rank(y, x)$ と表す。 x が現れないなら、 x をクエリ y の同義語候補と見なさない。

言い換えれば、クエリ x に連続しているアークの中で $rank(x, y)$ 番目に重いのは x から y までのアークである。反対に、クエリ y に連続しているアークの中で $rank(y, x)$ 番目に重いのは y から x までのアークであるといえる。

x から y までランクスコア(RS(x, y))は以下式で定義する:

$$RS(x, y) = A * \log(rank(x, y)) + \log(rank(y, x)) \quad (5)$$

式(5)の A は x から y までの同義語候補リスト順位と y から x までの同義語候補リスト順位を相互考えるための関数である。

式(5)を用いて、クエリ x のオリジナル同義語候補リストにある各候補語 y_i の $RS(x, y_i)$ 計算してから、 $RS(x, y_i)$ の昇順で x の全候補語をもう一度ランク付けする。

3.3 具体的成果

実験では四種類の文脈情報特徴の有効性を評価する: 主語と目的語の動詞、名詞の形容詞、および前後連続している隣接語である。隣接語の場合は、最も役に立つ文脈情報特徴を得られるために、前後 1 から 3 までの窓サイズをテストした。これから、それぞれ隣接語の窓サイズは隣接語 1、隣接語 2、および隣接語 3 と呼ぶ。四種類文脈情報特徴を別々に用いた同義語抽出の実験結果は表 2 で示している。

表 2 各文脈情報特徴の性能と有効性の順番

文脈情報特徴	平均精度	有効性の順番
主語の動詞	0.055	4
目的語の動詞	0.139	1
名詞の形容詞	0.113	2
隣接語 1	0.105	3
隣接語 2	0.104	
隣接語 3	0.091	

実験結果によって、別々に各文脈情報特徴を用いたら、目的語の動詞が同義語獲得に最もよく働いた(13.9%)。名詞の形容詞が 2 番目で、隣接語 1 が 3 番目である。主語の動詞の有効性は最低で、5.5%しかないので、目的語の動詞と名詞の形容詞の性能の半分未満である。隣接語の場合、隣接語 1 の性能が一番高い。隣接語 2 は隣接語 1 よりわずかに下がったが、隣接語 3 はさらに低くなった。つまり、本研究の実験結果によって、隣接語を用いて同義語獲得する場合、隣接語 1 は一番役に立つ文脈情報特徴で、より大きい窓がより役に立つ特徴を提供するとは言えないことを示した。

隣接語 1 が異なった窓サイズの中で最もよく効いたので、文脈情報特徴を組み合わせるときは目的語と主語の動詞、名詞の形容詞、および隣接語 1 の四種類文脈情報特徴で実験を行った。しかしながら、すべて四種類文脈情報特徴の組み合わせは目的語の動詞だけを用いるときより性能が低い。このことから見ると同義語獲得するために各文脈情報特徴の重要性が異なっているので、組み合わせるときはそれぞれの文脈情報特徴に異なる重みを付けるべきである。

続いて、文脈情報特徴は異なる重みで組み合わせを実験した。この実験の性能は目的語の動詞だけを用いるときより高めたため、各文脈情報特徴が表 1 に示されている有効性順で一つずつ加えて、重みを調整して、実験をした。実験結果は表 3 に示されている。

表 3 異なる重みを付けた各文脈情報特徴の組み合わせの性能

文脈情報特徴	重み	平均精度
目的語の動詞	1	0.1387
目的語の動詞+名詞の形容詞	2+1	0.1717
目的語の動詞+名詞の形容詞+隣接語 1	4+2+0.3	0.1739
目的語の動詞+名詞の形容詞+隣接語 1+主語の動詞	4+2+0.3+0.01	0.1741

本研究で提案した自動同義語候補選択ためのランク閾値の決める手法 RTS の有効性が分かるために、表 3 の実験結果中に性能が一番高い組み合わせ(最後のケース、これから「組み合わせ」と呼ぶ)で様々な同義語候補ランク類似度閾値をテストしてみた。RTS によって自動的に選択された候補ランク閾値と比較された結果は図 4 で示されている。図 6 の横軸は全コーパスに適用する人手で決めた類似度閾値であるが、Y 軸は様々な同義語候補ランク類似度閾値が適用された F-measure で表された性能である。凡例の「RTS」は提案した同義語候補選択ランク閾値の決める手法 RTS で、つまり式(3)によって自動的に選択された RT を適用された F-measure である。

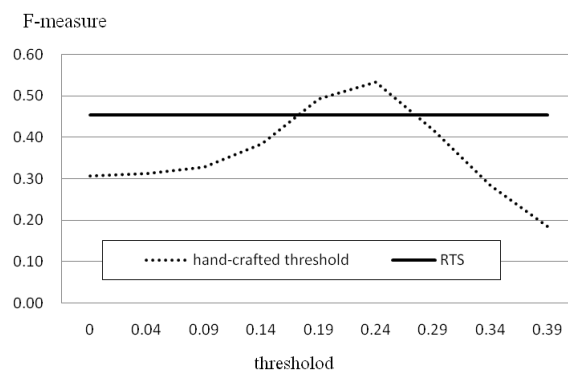


図 6. 自動同義語候補選択ためのランク閾値の決める手法 RTS と人手決めた類似度閾値を適用された性能比較。

実験結果によって、自動同義語候補選択ためのランク閾値の決める手法(RTS)が比較的によく寄与することを示している。RTS での性能が一番高くないが、RTS を用いない場合より RTS を用いるほうが同義語獲得に有効的な傾向になることが分かった。それに、閾値の人手決めが難しいが、RTS はクエリ単語によって自動的に選択できる。各クエリ単語に対して正解としての同義語候補選択ためのランク閾値が分からないため、本研究の実験結果は実用的に重要であると考えられる。

次に、提案した同義語候補の相互ランキング法(MRM)の有効性が分かるため、cosine の類似度を基づく各クエリ単語に RTS を適用した上さらに全体に MRM を適用して、実験した。RTS の RT と MRM の RS は式(3)と(5)の通りで計算する。実験結果は表 4 に示されている。

表 4 の中、「組み合わせ」という行は異なる重みを付けた四種類の文脈情報特徴の組み合わせを意味する。「--」という列は提案した RTS と MRM を用いないときの性能で、「RTS」列は RTS 方法での性能で、「MRM」は RTS を用いた上にさらに MRM を加えて適用した性能を意味する。「--」列の場合は、全部クエリに対して類似度閾値は 0 に設定されている。つまり、類似度は 0 以上だったら、同義語候補になるという仮定している。

表4 提案した RTS と MRM を用いる前後
各文脈情報特徴と組み合わせの性能比較。

文脈情報特徴	平均精度		再現率	F-measure	
	RTS	MRM		RTS	MRM
主語の動詞	055		.935	104	
	.112	.168	.714	.193	.272
目的語の動詞	139		.909	241	
	.235	.268	.818	.365	.404
名詞の形容詞	113		.916	202	
	.225	.251	.727	.344	.373
隣接語 1	105		.851	187	
	.131	.178	.520	.209	.265
組み合わせ	174		.935	294	
	.293	.314	.818	.432	.454

実験結果によって、候補ランク閾値は低い時、同義語獲得の再現率は高いが、クエリ単語に対応する比較的到低い類似度をもっている同義語候補は同義語獲得の精度は低い。逆に、候補ランク閾値は高かったら、比較的到低い類似度をもっている同義語候補は無視されたので、同義語獲得の精度を向上させたが、再現率は劣化させた。式(3)に基づく各クエリ単語対して自動的に同義語候補の数を決めた RT を適用することによって、クエリ単語に対応する比較的に高い類似度をもっている同義語候補だけが同義語として扱われる。したがって、表 4 の通りに平均精度は全体的に改良させた。結論として、RTS を用いて再現率は少し劣化させたが、総合的な性能指標 F-measure は改良させた。

続いて、クエリ単語と同義語候補対の関係を考えて、さらに提案した同義語候補の相互ランキング法(MRM)を適用することによって、平均精度と F-measure の両方は全体的に改良された。表 4 の平均精度は関数 A が最適化された式(5)で計算された RS によってのものである。実験結果によって、提案した RTS の上さらに MRM を応用したところ、同義語獲得の性能を最も向上させたことを示した。

4 成果要覧

査読付き論文リスト

[査読付 1] Yuxin Wang, Keizo Oyama: Framework for Building a High-Quality Web Page Collection Considering Page Group Structure, The 9th Asia-Pacific Web Conference and the 8th International Conference on Web-Age Information Management (APWeb/WAIM'07), Huang Shan, China, Jun. 16-18, 2007, pp.95-107.

その他の発表論文リスト

[発表1] Yuxin WANG, Nobuyuki SHIMIZU, Minoru YOSHIDA, and Hiroshi NAKAGAWA: Contextual Information Based Technical Synonym Extraction, 言語処理学会 第 14 回年次大会, B4-4, 2008 年 3 月 17 日—21 日.

レポート文書の文のカテゴリー分け

鍋島 彰崇

1 概要

本研究は、平成 19 年度情報大航海プロジェクト「新総合安全運航支援システム」における 1 つの技術開発テーマ(レポート文書の文のカテゴリー分け)として行われたものです。

本技術研究は、分類したいレポート文書を設定するだけで容易に文書内の文の分類を行い、大量レポート分析の効率化・精度向上に貢献する技術開発を行ったものである。ここでは、その研究における技術開発についての概要と特徴について述べる。

開発した技術の概要は、分野固有情報を用いずに、類似度を用いた機械学習によって文書中の文の分類を行うものである。茶筌 (chasen) などの形態素解析ツールを使用することで、容易に類似度計算が行えるので、文の分類を比較的簡単に行うことができる利点がある。

分野によらない汎用的な手法なので、他分野など応用範囲が非常に広いものと考えられる。その一方で、本研究における技術は、単純な手法であるので、他手法との組み合わせなどの応用にも幅広く対応できるものであるといえる。

提案する技術は、茶筌 (chasen) などの形態素解析ツールから得られる単語頻度や品詞頻度等から類似度計算するのみで、機械学習によって文書分類を行う方法の開発を行ったものである。従来の手法や先行研究では、個々の分野に依存した情報の利用や日本語の意味情報を用いて行っているものが多い。一方、本研究においては、分野固有情報や日本語の意味情報を用いずに、文と文の関係の類似度計算のみで、分類対象の文書中の各文を 1 文単位で分類を行う。分類の流れとしては、最初に分類したい文書 (1 文や複数の文) を参照文書 (図 1 では、文書 X、Y、Z) として設定する。分類したい文書は、ある程度の数が必要になるので、ある程度の数参照文書が設定される。このすべての参照文書を参照文書集合と定義する。次に、この 1 つの参照文書を 1 つの素性として、参照文書集合と教師データを作り出すための文書 (図 1 の正例文書、負例文書) との類似度を計算して、参照文書の重みを機械学習に学習させる。さらに参照文書集合と 1 文単位の判定すべき文 (図 1 の判定すべき文書 1) との類似度の値から機械学習で判定することで、文の分類をする。

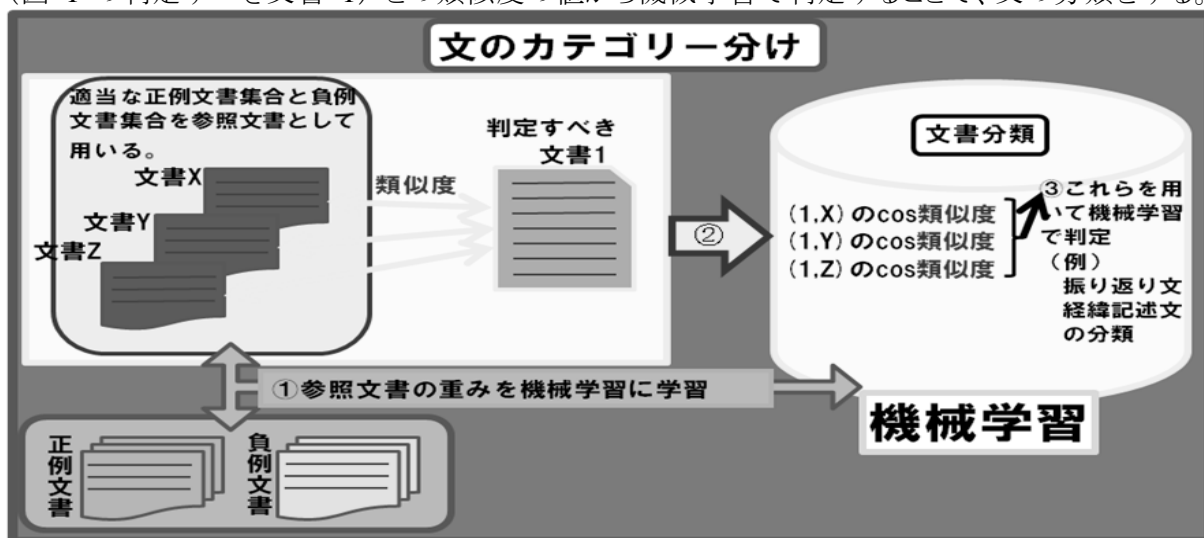


図 1 本研究における技術の概略図

2 レポート文書の文のカテゴリ分け

2.1 背景

航空業務の運航において、パイロットなどの当事者からの声(レポート)を考慮することは、安全を確保する上でも必要不可欠である。このレポートに記載されている内容は、多岐にわたり、人手で解析するために人手で分類することは、量的に不可能である。そのために、このレポート文書をいかに分析するかが重要な課題となっている。

そのためには、レポート中の文書を区別し分類することが必要である。このような分類アルゴリズムは従来から研究されているが、ここでは、対象とする分野固有の情報を利用せずに、レポートに記載されているテキスト相互間の関係を利用した技術の開発を行う。本技術研究で提案する文書分類技術は、分野固有の情報を用いなくて、航空以外の、鉄道や道路交通分野、電力、その他の産業分野などの安全管理の支援技術として他分野などに応用が期待できる。

2.2 内容

2.2.1 アプローチ

本技術開発のアプローチは、対象とする分野固有の情報を利用せずに、レポートに記載されているテキスト相互間の関係を利用して、技術開発を行う。また、このアプローチのための提案手法であるのが参照文書集合であり、これについて、以下で説明を行う。

2.2.2 参照文書集合

ここで、提案手法である参照文書集合について説明する。(図2参照)

まず、通常の機械学習で行われているように、教師データ用の文書集合の正例と負例(本研究においては、正例 100 例、負例 100 例)を用意する。また、教師データ用の文書集合とは別に、参照文書集合(本研究では、正例 100 例)を用意する。

そこで、1つの参照文書の特徴量(例:頻度の高い単語や品詞頻度)を

$$W_a = (W_{a1}, W_{a2}, \dots, W_{an})$$

とし、教師データ用の文書集合の1つの文書(正例)の特徴量を

$$W_b = (W_{b1}, W_{b2}, \dots, W_{bm})$$

とする。この cos 類似度は、

$$\cos \text{類似度} = \frac{W_a \cdot W_b}{|W_a| |W_b|} = \frac{W_{a1} \cdot W_{b1} + \dots + W_{an} \cdot W_{bm}}{\sqrt{W_{a1}^2 + \dots + W_{an}^2} \sqrt{W_{b1}^2 + \dots + W_{bm}^2}}$$

となる。これをすべての参照文書(参照文書集合)と教師データ用の文書集合の正例で計算して正の類似度とする。同様に、参照文書集合(正例)と教師データを作り出す文書集合の負例との cos 類似度を計算し、その値を負の類似度とする。この正の類似度を正例の教師データとして、また、負の類似度を負例の教師データとして機械学習に学習させる。

一方、参照文書集合(正例)と1文単位の判定すべき文との cos 類似度を計算する。この値を機械学習に与えて、この判定すべき文を判定する。これより、判定する方法としては、1文単位の判定すべき文と参照文書の全体である参照文書集合との類似度が、正の類似度と負の類似度のどちらに近いかで判定するものである。

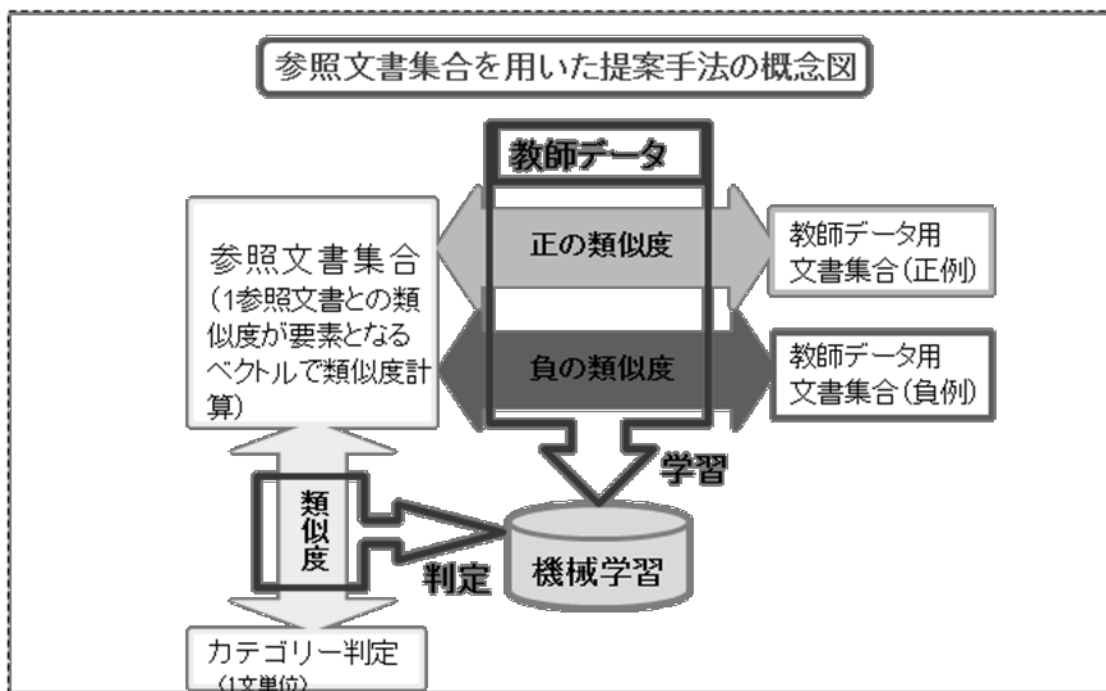


図 2. 参照文書集合を用いた提案手法の概念図

2.2.3 実験結果

これより、本技術開発で行った実験結果について述べる。

まず、従来から文書分類において一般的に用いられている **bags-of-words** についての説明とその実験結果、そして、本技術開発の提案手法である参照文書集合の実験について説明する。なお、実験対象の JAL-FDM コーパスは、航空業務用のテキストである。本研究でのテキストの分類は、振り返り文(主観的記述)と経緯記述(客観的記述)への 2 分類という特殊なタスクであるため、参考にできる主観的記述と客観的記述の文の分類に関する先行研究が非常に少ないために、従来法として一般的に広く用いられている **bag-of-words** による実験を参考例として比較の対象に用いた。

これより、本技術開発で行った具体的成果について述べる。

従来法 (bags-of-words)

文書分類において、一般的に用いられているのが **bag-of-words** 法である。これは、分類する対象の文書群の中で、単語の頻度を抽出し、頻度の高い単語を **feature** として機械学習などで文書を分類するものである。

本研究では、データの文書として、JAL-FDM コーパスを使用する。この中には、パイロットやクルーなどの声をレポートとしてまとめた Crew が書いた文書と、パイロットや機長に対してインタビューを行い、その現場の声を編集者によってまとめた編集者が書いた文書がある。また、Crew が書いた文書と編集者が書いた文書の内容としては、振り返り文や経緯記述文、環境、問題事象など様々な内容の文書で成り立っている。

ここで、本研究においては、振り返り文と経緯記述文の 2 種類を使って文の分類を行う。振り返り文と経緯記述文の分類を、**bags-of-words** で行った実験について説明をする。まず、すべての振り返り文から、形態素解析ツールの Chasen などから、文の内容を特徴として表していると考えられる頻度の高い単語を抽出する。なお、頻度の閾値としては、頻度の高い順に、まず、100 語、200 語を取り出し、実験データ文の数から、頻度の高い単語を増やした場合は、頻度を基準にとり、474 語(頻度 4 回)、755 語(3 回)、1708 語(2 回)とする。この頻度を以下で示す実験での機械学習の **feature** で

の単語として設定する。また、機械学習(SVM)としては、一般的によく用いられる libSVM を用いて文書分類を行う。以下に、実験条件と実験結果を示す。

○実験条件

機械学習 SVM(libSVM)

教師データ 200(正例:100 負例:100)

評価データ 400

振り返り文/経緯記述文を1文単位で判定する。

○実験結果

●feature:頻度の高い単語 100 語

Crew が書いた文書 55.5～61% (平均 57.47917)

編集者が書いた文書 51.25～57% (平均 54.42647)

●featureを増やした場合(Crew が書いた文書)

200 語:55～56.5% 474 語:51%

755 語:57.75% 1708 語:58%

この実験結果から、Crew が書いた文書のほうが、編集者が書いた文書より高い精度であることが分かる。これは、Crew が書いた文書の場合は、振り返り文として、反省や改善点などの内容であるが、経緯記述文は、過去に起こったことを順に記述しているために、振り返り文と経緯記述文とでは、使用される単語が少し異なってくるために、編集者が書いた文書に比べて、若干の精度の違いが生まれたものであると推察される。また、頻度の高い単語を増やした場合(Crew が書いた文書)、精度がほとんど向上しないことが分かる。これは、頻度の高い単語の順序の中に、振り返り文をうまく特徴づける単語が少ないと考えられる。また、すべての日本語文の文書で良く使われる助詞などを除外しないで使用している点もある。

参照文書集合の実験

ここでは、従来法(bags-of-words)に提案手法である参照文書集合を適用した場合について説明する。

まず、教師データを作り出すための文書集合[教師データ用の文書集合](正例と負例)と参照文書集合を設定する。ここで、従来法である bags-of-words によって、教師データ用の文書集合(正例と負例)と参照文書集合のそれぞれの文書の bags-of-words による頻度の高い単語の個数を抽出する。

次に、教師データ用の文書集合の正例と参照文書集合の頻度の高い単語の個数から cos 類似度を計算し、この値を正の類似度とする。同様にして、教師データ用の文書集合の負例と参照文書集合の頻度の高い単語の個数から cos 類似度を計算し、この値を負の類似度とする。この正と負の類似度を機械学習に学習させる。

一方、1 文単位の判定すべき文と参照文書集合の文書との類似度を計算し、機械学習で判定する。

以下で、本研究での実験条件について示す。

○実験条件

機械学習 SVM(libSVM)

教師データ 200(正例:100 負例:100)

評価データ 400

参照文書集合 100(1 文単位、複数文単位)

振り返り文/経緯記述文を1文単位で判定する。

(※なお、複数文単位とは、1文以上の複数の文を一つの塊とみて、一つの単位とする。)

●実験結果 1

feature: 頻度の高い単語 100 語による類似度

参照文書集合: 1 文単位 100

精度: Crew が書いた文書 79~80.25% (平均 79.44444)

編集者が書いた文書 55.75~57.25% (平均 56.41667)

●実験結果 2

feature: 頻度の高い単語 100 語による類似度

参照文書集合: 複数文単位 100

精度: Crew が書いた文書 63.25~66.25% (平均 64.89286)

編集者が書いた文書 54.25~55.75% (平均 54.62222)

実験結果から、編集者が書いた文書の頻度の高い単語は、振り返り文や経緯記述文のどちらかで特徴的に使われる単語がほとんど無く、名詞、動詞、助動詞としても相関性がなく分布している傾向があり、これにより、精度が向上しないものと考えられる。また、Crew が書いた文書の頻度の高い単語では、振り返り文/経緯記述文の両方によく使われる単語が頻度上位にきているので、複数文を参照文書集合とするとそのような単語がカウントされやすくなり、その影響によって精度が低下した。このよく使われる単語には、助動詞(「た」、「で」、「ます」等)が頻度上位に多数あり、続いて動詞、名詞の順の傾向がある。

さらに、頻度の高い単語の個数を増やした場合についての Crew が書いた文書の場合の実験結果を示す。

参照文書集合: 1 文単位 100

feature: 頻度の高い単語 200 語	精度: 84.25~87.5% (平均 85.80556)
500 語	精度: 84.75~88.75% (平均 86.72222)
1000 語	精度: 85.75~89.5% (平均 87.5)
2000 語	精度: 86.25~90.25% (平均 88.13889)

実験結果から、頻度の高い単語の個数を増やしていくことで、徐々に精度が向上して、90%以上に達していることが分かる。これは、参照文書集合を適用したことによって、bag-of-words による値の類似度が有効には働いていることを示しているものであると考えられる。

正の類似度と負の類似度の比較

参照文書集合の適用によってえられる類似度において、正の類似度と負の類似度の比較について説明する。

●参照文書集合を用いた bags-of-words の頻度の高い単語による類似度比較

ここでは、正の類似度と負の類似度を比較するために、1 文単位の振り返り文の正例と負例の類似度を計算し、グラフ(図 3~5)に示して考察を行う。

○ feature の頻度の高い単語を多くすると、負の類似度の値が全体的に小さくなる。これは、振り返り文と 経緯記述文の頻度の高い単語がそれぞれ違う傾向のために、振り返り文(正例)と経緯記述文(負例)の類似度は、小さく計算される。

これによって、正の類似度負の類似度との値の差が大きくなり、精度が向上していると考えられる。

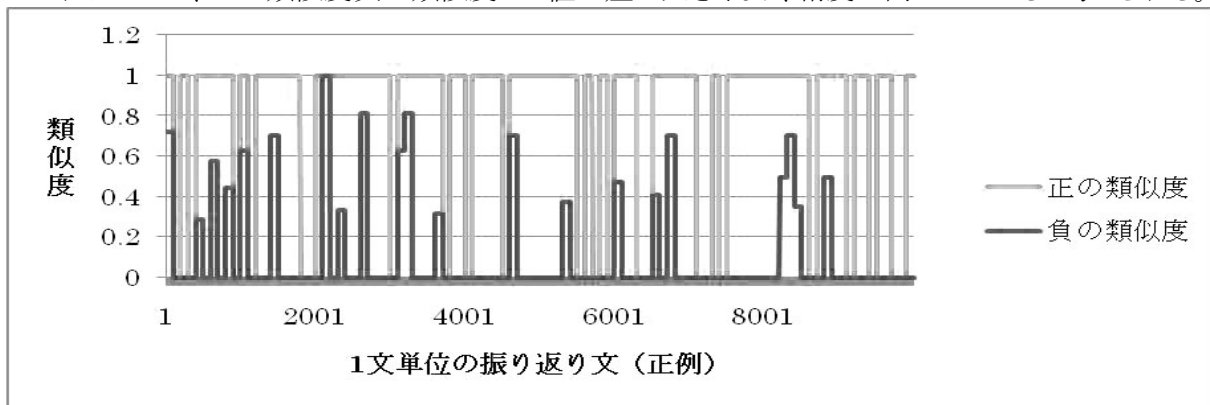


図3 頻度の高い単語 100 の場合の正と負の類似度比較

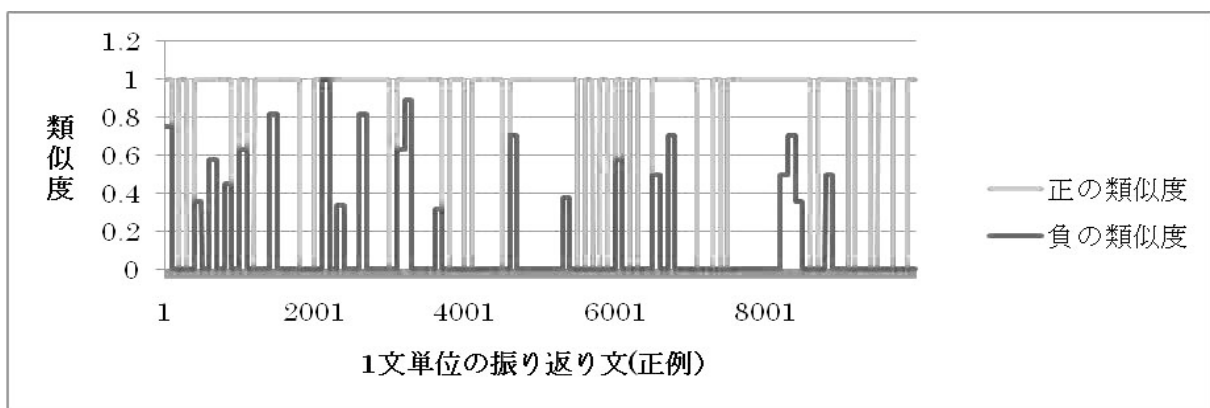


図4 頻度の高い単語 500 の場合の正と負の類似度比較

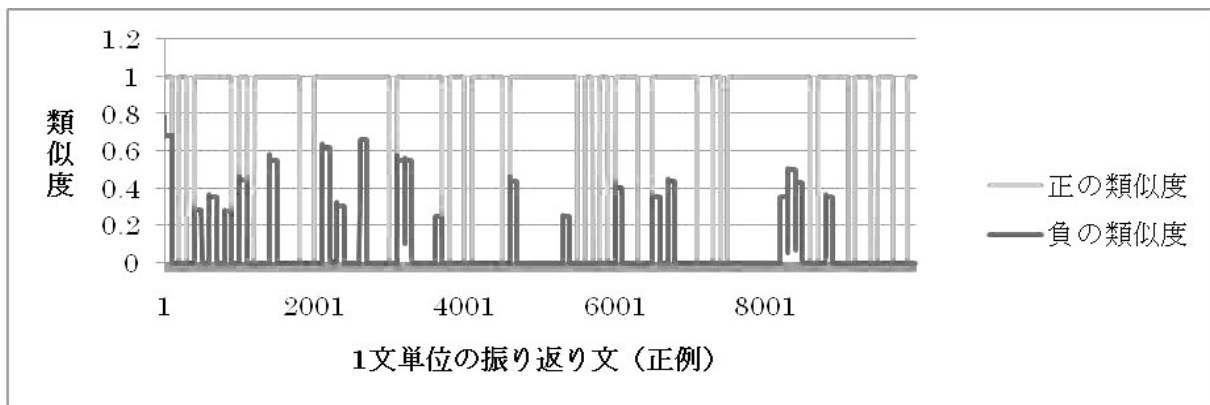


図5 頻度の高い単語 2000 の場合の正と負の類似度比較

2.3 具体的成果

本技術研究は、参照文書集合の方法を考案して、分類したいレポート文書を設定するだけで容易に文書の文の分類を行い、大量レポート分析の効率化・精度向上に貢献する技術開発を行った。これによって、従来、一般的に用いられている bag-of-words の頻度の高い単語による方法で 50~60% 台の文書分類の精度を、参照文書集合の方法を組み合わせることによって、約 80%以上の精度で文書の文を分類する方法を開発し、検討を行った。この参照文書集合による方法は非常に簡単で

精度が向上するために、別の方法との組み合わせなどを含めて応用について検討を重ねる必要があると考えられる。

3 成果要覧

特記事項

[特記 1] 鍋島 彰崇: レポート文書の文のカテゴリー分け, 平成 19 年度情報大航海プロジェクト「新総合安全運航支援システム」 報告書

キャンパスネットワークング研究部門

若原 恭

ネットワークの設計・運用・保守・管理の合理化を図る技術
— アドホックネットワーク、ネットワークセキュリティ、車車間通信、
次世代衛星ネットワーク、及び惑星間インターネット—

中山 雅哉

広域分散環境の高度基盤技術に関する研究

加藤 朗

ネットワーク運用に関する研究

小川 剛史

放送型データ配信を用いた情報システム構築に関する研究

中村 文隆

高帯域遅延積環境における高効率転送

中村 誠

教職員のための情報共有サイトの開発と高品質なインターネット中継システムの研究

関谷 勇司

DNS の安定運用に向けた分散計測システムの設計と構築に関する研究

ネットワークの設計・運用・保守・管理の合理化を図る技術

— アドホックネットワーク、ネットワークセキュリティ、車車間通信、次世代衛星ネットワーク、及び惑星間インターネット—

若原 恭

1 概要

情報通信ネットワークは、そのアプリケーションの急激な発展とともに、転送情報量の規模的な拡大に加え、機能面の質的な変革も必要となる。このような認識のもと、情報通信に係わる各種リソースの最大限の利活用によって、情報通信ネットワークの設計・運用・保守・管理をできる限り合理的なものにすることを目標にしている。具体的には、以下に示す諸技術に焦点を当て、理論と実験の両面から研究を行っている。

- アドホックネットワークの品質制御

移動可能な無線ノードのみから構成されるアドホックネットワークは、今後の本格的なユビキタス社会において必須になると考えられている。そのようなアドホックネットワークでは、運用と制御に係わる多くの基本技術が課題として山積している。そのうち、品質制御に係わる実用的な基本技術の確立を目指し、最大の情報転送能力を持つと判断される情報転送ルートの探索法、情報転送ルートにおいて最大情報転送能力を実現する方法等を検討している。

- ネットワークセキュリティ技術—無線ネットワークにおける中間者攻撃の検知—

無線ネットワークでは、転送情報を傍受したり、他のノードになりすましたりすることが比較的容易に実現できる。特に、存在を隠して情報転送の中継を行う中間者攻撃は大きな脅威となる。そこで、このような中間者攻撃を検知する手法を検討し、安全な無線ネットワークの構築と運用に資することを目指す。

- 車車間通信を応用した緊急通信

21世紀に入っても車社会が発展し続けているが、交通事故や交通渋滞等の課題の解決は期待通りには進んでいない。しかしながら、昨今発展の著しい情報処理・情報転送技術を応用することによってこれらの問題を解決できる可能性が大きくなってきた。そこで、車車間通信によって、交通事故や道路状況の変化等の情報を伝達する緊急通信の実現技術について検討を進めている。その目標は、緊急情報を、可及的速やかにかつ確実に必要な車両に届ける技術の確立にある。

- 次世代衛星ネットワークにおける衛星障害対策

低軌道周回衛星システムは、地震やテロ等の大災害に強く、グローバル性や同報性等の特長を持つことから、次世代の衛星ネットワークとして有力視されている。そのようなシステムの運用において衛星が故障した場合の対策方法について検討を進めている。こ

の方法によって、安定した運用が可能で衛星障害に強い低軌道周回衛星システムの実現を目指す。

- 惑星間インターネットのためのパケット転送制御

人類が宇宙に展開する時代においては惑星間の情報転送が必要になるが、そこでは伝送品質に加え、伝搬遅延時間が極めて大きいため情報が届くまでに要する時間が著しく大きくなるという問題がある。この問題に対処するため、中継箇所におけるパケット転送制御の工夫によって、情報転送所要時間を削減する検討に取り組んでいる。その目標は、地上或いは地球周辺の衛星ネットワーク等を対象にこれまでに研究されてきた技術に比較して、情報転送所要時間を大幅に短縮することにある。

2 アドホックネットワークの品質制御

2.1 背景

アドホックネットワーク MANET(Mobile Ad-hoc Network)は、ルータやアクセスポイント等のインフラ設備を用いることなく、移動可能な無線端末ノードのみから構成されるネットワークで、例えば、イベント会場や旅行先での情報交換、走行中の車同士の間での情報交換等、近い将来に実現が予想される本格的なユビキタス社会において必須になると考えられている。このため近年研究が活発になってきているが、例えば、所要の品質を維持したパケット転送法、無線リソースの有効利活用によるパケット転送情報量の拡大等、未解決の技術課題も多く、これらの解決が急務となっている。

2.2 内容

MANET を利用して情報を転送する際には、MANET を構成するノードの処理能力や動き等に応じて転送情報の品質(QoS: Quality of Service)が動的に変化するため、情報転送の開始時に所望の品質が達成できるか否かを判定できることが必要となる。そこで、品質として最も基本的な情報転送能力(伝送帯域)を取り上げ、送信ノードと宛先ノードを結ぶ複数のルートの中から最大情報転送能力を持つルートを探査する方法を検討した。この探査では、一般に候補ルート数が膨大になることから、探査処理が速やかに完了することと正確さが要求される。検討の結果、各ルートが持つ情報転送能力を、そのルートを構成するリンクが持つ情報転送能力の最小値で近似し、その最小値が最大となるようなルートを選択するという原理に基づく方式を考案した。そして、この方式の有効性を明確にするため、格子状に配置されたノードから構成されるネットワークをモデルとして実験評価を行い、正確な情報転送能力を計算する方式では計算量が膨大となり実用化には適していないこと、及び考案した方式によって、計算量を飛躍的に削減でき、しかも得られる情報転送能力値はほぼ正確で誤差は数%程度以下であり、実用上問題ないことを確認することができた。

次に、このようにして情報転送能力が最大と判断されたルートにおいて実際に情報転送を実現するため、そのルートを構成する各リンクで帯域を割り当てる必要があり、その方法を具体的に検討した。この検討では、MANET の情報転送方式として、一般に情報転送容量が大きく、特に品質の高い情報転送に有効であると考えられている時分割多重方式 TDMA(Time Division Multiple Access)を前提とする方針とした。そして、各無線リンクのタイムスロットの情報転送方向、即ち電波伝搬方向を考慮してタイムスロットを割り当てることによって、無線電波の干渉を回避したうえでタイムスロットの有効利用を最大限に図り、情報転送能力を拡大することが可能な原理を考案した。更に、この原理に基づくタイムスロット割当アルゴリズムを詳細に設計してプログラム化し、実験評価を行った。その結果、ネットワーク規模の大小に依らず高い精度で情報転送能力が導出できること、既存の手法に比較してその情報転送能力はかなり大きく、特に割り当て済みのタイムスロットが多いほどその効果が大き

いこと、及び計算時間が極めて小さく実用上問題にならないことを確認した。情報転送能力に関する改善効果の一例を図1に示す。

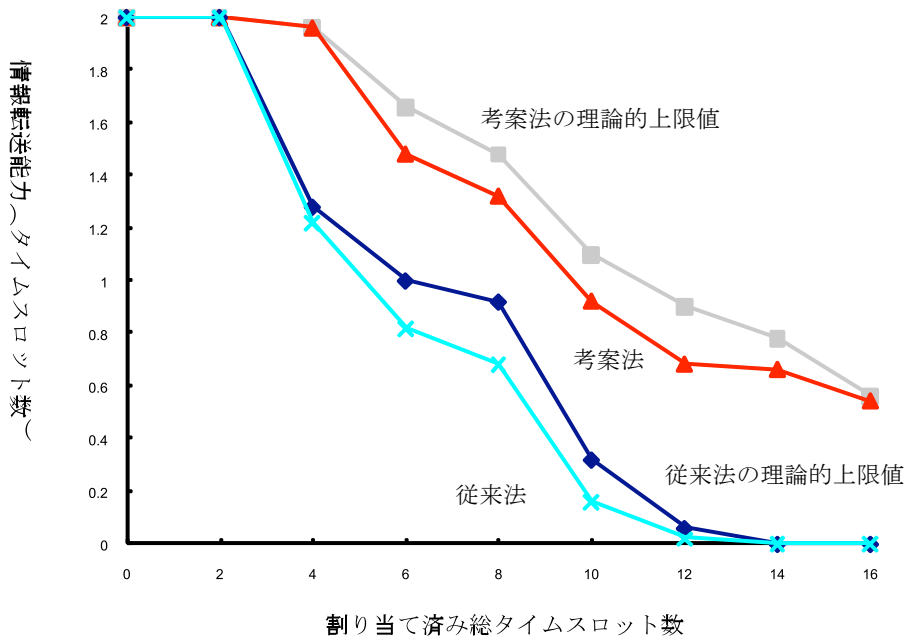


図1 考案したタイムスロット割当法の効果

2.3 具体的成果

前節で述べた諸技術は、今後の期待が大きいアドホックネットワーク MANET が現在抱えている基本問題の解決を目指すもので、それぞれ効果が明確であることを確認できており、MANET の実用と発展に寄与するものと期待される。

最大情報転送能力を持つルートを選択方式については、一つのアドホックネットワーク例でその有効性を確認することができ、来年度に学会発表をするべく論文投稿の準備を進めた。

与えられたルートにおける情報転送能力の導出とタイムスロット割当法については、精度が高くかつ導出に要する計算量が少ないことを実証でき、導出法の原理とアルゴリズム及びその定量的評価結果を論文誌論文、国際学会及び研究会等で論文発表した[査読付 2, 3、発表 1, 3]。

3 ネットワークセキュリティ技術

—無線ネットワークにおける中間者攻撃の検知—

3.1 背景

現在広く普及しているインターネット技術を無線ネットワークに応用した場合、隣に位置するノードがパケットを中継転送していたとしても、その存在を認識することなくパケット転送を続けてしまうことがあり、逆に、そのような性質を悪用し、他ノード宛のパケット情報を盗聴・改ざんしたりすることが可能となる。このような攻撃は中間者攻撃と呼ばれ、特にアドホックネットワークのように、ノードの移動によってネットワーク構成が動的に変化する場合に中間者攻撃の脅威が大きくなる。そこで、中間者攻撃への対処として、中間者攻撃

を検知するという基本技術について検討を行った。

3.2 内容

アドホックネットワークのように動的に構成が変化する無線ネットワークでは、一つのノードがパケットを直接授受できる相手ノードは2個以上になることが少なくない。その結果、中間者攻撃を行うノードからパケットを受けた場合、攻撃ノードを経由せずに同一のパケットをその直前に送信ノードから直接受信できる場合があり、そのような2重受信によって中間者攻撃を検知する方法を考案した。この方法によって中間者攻撃が検知できるためには、送信ノード・受信ノード・攻撃ノードが互いに電波が届く範囲内にある必要があり、そのような関係を安全三角形と命名した。この原理を図2に示す。一方、安全三角形が形成できない場合は、例えばGPS(Global Position System)によって、各ノードが位置情報を得られるものとし、送出すべきデータと共に位置情報もパケットに含めて転送することとし電波到達可能距離を超えてパケットが届いたときは中間者が存在していたと判定することを可能とした。このように、安全三角形と位置情報を活用した中間者攻撃検知法を安全三角形法と命名した。

この安全三角形法の中間者攻撃検知能力を明らかにするため、机上検討に加え、簡単なシミュレーション実験を行って評価した。その結果、理論的には誤検知確率は常に0であるものの、GPSを利用しない場合にはノード密度や電波到達距離の減少に伴って検知漏れが漸増する等の定量的特性を明らかにした。

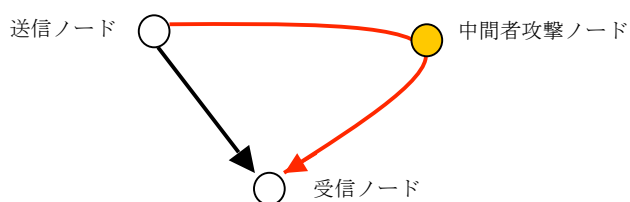


図2 安全三角形法の原理（直接及び中継の2重受信による中間者攻撃の検知）

3.3 具体的成果

無線ネットワークにおける中間者攻撃を検知する安全三角形法の基本原理と評価結果を論文としてとりまとめ学会の研究会及び大会で発表した[発表 2, 5]。今後、GPSによる位置情報測定誤差やノードの移動性を考慮した評価と改善を進めることによって、安全三角形法の実用的な技術確立ができるとの見通しを得た。

4 車車間通信を応用した緊急通信

4.1 背景

今日の車社会においては、交通トラヒックの安全性向上、運転や乗車の快適化、環境への配慮等が強く要求されている。そのうち、安全性向上については情報技術の応用が特に期待されており、近年の無線通信技術、パケット転送技術、衛星応用技術等の発展に伴い、車車間通信も経済的に実現できる見通しが得られるようになってきた。このような見通しの下、特に無線パケット転送技術の応用によって、道路交通トラヒックの安全性向上を図るための研究に取り組んだ。

4.2 内容

道路交通トラヒックの安全性向上を図るため、交通事故や道路状況等の情報を関係する車に逸早く伝達する緊急通信の検討を行った。その目標は、所定の範囲内の車に対して、緊急情報をできる限り確実にかつ迅速に届けることにある。緊急通信の実現には、交通事故や道路状況等の情報を得るためのセンサーを車または道路に設置しておき、そのようなセンサーから得られた緊急情報を、道路に設置した無線基地局または車の中継によって所定の範囲内の車に伝えていくことになる。しかし、設備投資額の観点からは、広範な道路でセンサーや無線基地局を設置することが容易でないため、車にセンサーと緊急情報の中継機能を持たせ、車車間通信を応用することが現実的な解決策となる。

そこで、車車間通信によって、緊急情報をできる限り確実にかつ迅速に伝達する技術の検討を行った。その結果、最近低廉化してきた無線 LAN 技術をそのまま適用するとパケットトラヒックの輻輳に基づく遅延時間の増大が大きな問題となるため、できる限り遠方の車のみ緊急情報を中継するように自律分散的に中継車を選択することによって冗長なパケットトラヒックを低減化すると共に、いったん受けた情報が緊急情報であることを認識した場合は直ちに中継を開始することによって伝達に要する時間を削減する方式を考案した。この方式には、複数の無線チャネルを利用することによって効果を大きくできるという特徴があり、その有効性を明らかにするため、簡単な車走行モデルを仮定してシミュレーション実験を行った。その結果、原理的には、例えば、0.1 秒以内に 1km 以内の範囲の全車に緊急情報を伝達することができ、実用性が達成できるとの見通しを得た。

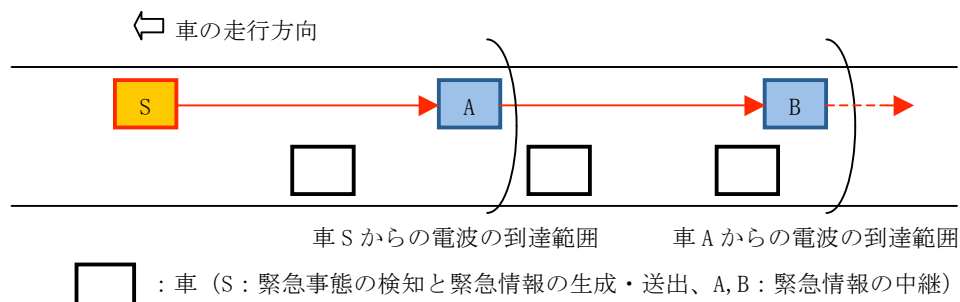


図 3 最小台数の中継車による緊急情報の伝達

4.3 具体的成果

車車間通信による緊急情報の転送法に関する検討結果を論文としてとりまとめ、電子情報通信学会の大会で論文発表した。[発表 7]

また、これまでの検討結果について興味を持ち始めた自動車製造企業と、今後の共同研究の可能性について検討を開始した。

5 次世代衛星ネットワークにおける衛星障害対策

5.1 背景

地震やテロ等の大災害に対する強靱性やグローバル性等の理由から低軌道衛星 (LEO: Low Earth Orbit) 周回通信システムが次世代のネットワークとして有望視されている。パケット転送方式に基づく LEO 衛星通信システムを実用化するためには、様々な技術課題を解決する必要があるが、運用という観点からは、衛星が故障した場合の対策技術を開拓することが必須である。この対策については、衛星コストが小さくないことから、予備衛星を持ってない場合や予備衛星の故障時の対策も必要となるが、このような対策についてはあまり系統的な

研究がなされていないのが現状である。そこで、LEO 衛星通信システムにおいて衛星が障害となった場合の対策技術の検討を行った。

5.2 内容

LEO 衛星通信システムにおいて衛星が故障したときは、故障衛星の近傍で正常に動作している衛星を幾つか移動し、故障した衛星がカバーしていた地上エリアをカバーできるようにするとともに、移動した衛星が持つべき衛星間伝送リンクを再設定する衛星再配置法の検討を昨年度に引き続き進めた。具体的には、衛星の再配置の最適化をまず行い、その結果に基づいて衛星間リンクの最適化を図るという 2 ステップ設計法に基づき検討を行った。第 1 ステップでは、特に地上サービスエリアの最大化を図る方針としたが、移動する衛星個数、サービス停止時間や停止回数等も考慮して総合判断する方針とした。第 2 ステップの設計では、衛星間リンクの距離、ルーティングの容易さ、情報転送容量等を考慮して総合判断する方針とした。

以上の設計アプローチを、我が国で研究開発を進めてきた LEO 衛星通信システムである NeLS (Next generation Low earth orbit Satellite System) を対象として、一つの衛星が故障した場合に適用し、評価を進めた。NeLS は 120 個の衛星と 240 本の衛星間リンクで構成される。評価の結果、故障衛星と同一軌道上の前後にある 2 個の衛星と 8 本の関連衛星間リンクのみを再配置することによって地上サービスカバー率をほぼ 100% に維持した上で衛星間リンク距離やルーティング上のホップ数の増大を抑えることが可能であることを明らかにし、本設計アプローチの有効性を確認した。

5.3 具体的成果

衛星故障時の対策技術に関する検討結果を、衛星分野の専門家が集中的に参加する国際学会 ICSSC で論文発表した。[査読付 1]

以前出願していた特許「地球非静止軌道衛星通信ネットワーク及びその中継装置」については登録がなされた。[特許 1]

6 惑星間インターネットのためのパケット転送制御

6.1 背景

かなり先の将来ではあるが、米国のコロニー計画のように、太陽系における地球以外の惑星での人類の活動が期待されており、それに伴って惑星間で様々な情報を交換するためのインターネットの研究が始まっている。惑星間インターネットにおけるパケット転送では、特に伝播遅延時間が大きく、例えば 10 分のオーダーにもなるという特性を考慮する必要がある。このような場合、地上のインターネットで採用されている転送制御では、適切な転送品質を達成できないことがある。即ち、惑星間インターネットでは新たなパケット転送制御技術の開拓が必要となる。特に、複数ルートのパケットが一つのルートに合流して転送されるような場合に転送誤りの訂正のため再送を行うと、パケットの最終到着までに要する時間が、合流地点の惑星におけるバッファ管理に大きく依存してしまうことがある。そこで、このような転送時間に着目して、適切なバッファ管理を含むパケット転送制御法についての研究に取り組んだ。

6.2 内容

地上系のインターネットを拡張して惑星間に拡大するという惑星間インターネットに関するこれまでの研究の成果として、宇宙環境に適したファイル転送用国際標準プロトコル CFDP (Consultative Committee for Space Data Systems File Delivery Protocol) が開発されている。しかし CFDP では中継

ノードでのバッファ溢れへの対応が規定されていないため、本研究では CFDP を前提として、惑星間インターネットに適したパケット転送制御方式を検討した。その結果、惑星間インターネットを「快適に使う」ためには、ファイル転送に要する総時間 FDT(File Delivery Time) の短縮が不可欠であり、特に伝送誤りが生じた場合の対策として再送を行う場合に、再送制御にかかる時間を最小化する方式が望ましいとの見通しを得て、そのような再送のために送出済みパケットを誤りなく伝送できたことを確認するまで蓄積しておくバッファがあふれた場合の処理方式として、次の3技術を考案した。

- ・既に送出したパケットを優先して廃棄する。この優先制御によって、送出後に伝送誤りが起きると元の送信ノードから送出し直す必要があるが、そのような確率は一般に小さいので、結果的に再送パケット数が少なくなる、という効果が得られる。
- ・到着するまでに要した伝送時間が小さいパケットを優先して破棄する。この優先制御によって、元の送信ノードから送出し直す場合に要する伝送時間が小さくて済むという効果が得られる。
- ・いったん再送することになったパケットと同一のファイルに含まれるパケットを優先して破棄する。この優先制御によって、パケット再送の対象となるファイルが集中し、結果的にそのようなファイルが削減できるという効果が得られる。

これら3技術を適用することで、バッファ溢れが生じる確率を小さくすること、誤りやバッファ溢れに起因するパケット再送の回数を削減すること、及びパケット再送時のパケット転送時間を削減できることによって、総合的な FDT を大幅に短縮できることをシミュレーションによって確認した。

6.3 具体的成果

惑星間インターネットにおけるファイル転送所要時間を削減することが可能なパケット転送制御方式について、電子情報通信学会の研究会と総合大会で論文発表を行った。[発表 4, 6]

7 成果要覧

受賞関連

[受賞 1] 若原 恭: 功労顕彰状, 電子情報通信学会・通信ソサイエティ, 2007年9月.

査読付論文リスト

[査読付 1] Taiki Kasuda, Jun-Ichi Mizusawa, and Yasushi Wakahara: Optimization of the Satellite Rearrangement against Satellite Failure in LEO System, Proceedings of 25th AIAA/ICSSC, AIAA-2007-3297, Seoul, Korea, Apr. 2007.

[査読付 2] Jianping Li and Yasushi Wakahara: End-to-end Bandwidth Guarantee in a Mobile Ad Hoc Network, Proceedings of 8th APRU Doctoral Students Conference, No. 224, Tokyo, Japan, July 2007.

[査読付 3] Jianping Li and Yasushi Wakahara: Time Slot Assignment for End-to-end Bandwidth Guarantee in a Mobile Ad Hoc Network, Journal of Communications, Vol. 2, Issue 6, pp. 56-64, Nov. 2007.

特許出願／取得

- [特許 1] 石井 良和, 湯地 洋子, 原川 孝夫, 若原 恭, 元吉 茂, 森川 栄久: 地球非静止軌道衛星通信ネットワーク及びその中継装置, 特許第 3987903 号 (特願 2002-037486 号), 2007 年 7 月.

その他の発表論文リスト

- [発表 1] Jianping Li and Yasushi Wakahara: Direction-aware Time Slot Assignment for Maximum Bandwidth in Slotted Wireless Ad Hoc Networks, 電子情報通信学会, 技術報告 AN2007-60, pp.63-68, アドホックネットワーク研究会, 2007 年 12 月.
- [発表 2] XiaoYang Zhang and Yasushi Wakahara: Sub-Marine Attack and its Defense in Ad Hoc Network Routing Protocols, 電子情報通信学会, 技術報告 IN2007-120, pp.7-12, 情報ネットワーク研究会, 2008 年 1 月.
- [発表 3] Yasushi Wakahara and Jianping Li: Time Slot Assignment for a Route with Maximum Bandwidth in a TDMA based Ad Hoc Network, Proceedings of International Symposium on Secure-Life Electronics, Tokyo, Japan, pp.223-228, Mar. 2008.
- [発表 4] 久保 淳, 中村 文隆, 若原 恭: 長遅延ネットワークにおいて効率の良いバッファ管理方式, 電子情報通信学会, 技術報告 NS2007-158, pp.155-160, ネットワークシステム研究会, 2008 年 3 月.
- [発表 5] 張 笑陽, 若原 恭: 安全三角形と GPS によるアドホックネットワークの中間者攻撃の検出, 情報処理学会, 全国大会, 3Z-8, 2008 年 3 月.
- [発表 6] 久保 淳, 中村 文隆, 若原 恭: 惑星間インターネットにおけるバッファ管理方式, 電子情報通信学会総合大会, B-7-105, p.182, 2008 年 3 月.
- [発表 7] Akkhara Pakornsiri and Yasushi Wakahara: Alarm Message Broadcasting in Vehicular Ad Hoc Networks (VANETs), 電子情報通信学会総合大会, BS-3-12, pp.S-32-S-33, 2008 年 3 月.

広域分散環境の高度基盤技術に関する研究

中山雅哉

1 概要

Internet に代表される広域分散環境は、今日の社会生活に不可欠な存在となっている。これまでは、計算機を主体としてノードが構成されてきたが、最近では、気象センサーや振動センサーなどの様々な小型デバイスもネットワークに接続される様になり、これらの小型ノードから得られる情報が社会生活に対する安心情報として利用されるようになってきた。一方で、ネットワークに接続されるノードの規模が爆発的に拡大することで、各ノードから得られる情報をデータ統合ノードに集約する際の中継ネットワークの輻輳や遅延に関わる課題や、データ統合ノードに集約された情報を広域に分散した多数の利用者に効率的に提供するための課題を解く技術が不可欠なものとなっている。また、ノード数の拡大に伴って、通常のノードを中継先とする DDoS 攻撃や SPAM などを早期に検知、防止するためのセキュリティ対策技術の必要性も高まっている。

さらに、各ノード間の接続は無線ネットワークの利用が一般的になり、どこでも誰でも簡単に近くのノード間でグループコミュニケーションを行うことができるアドホックネットワーク技術の研究が盛んになってきている。

本研究室では、これらのネットワーク基盤技術の研究に取り組んでいる。

2 自律分散ノード間での効率的な情報流通技術に関する研究

2.1 背景

今日の社会生活に不可欠な存在である広域分散環境 (Internet) には、最近では計算機だけでなく気象センサーや振動センサーなどの様々なデバイスがノードとして接続されるようになってきた。

これらの小型センサーは、自律分散ノードとして観測されたデータを集約ノードなどに常時送信する機能を持っており、集約ノードに集められた情報に基づいて、社会生活に対する安心情報としてネットワーク利用者に提供されるようになってきている。

例えば、気象庁が行っている緊急地震速報は、振動センサーから得られる地震に関する情報を、ネットワークを介して利用者に提供する仕組みの一例として挙げることができる。

2.2 内容

ネットワークに接続される自律センサノードの規模が爆発的に拡大すると、各ノードから得られた情報をデータ統合ノードに集約する際の中継ネットワークの輻輳や遅延に関する問題や、データ統合ノードで得られた集約情報を広域に分散した多数の利用者に提供するための技術が必要不可欠なものとなる。

これまで、国内外におかれた各地の気象センサー情報を集約する研究を行ってきたが、今年度はこれに加えて、3次元加速度センサーを用いて建物の振動情報の連続計測を行うことで都市基盤のリスク管理を行う研究を行っている [発表 1]。

一方、集約ノードに集められた情報を分散した場所の利用者に提供する技術として、皆既日食のインターネット中継 [査読付 1] を通じて、情報提供ノードの負荷分散技術に関する研究を行っている [発表 4]。

また、遠くに離れたノード間で TCP 方式でデータ転送を行うと、帯域・遅延積から、中継ネットワークの帯域を十分に使ったデータ転送を行うことができない問題が生じる。これに対して、本研究室では、エンドノード装置間で行われる一つのデータ転送を、中継ノードを用いて 2 つ以上のデータ転送に分割し、ノード間の遅延を仮想的に短縮させる方法の提案を行ってきた。この方式の有効性を PlanetLab 環境を用いて、世界中に点在したノード間でのデータ転送実証実験で明らかにした [発表 2]。

2.3 具体的成果

この研究では、気象センサーや振動センサーなどの多様な自律分散ノードから得られる情報の集約や日食中継などを通じた全国各地の観測データの共有に関して、幅広く教育・研究の成果が得られている [発表 1, 査読付 1, 発表 4]。

また、広域環境での効率的データ転送方式に関する研究についても、PlanetLab 環境を通して、世界中に点在したノード間でのデータ転送実証実験でその有効性を明らかにした [発表 2]。

3 安全で安定した社会基盤を構築するネットワーク技術の研究

3.1 背景

インターネットに接続されるノード数の拡大に伴って、通常のノードを中継先とする DDoS 攻撃や SPAM などの新たな問題が発生してきている。社会基盤として安全で安定したネットワークを構成するためには、これらの攻撃を早期に検知、防止するためのセキュリティ対策技術の開発が広く望まれている。

また、最近では、無線ネットワークの利用が一般的になっており、ノート PC をはじめとして、携帯型の端末を持ち歩きながら、どこでもネットワークに接続して利用できる環境が都市部では当たり前になってきた。しかし、無線 LAN の基地局がカバーする範囲は狭いため、移動しながら通信ができるネットワーク環境がどこでも構築される状況には至っていない。

そこで、近くの無線ノード同士がお互いに他のノードの通信を中継し合うことで、簡便に近隣のノード間でグループコミュニケーションを行うことができるアドホックネットワーク技術の研究が盛んに行われるようになってきた。しかし、グループ内の通信といえども他者の通信の中継を通常のノードが担うことから、不審者による通信妨害の検知や回避は不可欠であり、その解決は重要な課題となっている。

さらに、離島や山間部などでは地上網の整備が困難であり、衛星等を用いて社会基盤ネットワークを構築する必要性がある。

3.2 内容

本研究室では、安定した社会基盤を構築するネットワーク技術として、以下の 3 つのセキュリティ対策技術に関する研究と無線通信の効率化に関する研究を行っている。

セキュリティ対策技術に関する研究の 1 つめは、正常な通信に紛れる様な低レートで連続して行われる DoS 攻撃や、正常に動作する一般のノードに被害者のアドレスを送信元と偽装してパケットを送信するリフレクタ型 DDoS 攻撃の様に、従来の方式では検知が困難であった本来の攻撃元を、被害者側から容易に直接調べられるように改良した traceback 方式の研究である [査読付 3, 発表 3, 発表 7]。

2 つめは、利用者が望まない SPAM などの通信を排除する方式として、各利用者が主体となって、通信相手毎に利用するコミュニケーション手段に対する ID へのアクセス権限を管理する機構の提案である [発表 8]。

3 つめは、無線ノード同士が他のノードの通信を中継するアドホックネットワークにおいて、不審ノードの参加を早期に検知する方式に関する研究である [発表 6]。

また無線通信の効率化に関する研究は、通信の中継を行うノード間で協調制御を行うことができるメッシュネットワークにおいて、隣接ノードが扱っているデータフローの情報を集めて、他のフローとの干渉関係を考慮したタイムスロットの割当てを行う同期式 MAC 制御方式に関する研究である [査読付 2]。

3.3 具体的成果

セキュリティ対策技術のうち、正常な通信に紛れる様な低レートで連続して行われる DoS 攻撃や、一般のノードを中継させるリフレクタ型 DDoS 攻撃に対して、本来の攻撃元を被害者側から容易に直接調べられるように改良した traceback 方式の研究 [査読付 3, 発表 3, 発表 7] は、第 15 回 DPS Workshop において、学生奨励賞 [受賞 1] を受賞する成果を得ることができた。

国内外の各大学で行っている衛星を用いた社会基盤の構築と高等教育での利活用 SOI/SOI Asia project に関して APRSAF 国際会議で紹介した [発表 5]。

また、「高等教育機関の情報セキュリティ対策のためのサンプル規定集」の策定にも関わり、情報セキュリティの日功労者表彰 [受賞 2] を受賞した。

4 成果要覧

受賞関連

[受賞 1] 高田友則: 学生奨励賞, 第 15 回マルチメディア通信と分散処理ワークショップ実行委員会, 2007 年 11 月.

[受賞 2] 国立情報学研究所 (国立大学法人等における情報セキュリティポリシー策定作業部会) 及び電子情報通信学会 (ネットワーク運用ガイドライン検討ワーキンググループ): 情報セキュリティの日功労者表彰, 内閣官房情報セキュリティセンター, 2008 年 2 月.

査読付論文リスト

[査読付 1] 尾久土正己, 荻原文恵, 小澤友彦, 吉住千亜紀, 富田晃彦, 山田宏之, 明井英太郎, 石川雅一, 山本文治, 中山雅哉, 半田利弘: プラネタリウムにおける日食の全天周生中継, 地学教育, Vol. 60, No. 3, pp. 99-107, 2007 年 5 月.

[査読付 2] Kentaro Saito and Masaya Nakayama, The Spatial Reuse Slot Allocation with Reallocating Neighbor Links in the TDMA-Based Wireless Mesh Network, Proc. of IEEE VTC 2007-Fall, Sep.-Oct. 2007.

[査読付 3] 高田友則, 中山雅哉: 低レートの DoS 攻撃に対応する改良型 ICMP Traceback, 情報処理学会 DPSWS2007, 2007 年 10 月-11 月.

その他の発表論文リスト

[発表 1] 中山雅哉: インターネットの広域災害対策への活用- 安否確認システムとセンサユニットによる都市基盤リスク管理 -, 地域防災情報シンポジウム in 岩手, 2007 年 4 月.

[発表 2] 小林弘和, 中山雅哉: 中継ノードを利用した長距離 TCP データ転送の効率向上と中継ノード発見手法の比較, 信学技報 Vol. 107, No. 74, IA2007-4, pp. 17-22, 2007 年 5 月.

[発表 3] 高田友則, 中山雅哉: 改良型 iTrace 手法 (iTrace-PT 手法) の反射型 DDoS 攻撃への適用とその効果, 信学技報 Vol. 107, No. 76, IA2007-35, pp. 27-31, 2007 年 10 月.

- [発表 4] 工藤めぐみ, 井上博之, 中山雅哉: 低負荷な多チャンネルストリームコンテンツの選択方式に関する検討と実装, 情報処理学会 DPS 研究会, Vol. 2007, No. 117, pp. 31–26, 2007 年 11 月.
- [発表 5] Masaya Nakayama: SOI/SOI Asia Activities, Proc. of APRSAF14, Nov. 2007.
- [発表 6] Muhammad Arifin Ritonga and Masaya Nakayama: Evaluation of Manager-Based Architecture for Intrusion Detection System in Ad Hoc Networks, 信学技報 Vol. 107, No. NS-524, pp. 279–284, Mar. 2008.
- [発表 7] 高田友則, 中山雅哉: 偽装 iTrace パケットによる影響とその対策, 信学技報 Vol. 107, No. NS-524, pp. 341–346, 2008 年 3 月.
- [発表 8] 阪本裕介, 中山雅哉: OpenID を用いたユーザ指向型ホワイトリスト作成手法の提案, 情報処理学会 GN 研究会 Vol. 2008, No. 31, pp. 73–78, 2008 年 3 月.

ネットワーク運用に関する研究

加藤 朗

1 概要

コンピュータネットワークは、それを構築しただけでは実用目的に活用することはできず、きちんと運用管理されていることが必要である。通常ネットワーク管理では ping 等による機能監視や snmp などによる状態監視が有用であるが、教育・研究ネットワークの場合には、帯域を論理的に分割する VLAN や、さらに分割の単位を layer-1 まで押し下げた Lambda Network のような利用形態がとられることがある。これは、一般のインターネットでは十分なセキュリティ上の対策をとることが求められるが、VLAN や Lambda Network を用いた場合、限られた範囲の機材しか接続することができないため、セキュリティ上の対策を軽減、あるいは場合によっては考慮しなくてすむ。セキュリティ対策は多くの場合、パケットの検査が必要で、そのための機器のコストのみならず、それらによって発生する可能性があるパケット損失、ジッタなどを回避できるため有用である。ここでは、これらに関する以下の研究活動の概要について報告する。

- 広帯域ネットワーク運用に関する研究

帯域を静的にあるいは準動的に割り振り、帯域を保証するネットワークサービスは lambda networking と呼ばれている。これは対地まで layer-1 や layer-2 の回線を設定し、ユーザはその帯域を占有できるようなサービスであるが、このようなネットワークの構築や運用は通常のインターネットのそれとは異なる側面がある。これらについて、実際のネットワークの運用管理を通じて必要な機能やユーザのサポートなどの面について検討する。

- 研究・教育ネットワークにおける経路制御の運用に関する研究

研究・教育ネットワークは特に 2005 年末にアジアとヨーロッパを直接接続する目的で運用が始まった TEIN-2 によってその相互接続のトポロジは一層複雑になった。また、研究・教育ネットワークは営利目的で運用されているわけでないため、機器や回線障害によって発生する障害に関しては、商用ネットワークより有機的に相互にバックアップしあうことが日常行われている。このため、しばしば宛先への経路が最適ではないリンクを経由することも多く発生している。ここでは、各研究・教育ネットワークのポリシーを尊重しつつ相互バックアップが可能な経路制御方式について考察する。

- DNS サーバ運用に関する研究

DNS はインターネットを利用する上で欠くことができないサービスであり、ネットワークアーキテクチャ的にはアプリケーションの一つに過ぎないが基盤サービスとして位置づけられており、その可用性は重要である。ここでは Anycast による運用方式や IPv6 への対応について考える。

2 広帯域ネットワーク運用に関する研究

2.1 背景

最近の大規模科学では、実験装置や観測装置が生成した大量のデータを研究者がネットワーク経由で入手し、研究を行うスタイルが多く見受けられるようになった。これは特に実験装置や観測装置が観測上の問題や経済的問題で限られた地点にしか設置できないことによるが、特に国際プロジェクトではデータの消費者である研究者との間の距離は非常に大きくなる傾向にある。このような長距離大容量のデータ転送を効率的に実施するのはそれほど容易なことではない。特に世界的な情報基盤として確立しているインターネットは、その帯域は他の種々の活動と共有しているため、帯域の占有は公平でなければならないなどの制約がある。また多くのネットワーク機材はこのような長距離大容量データ転送を前提として設計されているとは限らないため、バッファ容量不足によるパケット損失などが発生し、転送効率が著しく低下する可能性もある。

また、実時間高品位画像転送のようなアプリケーションを考えた場合、再送を実施する時間的余裕はないため、パケット損のないネットワークが必要になる。特に対話型アプリケーションを考えた場合、圧縮・伸張に必要な時間を節約するため非圧縮な動画転送が必要になるが、この場合、HDTVで1.5Gbpsの帯域が、また4Kデジタルシネマの場合には6Gbpsの帯域が必要になる。

このような、少数のユーザが長距離大容量のデータ転送をパケット損を最小にかつ効率的に実施するためには、インターネットのような帯域を共有するネットワークサービスではなく、同じTCP/IPを使ったとしても帯域的には別のネットワークを利用することが考えられる。対地が固定している場合にはOC-192cあるいはSTM-64などの9.6Gbps専用線を調達すればよいが、予備実験やデモンストレーションなどの短期間の利用に専用線を調達するのは効率的ではない。

近年、多くの研究・教育ネットワークでは、予算の伸びを上回る帯域需要が予期されているため、オランダのSURFnetやアメリカのNLR (National Lambda Rail)、スカンジナビア各国を接続するNORDUnet、ヨーロッパの主要各国を接続するGeantでは、近海および陸上区間に関しては光ファイバーケーブルの芯線借りを行い、DWDM装置を用いることで9.6Gbpsあるいは10Gbpsの回線を多数得ることができ、必要な帯域を得られるように確保している場合も少なくない。

これらのネットワークを相互接続することができれば、単一の研究・教育ネットワークでは実現できなかった長距離広帯域専用ネットワークを設定することが可能になる。この趣旨に賛同したネットワークはGLIF¹ – Global Lambda Integrated Facility – という活動を共同で行っている。当然そのネットワークの主要ユーザに優先権があるが、空いている時間帯は短期間の利用に関しては請求書の発行をせずに帯域を総合に利用できることが申し合わされている。この枠組みを用いることで、単一研究・教育ネットワークの地理的範囲に留まらない範囲で専用の帯域を確保することができ、各種実験やデモンストレーションに数多く活用されている。

2.2 内容

IEEAF 太平洋回線はSeattle市内のデータセンターにあるPacific Northwest Gigapop²とNTT大手町ビルを接続するOC-192cの回線であり、Tyco Telecom (現在はVSNLを経てTata Communications)から提供を受けている。東京側はONS-15454を経由してNI-40Gの10GE WANPHYポートに接続されている。このONS-15454、NI-40G、および低速系を収容するNI15000は東京に於けるGLIFの拠点の一つであるT-LEXを形成している。我が国の研究・教育ネットワークは、市内区間を除くとDark FiberやWDMを使用していないため、国際回線のOC-192c SONET回線をそのまま交換することはできない。そのため、NI-40Gで10GE LANPHYに変換し、layer-2で交換を行っている。

¹<http://www.glif.is/>

²<http://www.pnw-gigapop.net/>

一方 Seattle 側は、Pacific Northwest Gigapop に設置された CANARIE 所有の HDXc を経由して RX-4 に 10GE WANPHY によって接続されている。RX-4 は Pacific Wave の Catalyst6500 に 10GE LANPHY で接続されているため、これを經由して US 国内への layer-2 による接続が可能である。さらに、利用許可やスケジュールの調整は必要であるが、Seattle - Chicago 間の NLR の一波を用いた TransLight を用いて Chicago の StarLight に延し、そこからヨーロッパ方面へと回線を延伸することも可能である。

GLIF では毎年 2 日間のワークショップを開催し、計画されているあるいは終了した実験やデモンストレーションの報告をし、また相互接続やトラブルシュートに必要な情報交換を行っている。2007 年は Czech の Prague で 9 月に開催された。また GLIF の技術系のワーキンググループは、年に一度では不十分であるため、2008 年に Joint-Techs や APAN などが Honolulu で共同で開催した TIP2008 の一貫として中間会合を行っている。また、各ネットワークの帯域を相互接続することが GLIF の特徴であり、相互接続を円滑に提供することが重要である。そのため、これらの相互接続地点を GOLE - GLIF Open Lambda Exchange - と呼び、GOLE の担当者はさらに電話会議を定期的に行い、特にデモンストレーションが集中する SC07 などの会議に備えている。

また Nevada 州 Reno 市で開催された SC07 では、会場と我が国を接続できる広帯域ネットワークの状況について確認し、トラフィック集中が懸念されていた区間の利用者の時間割り当て方針について確認を行った他、IEEAF 太平洋回線を用いて会場と東京大学理学部旧一号館を 10Gigabit 級ネットワークで接続を行い、Data Reservoir チームによる並列 TCP 転送実験に貢献した。

2.3 具体的成果

2006 年末に本学情報理工学系研究科平木教授が率いる Data Reservoir チームによる Internet2 Land Speed Record <http://www.internet2.edu/lsr> を IPv4 に引き続き IPv6 でも更新したのに伴い、Prague で開催された GLIF の会合である Global Lambda Grid Workshop で、特にネットワーク面からの状況について講演を行い、回線等を提供して頂いた各ネットワークおよびその担当者に謝意を表すことができた。また翌月には、「情報化促進部門」として総務大臣表彰を受けている。

Lambda Networking は効率的なデータ転送を行うためには有効な技術であるが、一方低い layer でのサービスであるため、回線を確立する際に障害が発生した場合、多くの場合障害は設定ミスに関連することが多いが、その障害個所を特定するのは困難である。そのため、機器の設定や送受されたパケット数やエラーの発生頻度などの情報が web 等の手段でユーザに公開されていれば、障害個所の特定に有効である。また、全く通信ができない場合はともかく、非常に低い割合でパケット損が発生している場合、回線やネットワーク機材に起因するものか、受信ホスト内部でパケットを紛失しているのかを区別することは非常に難しい。そのため、受信端(もしくは回線の途中)に光分岐器を予め設置しておき、通信に影響を与えないようにパケットを観測し、TCP/RTP のシーケンス番号などをチェックすることは、Lambda Network をサポートする上で重要であると考えられ、現在 FPGA を用いた観測機材の開発を行っている。

3 研究・教育ネットワークにおける経路制御の運用に関する研究

3.1 背景

研究・教育ネットワークは商用 ISP とは異なり、直接の顧客関係にない場合でも通過することが可能であったり、あるいはバックアップとして機能することを双方で確認しているケースなど、同じ BGP で経路交換を行っているとは言え、商用ネットワークの場合と状況は異なる。特に最近では相互接続が密になり、2 地点間に存在する可能な経路は複数ある場合も少なくない。またそれ以外に商用 ISP 経由の接続性もある場合も多く、これらを正しく使い分けるのは容易ではない。特にバックアップを考えた場合、障害が発生しても意図通りの経路制御を行うような挙動を各ルータに設定するのは、容易

ではなく、複雑な設定が必要になる場合が多い。

現在の BGP はこのような要望を簡単に解決できる枠組みを提供していないため、定常的な経路の監視を怠った場合、経路が著しく遠回りになってしまうことも珍しくない。実際に東京大学と韓国の KREONET2 との通信がシンガポール経由になっていたこともある。このような状況に対して、経路制御ポリシーを明確にし、それに対応した設定を自動生成するなどの対策が必要になっている。

3.2 内容

研究・教育ネットワークに関しては、グローバルな相互接続が実施されているにも関わらず、これらの問題を共通に議論し解決する枠組みがなかった。2006 年には、関係する研究・教育ネットワークのオペレータのフォーラムである RENOG が設立され、Internet2 のメンバ会合の際にミーティングが開かれるようになった。

RENOG の一つの議題は複雑な経路制御ポリシーを実装するための方法論であり、各研究・教育ネットワークの通過ポリシーの明確化およびその定式化が APAN-JP によって行われている。

これに並行して経路の広告ポリシーの扱いに関して検討を行った。一つの可能性は、BGP Community を用いて、各経路の発信元で、その経路に対する取り扱いの希望を記述することである。ただし、冗長な経路を避けるためには、特定の研究・教育ネットワークから受信した場合には経路を採択する、さらに広告するなどの指定も必要になるが、比較的フレキシブルな経路の取り扱いに関する要望を記述することができ、実際の経路情報の収集解析機構と併用することにより、バックアップを含めて不適切な経路が発生した場合には発信元によってある程度の修正が可能になる。

3.3 具体的成果

提案中の方式は、BGP Community を用いて、対象研究・教育ネットワークにおいて、指定した隣接研究・教育ネットワークから経路を受信した場合の挙動を、他に優先して採用する (Local_pref を通常より大きくする)、他と同等に扱う、他があればそちらを優先する (Local_pref を通常より小さくする) などの指定を可能にする。また、その研究・教育ネットワークが指定した研究・教育ネットワークに経路を広告する際の指示として、広告しない、AS 番号を複数回 ASPATH に追加して広告する、そのまま広告する、などの指示も可能である。

この方式は、操作対象となる研究・教育ネットワークに加えて、経路の受信元あるいは広告先の隣接研究・教育ネットワークを指定することが必要になる。現在この目的に用いられている 16bit の AS 番号を BGP Community に二つ指定することはできない。そのため、8bit 程度の研究・教育ネットワーク番号の割り振り・管理が必要になる。この場合、残る 8bit 程度の情報によって受理や広告の挙動を指定することができる。

この方式に関しては、RENOG ミーティングで提案を行い、関係者と議論を行った。またこの BGP Community 値に対応した動作をするようにオペレータがマニュアルで設定作業を行うことは、かなり複雑なものになることが想定されている。これを解決するため、ACL や Local_pref などを用いたルータの設定の自動生成の枠組みに関して検討を進めている。

4 DNS サーバ運用に関する研究

4.1 背景

インターネット上のいろいろな資源に名前をつけ、それに対応した IP アドレスなどの情報を得る仕組みは現在は DNS - Domain Name System - で定義されている。DNS はインターネット上で提供されているアプリケーションの一つに過ぎないが、ユーザがインターネット上の資源を陽に指定する殆どどの場面で DNS で定義された名前を用いている。そのため、DNS はインターネットを有用たら

しめている基本的な基幹サービスとして分類されている。そのため、DNS が安定に可動していることは、インターネットを実用目的で利用する場合には非常に重要である。

DNS では、名前解決を行う際には、数回の問い合わせ・応答プロセスを経て目的の名前に対応する必要な情報を得ることができるようになっている。この際、cacheに残っている以前の検索結果を可能な限り利用して、名前解決に必要なメッセージ数の削減、名前サーバへの問い合わせ数の軽減、および最終的な情報が得られるまでの時間の短縮につとめている。

DNS では、DNS サーバの負荷の低減や応答が得られるまでの時間の削減のために、極力 TCP ではなく UDP を用いることになっている。フラグメントを避けるために、DNS では UDP によるメッセージ長を従来は 512byte に制限しており、これが同一ドメインに対する DNS サーバ数の上限を決定している。この制約は、IPv4 に加えて IPv6 をサポートしなければならない状況ではより顕著である。そのため、EDNS0 という UDP のメッセージ長の制約を緩和する拡張が提案、実装されているが、現実の問い合わせのうち、EDNS0 に対応したものは半分程度とあまり多くはないのが現実である。

Root や TLD - Top Level Domain - のように他に比べて重要なノードを担当する DNS サーバは、特に高い可用性が要求されるが、上記の制約により設定できる DNS サーバ数には限界があり、単一サーバを負荷分散器を併用したサーバクラスタで実現することによる能力改善が唯一の方法であった。これに対し、インターネット上の異なった地点から同一経路情報を広報し、問い合わせを最も近くのサーバに誘導する方式が注目され、RFC3258 として定義されている。この方式は、一般には Anycast と呼ばれており、複数のサーバが異なった地点で並行稼働していることから、単に処理能力の向上に留まらず、可用性を向上し、また経路制御システムでは遅延が経路判定の材料に用いられていないために限界はあるものの、応答時間の改善も期待できることで、これらの上位ノードに対応したサーバで用いられている。

4.2 内容

IPv4 アドレスの未使用空間が徐々に減ってきており、2010 年頃には売切れる可能性が指摘されている今日この頃、IPv4 アドレス利用を抑制あるいは不要な IPv4 アドレスを再利用することは重要であるが、利用中のアドレスの変更にはコストのかかるアドレス変更が必要になるため、IPv4 アドレスの多少の延命措置としては機能するものの、究極には IPv6 に移行することが必要であると考えられており、これは DNS に対しても例外ではない。DNS における IPv6 アドレスを表現する AAAA レコードが最初に RFC1885 として定義されたのは 1995 年であり、現在では殆どどの DNS サーバの実装においてサポートされている。一方、IPv6 を用いた名前解決には、DNS サーバのアドレスを従来は IPv4 のみだったのに対して IPv6 を追加する必要がある。

TLD 以下の DNS の IPv6 対応は、IPv6 による問い合わせをサポートした DNS サーバ、つまり bind9 や bind8.4 を用いることに、2001 年頃から可能だった。それに対して TLD サーバでは、Root ゾーンに変更が必要なことから若干時間がかかり、.JP や .KR などの TLD サーバの一部に AAAA レコードが追加されたのは 2004 年になってからであった。一方 Root DNS サーバに関しては、パケット長などの問題に関連し、Root DNS サーバ運用者間では 2000 年頃から議論されてきたものの、なかなか実現できなかった。これは、その変更に対して最終的な責任をもつ ICANN/IANA が比較的保守的な運用をしていたことも一因である。

インターネットとは分離された実験室でのテストでは、良く知られた幾つかの DNS の実装に関しては問題がないことは早いうちから示していたが、Root DNS サーバの名前とアドレスを列挙したパケットが 512byte 以下に収まるためには、追加できる AAAA レコードは僅かに二つと、実用上は不十分だった。このため、anycast を用いることによって、サーバ数の事実上の拡大と IPv6 ネットワークのトポロジが IPv4 のそれに比べてまだかなり粗いことに対応する必要があると考え、M-Root DNS では、東京、San Francisco、Paris と、インターネットユーザが密な北米、ヨーロッパ、アジア太平洋での IPv6 対応可能なサーバの展開を 2004 年から行ってきた。また、EDNS0 によるメッセー

ジ長の制約の緩和も期待された。

これらの状況および IPv4 アドレスの残余の枯渇が次第に深刻になってきたこともあり、2007 年 1 月に ICANN から sac018 として Root DNS サーバにおける IPv6 対応の方向性が示された。その後、2007 年末に ICANN 理事会によって承認が得られ、2008 年 2 月 4 日 (アメリカ太平洋時間) にそれまでに申請があった 6 つの Root DNS サーバの AAAA レコードが追加されることになった。

M-Root DNS サーバでは、基本的な準備は完了していたため、2007 年 8 月に AAAA レコードの追加申請を行っており、ICANN 理事会の決定後、経路制御や監視系などの若干の追加作業を行い、最初の 6 つの Root DNS サーバの一つとして IPv6 による DNS サービスの提供を、IPv6 がサポートされていない Seoul 以外の 3 つの anycast ノードで³開始した。

4.3 具体的成果

2008 年 2 月 4 日以前にも M-Root DNS サーバはその IPv6 アドレスを公表していたため、合計で毎秒 2 パケット程度の問い合わせを定期的に受けていた。これらのうちの相当分は、実際の DNS の問い合わせではなく、DNS のパケットを送信して行う観測活動によるものであることが判明している。実際に AAAA レコードが追加されたときには、比較的短い時間でこの数字は毎秒 50~100 程度に増加した。M-Root DNS サーバにおける全問い合わせは毎秒 13,000 程度であるから、IPv6 による問い合わせは全体の 0.5% 程度となる。これは、幾つかの IX の IPv4 と IPv6 のトラフィックの比が 0.1% に達していないことを考えると、絶対量としては大した量ではないが、相対的には通常のトラフィックの比率を凌いでいる。

この理由として、DNS のトラフィックは殆んどが単一の問い合わせに対して単一の応答を返す形で終了しており、ストリームや P2P などの大量のデータのやり取りを行うサービスではないこと、また、問い合わせを送る側が IPv6 に対応しており、IPv6 での接続性がある場合には、IPv6 をまず試す、という実装が多いことに起因していると考えられる。

IPv6 化を行う前には、問い合わせが増加する、TCP による問い合わせが増える、などを懸念する向きもあった。M-Root DNS サーバは、DNS OARC⁴ – Operation, Analysis, and Research Center – が主催する “Days in the life (DITL) プロジェクト” として、共通な一定時間のトラフィック情報の同時収集に 2007 年から参加しており、2008 年の DITL は、日本時間では 3 月 19 日から 21 日に掛けて実施された。このデータの一部を 2007 年 1 月 8 日から 10 日のものと比較すると、IPv6 化による影響をある程度解析することができる。2007 年に関しては TCP による問い合わせ数は全体の 0.53% あったのに対して、IPv6 化が実施されたあとの 2008 年のデータでは 0.41% となっている。また、ある特定の 1 時間に TCP による問い合わせを送信した発信元 IP アドレス数は 2007 年は 58,388 だったものが、2008 年では (ただし同時刻) 45,644 と減少している。

これらの数字は Root DNS サーバ全体のものではないため、これから直ちに IPv6 化によって TCP による問い合わせは 3/4 に減少したとすることはできないが、IPv6 化によって懸念された TCP による問い合わせは、目だった増加はしていないことは推定できる。

また、M-Root 全体のトラフィックをみた場合、およそ半分が EDNS0 に対応した問い合わせであったが、IPv6 による問い合わせは約 92% が EDNS0 に対応していることが分かっている。100% になっていないのは、測定用の問い合わせが EDNS0 に対応していないことも含まれているが、これらによる誤差は僅かであり、IPv6 には対応しているが EDNS0 に対応していない DNS クライアントが相当数あることが想定されているが、詳細は今後の調査が必要である。

³正確には東京には anycast ノードは 3 つの主要 IX に対応して 3 つあるので、5 つと言うべき

⁴<https://oarc.isc.org/>

5 成果要覧

査読付論文リスト

[査読付 1] 小原泰弘, 今泉英明, 加藤朗, 中村修, 村井純: 広範なトラフィック要求に対応する負荷分散経路計算アルゴリズム, 情報処理学会論文誌, Vol. 48, No. 4, pp. 1627-1640, 2007.

招待講演 / 招待論文

[招待 1] 平木敬, 稲葉真理, 菅原豊, 吉野剛史, 玉造潤史, 加藤朗: Internet2 Land Speed Record 長距離 TCP 通信高速化への挑戦, 情報処理学会誌, Vol. 49, No. 2, pp. 55-62, 2008.

受賞関連

[受賞 1] データレゼボワール IPv6/IPv4 TCP データ転送チーム: 総務大臣表彰「情報化促進部門」, 総務大臣, 2007年10月.

その他の発表論文リスト

[発表 1] Akira Kato: The WIDE Internet, WIDE-KT workshop, Seoul, Korea, Apr. 2007.

[発表 2] 加藤 朗: GLIF と相互接続, プレスコンファレンス, 東京大学情報理工学系研究科, 東京, 2007年5月.

[発表 3] Akira Kato: Lambda Network Operation, Light Path to India Workshop hosted by the Government of India and Keio University, New Delhi, May 2007.

[発表 4] Akira Kato: T-LEX : A Lambda Facility in Tokyo, Light Path to India Workshop hosted by the Government of India and Keio University, New Delhi, May 2007.

[発表 5] Akira Kato: JP R&E Networks, AI3 Workshop, Phnom Penh, Cambodia, May 2007.

[発表 6] Akira Kato: CJK Collaboration on Future Internet Research Activities, The 2nd Future International Workshop, Seoul, Korea, July 2007.

[発表 7] 加藤 朗: How the Internet can survive?, 第 20 回 Japan Network Operators Group ミーティング, 帯広, 2007年7月.

[発表 8] Akira Kato and Tomoki Murota: Microscopic Traffic Analysis, 8th CAIDA-WIDE Workshop, Chicago, Illinois, July 2007.

[発表 9] Akira Kato: T-LEX Update, Lambda BOF, The 24th APAN Meeting, Xi'an, China, Aug. 2007.

[発表 10] Akira Kato: Internet2 Land Speed Records in the GLIF environment, 7th Annual Global LambdaGrid Workshop, Prague, Czech, Sep 2007.

[発表 11] Akira Kato: Toward Collaboration of Future Internet Research, Core University Seminar, Cheju, Korea, Oct. 2007.

[発表 12] 加藤 朗: Root DNS サーバ, DNS-Day, IP Meeting 2007, 東京, 2007年11月.

[発表 13] 加藤 朗: 移行の科学, 第 12 回 分散システム/インターネット運用技術シンポジウム- 柔らかなサービスを支える技術 -, 情報処理学会分散システム/インターネット運用技術研究会, 東京, 2007年11月.

- [発表 14] Akira Kato: What we are doing on the other side of the globe, Seattle Workshop of IEEEAF Africa Global Quilt Initiative, International Educational Equal Access Foundation, Seattle, Washington, Nov. 2007.
- [発表 15] 加藤 朗: IPv6 Operation in Root/ccTLD Servers, 第 40 回 DISTIX 全体ミーティング, 東京, 2008 年 1 月.
- [発表 16] Akira Kato: T-LEX Update, GLIF Tech/Contron-plain Interim Meeting, Honolulu, Hawaii, Jan. 2008.
- [発表 17] 加藤 朗: Root DNS への AAAA レコードについて, 第 21 回 Japan Network Operators Group ミーティング熊本, 2008 年 1 月 (国際電話による遠隔講演).
- [発表 18] Akira Kato: IPv6 Operation in Root Servers, International Concerence on Computer Applications 2008, University of Computer Studies, Yangon, Yangon, Myanmar, Feb. 2008.

放送型データ配信を用いた情報システム構築に関する研究

小川 剛史

1 概要

放送型データ配信は同報性に優れ、多数の相手に同一の情報を同時刻に提供することができる。この放送型データ配信を応用し、多数のクライアントの存在が想定される情報システムの構築に関して研究を推進した。具体的なテーマは以下の2つである。

- 放送型データ配信を用いた大規模仮想空間システム構築[査読付 5]
- 放送型データ配信を用いた自律移動型センサネットワーク[査読付 2, 発表 5]

また上記以外に、協調作業環境に関する研究[査読付 4, 発表 1, 2]や3次元ユーザインタフェースに関する研究[査読付 3, 発表 3, 7]、人のコンテキスト抽出に関する研究[査読付 1, 発表 4]、教育用計算機システム構築に関する研究[査読付 6]を行った。以下では、放送型データ配信に関連する2つのテーマに関して詳細に述べる。

2 放送型データ配信を用いた大規模仮想空間システム構築

2.1 背景

人々のコミュニケーション支援や行動解析、軍事シミュレーション、多人数型オンラインロールプレイングゲーム(MMORPG)など、ネットワーク上で利用する共有仮想空間システムが多数存在している。これらのシステムの多くは、クライアント・サーバ型のアーキテクチャを用いて構成され、サーバが仮想空間を構成するモデルデータや、仮想空間内で発生するイベントに関するデータをすべて管理し、クライアントがデータを要求した際にサーバが必要なデータを提供する。MMORPGのように数千人が同時に参加するようなシステムでは、クライアント数の増加に伴うサーバ負荷の増大や、大規模な仮想空間に関する情報を取得する際のクライアント・サーバ間のトラフィックの増大などが問題となっている。

これらの問題を解決するため、プッシュ型放送を用いてクライアントへ仮想空間の構成データを提供する放送型サイバースペースを提案してきた。プッシュ型放送では、クライアント数が増加してもサーバにおいてデータを放送する部分に関して変更を必要としないため、多数のクライアントが仮想空間を利用してもサーバにおけるスループットの低下が少ない。さらにクライアントでは、必要なデータを必要なときに取得するだけでよく、常に放送されるデータを監視する必要がないため、データ受信のためのオーバーヘッドはほとんどない。しかし一方で、クライアントにとって必要なデータが必要なときに放送されるとは限らないため、待ち時間が発生するという問題点がある。これまでに、まず比較的更新が頻繁には行われない仮想空間モデルの配信に注目し、仮想空間を複数の部分空間に分割して、部分空間ごとのモデルをデータアイテムとする放送スケジューリング手法を提案した。提案手法では仮想空間内を行動するアバタの移動に基づいてスケジューリングすることで、必要なモデルデータをクライアントが受信するまでの待ち時間を削減した。

仮想空間を構成するデータには、空間構造を決定するモデルデータと、ユーザが操作するアバタの位置や属性、アバタ同士の会話に関するデータなどユーザのインタラクションによるデータが存在する。このようなインタラクションに関するデータは、仮想空間における一貫性を保持するためにも、

即座に全クライアントへ配信される必要があるが、これまでに提案している手法ではこれらのデータを対象としていなかった。そこで、本研究では、放送型サイバースペースにおいて、アバタの位置など更新頻度の高いデータを提供するためのスケジューリング方式について検討を行った。

2.2 内容

想定するリアルタイムデータは、オブジェクトの座標や、形状、テクスチャといった実データもしくはその差分データで、オブジェクトがアバタの場合には、テキストや音声による会話データも含む。これらのデータをオブジェクトごとにデータアイテムとしてまとめ全クライアントへ放送する。これまでに提案した放送型サイバースペースでは仮想空間

の部分空間を構成するモデルデータをひとつのデータアイテムとしてスケジュールを決定し、繰り返し放送していた。本研究では、図 1 に示すようにモデルデータを配信するための帯域と、リアルタイムデータを配信するための帯域に、全帯域を分割して利用することを想定している。

放送型サイバースペースでは、多数のクライアントのビューに投影されるオブジェクトのデータを配信すると、1回の放送で多数のクライアントのビューを更新できるため、多くのアバタ(ユーザ)から見られている(人気度の高い)オブジェクトの更新頻度を上げることが、システム全体での更新頻度向上につながる。各クライアントでは、アバタの近くに存在するオブジェクトは、移動させたり、話しかけたりといったインタラクションが発生する可能性が高いため、一貫性を保つためにも常に最新の状態で表示される必要がある。また、近くのオブジェクトは、ビュー内でも大きく表示されるため、視覚的にも重要なオブジェクトであると言える。そこで、オブジェクトの人気度とアバタとの距離からオブジェクトの重要度を算出し、その重要度に基づいて放送するオブジェクトを決定するPIDR(Popularity/Distance regarding Remaining time)方式を提案した。

PIDR 方式では、式(1)を用いてオブジェクト*i*の重要度 $Priority_i$ を算出し最も重要度の高いオブジェクトのデータを放送する。

$$Priority_i = \frac{P_i}{D_i R_i} \cdot balancer \quad (1)$$

P_i はオブジェクト*i*の人気度でオブジェクト*i*を視野におさめているアバタ数を、 D_i はオブジェクト*i*に最も近くにいるアバタとの距離を、 R_i はオブジェクト*i*のデッドラインまでの残り時間を示している。オブジェクトのデッドラインとは、そのオブジェクトのデータを次に放送しなければならない時刻のことで、オブジェクトごとに決定されるものとする。 $balancer$ は特定のオブジェクトデータのみが繰り返し放送される現象を回避するためのパラメータで、式(2)を用いて計算する。

$$balancer = \frac{E_i}{U_i} \quad (2)$$

E_i はオブジェクト*i*が前に放送された時刻からの経過時間を、 U_i はオブジェクト*i*に要求される更新頻度から計算した最大放送間隔を示している。

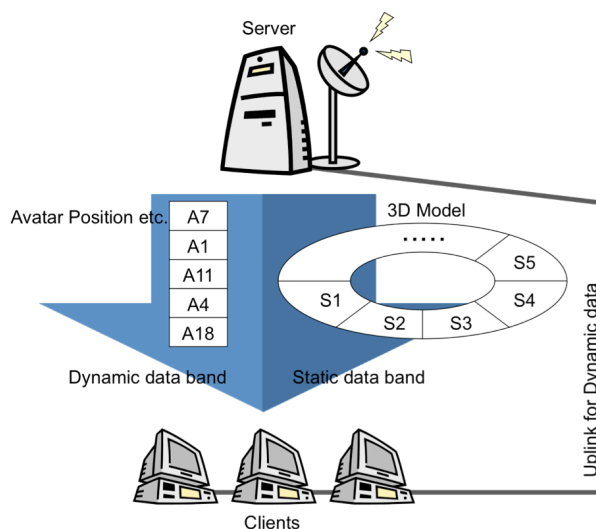


図 1 放送型サイバースペースの概要

仮想社会のシミュレーションや3次元仮想空間を用いたショッピングモールなど、ユーザが自由に活動する空間を想定して、提案方式の有効性を検証した。このような仮想空間では、現実空間と同様に、ユーザの行動は自身の興味に加え、周囲の状況にも影響を受けると考えられる。具体的には、人が集まっている場所を発見するとその対象に関心をもち、同調して自分も近づくといった行動である。ユーザの操作するアバタの行動特性として同調行動を仮定し、仮想空間におけるオブジェクトの人気度に偏りがある環境を設定してシミュレーション実験を行った。シミュレーション実験では、すべてのデータアイテムをランダムにスケジューリングし、繰り返し放送する CycRnd 方式と、デッドラインまでの残り時間の少ないものを優先的に放送する LRF 方式、アバタとの距離とデッドラインまでの残り時間を考慮した IDR 方式、オブジェクトの人気度とデッドラインまでの残り時間を考慮した PIR 方式の4方式と提案方式の性能を比較した。結果の一例を図 2 に示す。図 2 はアバタからの距離ごとのオブジェクトの更新頻度を示しており、アバタから半径50メートル以内にあるオブジェクトについて PIDR 方式が最も高い更新頻度を実現していることがわかる。また PIDR 方式では、特にアバタに近いオブジェクトほど更新頻度が高くなっており、ユーザに提供される仮想空間のビューにおいて大きく表示される重要なオブジェクトが頻繁に更新されるという結果が得られた。

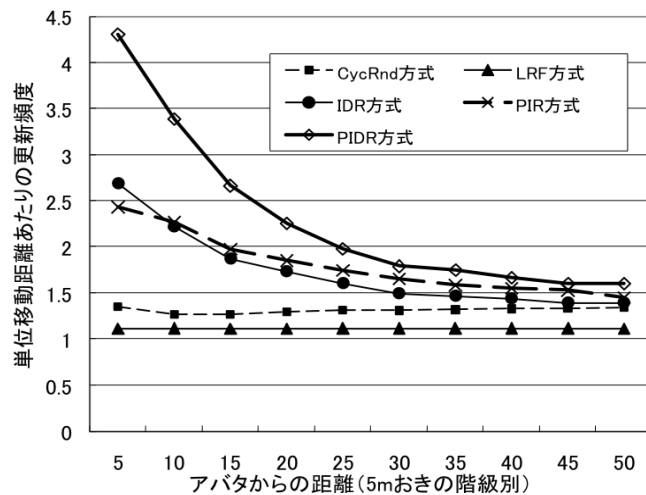


図 2 単位移動距離あたりの更新頻度

2.3 具体的成果

本研究で提案した方式では、同報性の高いプッシュ型の放送配信を利用しているにもかかわらず、人気度が高く、アバタの近くに存在するオブジェクトの更新頻度が高くなっており、各ユーザの視野に表示されるオブジェクトの移動や変化を滑らかに表示できることが分かった。この研究成果は情報処理学会論文誌にて発表した。

また、本研究に関連してマルチビューポートを用いた複数仮想空間の操作インターフェースに関する研究[解説 1, 発表 6]では、日本バーチャルリアリティ学会サイバースペースと仮想都市研究会より表彰され、招待講演を行った[受賞 1, 招待 1]。

3 放送型データ配信を用いた自律移動型センサネットワーク

3.1 背景

環境モニタリング、動植物の生体調査、建物内のセキュリティ管理などを目的として、センサノード(以下、ノードと呼ぶ)だけでネットワークを構築するセンサネットワークが注目されている。センサネットワークでは、各ノードが取得したセンシングデータをマルチホップ通信により基地局に転送し、収集している。従来は、ノードの位置が固定された環境を想定したものが一般的であったが、広大な領域をセンシングするためには多数のセンサが必要となることや、人が入れない汚染地域などではセンサの設置自体が困難であるという問題があった。

近年では、ロボティクス技術を用いてノード自身が移動してセンシングを行ったり、ネットワークを形成したりする移動型センサネットワークが注目を集めている。移動型ノードを用いれば、人が直接センサを設置することが困難な場所であってもセンシングすることができ、さらに広大な領域であっても

少数のノードでのセンシングが可能となる。しかし、ノード数に対して観測領域が非常に大きくなると、各ノードの無線通信範囲に他のノードが存在しない状況が発生する。この場合、基地局にセンシングデータを送信するために、各ノードは基地局と通信可能な位置まで移動する必要があり、多大な電力を消費することとなる。本研究では、図3に示すような疎な移動型センサネットワークにおいて、センシングデータを基地局へ収集するためのネットワーク(以下、収集ネットワークと呼ぶ)を構築する際の移動コストを低減する、放送型データ配信を用いたノード制御方式について検討を行った。

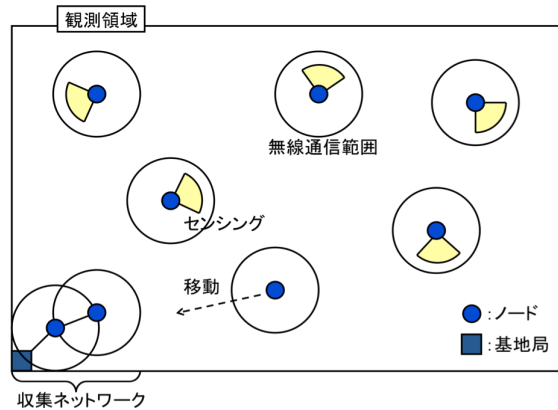


図3 疎な移動型センサネットワーク

3.2 内容

本研究では、プッシュ型放送を用いて、基地局とマルチホップ通信を行っているノードの位置を放送することで、各ノードは基地局へデータを転送できる最も近い地点に移動し、一時的なネットワークを形成するSR(Shortest Route)方式を提案した。また、収集ネットワークを形成するために移動している際に、通信可能となったノードと協調動作することでコストを削減するように拡張したSR-N(SR with Negotiation)方式を提案した。SR-N方式では、移動中に通信が可能となったノードと目的地情報を交換し、相手ノードの目的地と通信可能な位置に移動する方が、現在の目的地まで移動するよりも移動距離が短くなる場合に目的地を変更する。

(1) SR方式

SR方式では、収集ネットワークを効率よく形成するために、収集ネットワークに参加しているノードの位置を放送し、ノードはその情報に基づいて移動する。図4(a)のように、センシングを終えたノード a , b は、放送される収集ネットワークに参加しているノードのうち、最も近いノード c と通信可能な位置 D_a , D_b を目的地として移動する。図中では、点線矢印は移動経路を表している。また、 D_x はノード x の目的地を表す。

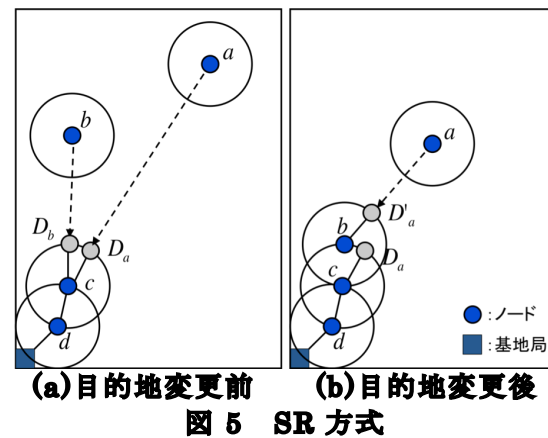


図5 SR方式

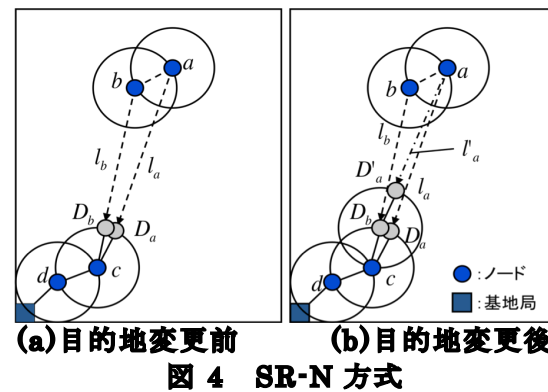


図4 SR-N方式

移動中、新たに収集ネットワークにノードが追加されると、放送されるデータが更新される。図4(b)のように、ノード b が新たに収集ネットワークに参加すると、ノード a はノード b が最も近いノードとなり、目的地を D_a から D'_a に変更する。このように、ノードはより移動距離が短くなるように目的地を変更しながら移動することで、消費電力を低減する。すべてのノードが収集ネットワークに参加した後、データの転送および中継が終了したノードからセンシング地点へと移動する。

(2) SR-N 方式

SR-N 方式は SR 方式を拡張したもので、ノードが目的地に向かって移動している間に他のノードと通信可能となったとき、目的地情報を交換し、連携して収集ネットワークに参加する。これにより収集ネットワークへ参加する際の移動距離を低減し、ノードの移動に要する電力の低減と、そのノードの後方から近づいてくる他のノードの移動コスト削減が期待できる。以下、SR-N 方式におけるノード動作を、単独ノード同士が出会った場合について説明する。

図 5 に単独ノード同士が出会った場合のノードの動作例を示す。 l_x はノード x の目的地までの距離を表している。ノード a, b が収集ネットワーク内のノード c に接続するために移動している途中で、互いの通信範囲内まで接近した状況を示している。ノード a, b は、以下に示す手順で目的地を決定する。

1. 通信可能となったノード a, b は、現在の目的地である D_a, D_b の座標を互いに交換する。
2. 各ノードの現在位置と目的地までの移動距離 l_a, l_b を計算し、移動距離の短いノードを長いノードの接続先候補とする。図 5 では、 $l_a > l_b$ のため、ノード b がノード a の接続先候補となる。
3. ノード a は、ノード b に接続できる位置 D'_a まで移動した場合の移動距離 l'_a を求め、 $l'_a < l_a$ ならば接続先をノード c からノード b に変更する。
4. 手順 3 でノード a が接続先をノード b に変更した場合には、ノード a はノード b に自身の接続先をノード b としたことと、その接続位置 D'_a を伝える。
5. 以後、ノード b は目的地を変更するとノード a に新たな目的地を伝える。ノード a は受け取った情報を基に自身の目的地を修正し、放送による目的地の変更は行わない。

提案手法の有効性をシミュレーション実験を行って確認した。シミュレーションで用いたパラメータを表 1 に示す。観測領域は2次元平面とし、基地局を領域の一角に配置した。各ノードのセンシング地点はランダムに割り当て、変更しないこととした。センシング地点をノードの初期位置とし、ノードはシミュレーション開始と同時にセンシングを開始する。各ノードは1回のセンシングに 1000 秒を要し、この間に 5Mbit のデータを収集した後、データ転送を開始する。ノードの移動コストを 1J/m とし、データの送信、および受信コストを、それぞれ以下に示す式 (3) および式 (4) によりモデル化した。

$$(k \cdot 50) + (0.1 \cdot k \cdot d^2) \text{ [nJ]} \tag{3}$$

$$k \cdot 50 \text{ [nJ]} \tag{4}$$

k [bit] は送受信データ量、 d [m] は送信距離を表す。1回のセンシングデータを 50m 離れたノードに送信する場合、送信コストは 1.5J となる。移動およびセンシングデータの送受信以外の動作に要するコストは非常に小さいものとして無視した。また、ノードはデータの送信と受信を同時に行えるものとし、電波衝突による遅延や、通信エラーはないものとする。表 2 に SR 方式および SR-N 方式と放送を用いない NB 方式における各ノードのコストとスループットを示す。

3.3 具体的成果

本研究で提案した方式を用いることで、疎なセンサネットワークにおいて移動コ

表 2 パラメータ

パラメータ	値
領域の大きさ	2,000[m] × 2,000[m]
ノード数	400
ノードの移動速度	1[m/s]
無線通信範囲	50[m]
通信速度	2[Mbps]
移動コスト	1[J/m]
シミュレーション時間	30,000[s]

表 1 評価結果

	NB	SR	SR-N
移動コスト[J]	11,523	9,773	8,445
送信コスト[J]	318	649	715
受信コスト[J]	61	132	134
スループット[kbps]	621	873	935

ストを低減しつつ、スループットを向上させることが可能であることを確認した。これらの研究成果は電子情報通信学会および情報処理学会のワークショップにて発表した。

なお本研究は、平成19年度科学研究費補助金(若手研究(B))の研究課題「放送型データ配信を用いた自律移動型センサネットワークに関する研究」(課題番号 19700092)として採択されている。

4 成果要覧

招待講演／招待論文

[招待 1] 廣瀬 康一, 小川 剛史, 清川 清, 竹村治雄: 複数の仮想空間を操作するマルチビューポートインタフェースのフレームワーク, 日本バーチャルリアリティ学会研究報告サイバースペースと仮想都市研究会, Vol.13, No.1, pp.1-7, 東京, 2008年2月.

受賞関連

[受賞 1] 廣瀬 康一, 小川 剛史, 清川 清, 竹村治雄: サイバースペース研究賞, 日本バーチャルリアリティ学会サイバースペースと仮想都市研究会, 2008年2月.

著書／解説

[解説 1] 廣瀬 康一, 小川 剛史, 清川 清, 竹村 治雄: 複数の仮想空間を操作するマルチビューポートインタフェースのフレームワーク ~複数の3次元仮想空間をシームレスに接続~, 画像ラボ(日本工業出版), Vol.18, No.4, pp.8-12, 2007年4月.

査読付論文

[査読付 1] 中村 友宣, 小川 剛史, 清川 清, 竹村 治雄: 利用者コンテキスト認識における電車内外判定に関する検討, 情報処理学会マルチメディア, 分散, 協調とモバイルシンポジウム(DICOMO2007)論文集, pp.1496-1501, 三重県鳥羽市, 2007年7月.

[査読付 2] 新城 達也, 北島 信哉, 小川 剛史, 原 隆浩, 西尾 章治郎: 移動型センサネットワークにおけるプッシュ型放送とノード間通信を用いた移動制御手法, 情報処理学会マルチメディア通信と分散処理ワークショップ, pp.111-116, 石川県加賀市, 2007年10月.

[査読付 3] Miranda Miranda Miguel, Takefumi Ogawa, Kiyoshi Kiyokawa, and Haruo Takemura: A PDA-Based See-Through Interface within an Immersive Environment, Proc. of the 17th International Conference on Artificial Reality and Telexistence (ICAT 2007), pp.113-118, Esbjerg, Denmark, Nov. 2007.

[査読付 4] 足立 智章, 小川 剛史, 清川 清, 竹村 治雄: ライブビデオと3次元実環境モデルを用いた遠隔協調作業支援システム, 日本バーチャルリアリティ学会論文誌, Vol.12, No.4, pp.549-558, 2007年12月.

[査読付 5] 小川 剛史, 永石 博憲, 原 隆浩, 西尾 章治郎: 放送型サイバースペースのためのオブジェクトの人気度と距離を考慮したスケジューリング方式, 情報処理学会論文誌, Vol.49, No.2, pp.739-748, 2008年2月.

[査読付 6] 榎田 秀夫, 小川 剛史, 町田 貴史, 中澤 篤志, 清川 清, 竹村 治雄: Diskless Linuxを用いた情報教育システムの開発とその評価, 情報処理学会論文誌, Vol.49, No.3, pp.1239-1248, 2008年3月.

その他の発表論文

- [発表 1] 四宮 龍, 小川 剛史, 清川 清, 竹村 治雄: 三次元物体の比較検討が可能な個人・共有空間連携型協調作業支援システム, ヒューマンインタフェース学会研究報告集, Vol.9, No.3, pp.31-36, 東京都文京区, 2007年6月.
- [発表 2] 四宮 龍, 小川 剛史, 清川 清, 竹村 治雄: テーブルトップ型協調作業環境における三次元モデルの比較検討用インタフェース, 日本バーチャルリアリティ学会研究報告, Vol.12, No.3, pp.97-102, 北海道旭川市, 2007年10月.
- [発表 3] 林 豊, 間下 以大, 小川 剛史, 清川 清, 竹村 治雄: 全方位視覚センサを用いた指先動作検出による円筒型ユーザインタフェースの提案, 電子情報通信学会技術研究報告, Vol.107, No.554, MVE2007-88, pp.43-48, 福岡県北九州市, 2008年3月.
- [発表 4] 中村 友宣, 小川 剛史, 清川 清, 竹村 治雄: 二酸化炭素センサによる鉄道車内混雑度推定を用いたウェアラブル学習システムのための利用者コンテキスト認識, 電子情報通信学会技術研究報告, Vol.107, No.554, MVE2007-89, pp.49-54, 福岡県北九州市, 2008年3月.
- [発表 5] 新城 達也, 北島 信哉, 小川 剛史, 原 隆浩, 西尾 章治郎: 移動型センサネットワークにおけるプッシュ型放送とノード間通信を用いたノードの移動制御手法に関する一考察, 電子情報通信学会データ工学ワークショップ(DEWS2008)論文集, 宮崎県宮崎市, 2008年3月.
- [発表 6] 立石 昂裕, 小川 剛史, 清川 清, 竹村 治雄: 複数の仮想空間を操作するマルチビューポートインタフェースの多人数対応に関する検討, 電子情報通信学会総合大会講演論文集, A-16-19, 福岡県北九州市, 2008年3月.
- [発表 7] 中嶋 拓野, 小川 剛史, 清川 清, 竹村 治雄: 歩行動作の速度に依存した自動非線形スケールリングによる仮想空間ナビゲーション手法, 電子情報通信学会総合大会講演論文集, A-16-20, 福岡県北九州市, 2008年3月.

高帯域遅延積環境における高効率転送

中村 文隆

1 概要

高帯域遅延積環境、特に惑星間通信のように Round Trip Time が $10^2 \sim 10^4$ s 程度のネットワーク環境において、効率の良いデータ転送を行うための方式について考察し、伝播遅延に基づいて優先度を設定するバッファ管理方式の考案と評価を行った。

2 背景

伝送路上に滞留するビット列、いわゆる帯域遅延積 (BDP) が長大となる Long Fat pipe Network(LFN) は地球の内外を問わず存在する。ただし、地上における Grid のような LFN では遅延を $d(\text{sec})$ のオーダーとしたとき $\log(d) \lesssim 0$ であり、逆に惑星間空間 (深宇宙) では $\log(d) \gtrsim 2$ であって、仮に BDP が同程度の地球内外の LFN があったとしても、両者におけるネットワークへの要求条件と実現方法は必ずしも同じではない。

また、現在の惑星間空間で運用されているネットワークは主としてサイエンスミッションのために用いられており、通信は基地局と飛翔体との 1 対 1 通信を基本とし、呼源の性質はポアソンの呼源ではない。

そこで、本研究では遅延 $d(\text{sec})$ が $\log(d) \gtrsim 2$ 、かつ、不特定多数のポアソンの呼源をもつようなネットワーク、例えば、惑星間におけるインターネットのようなネットワークを運用する際に、効率の良いデータ転送を行うための方式について検討した。

2.1 内容

現在、サイエンスミッションのための惑星間ネットワークは一部で運用されているものもあるが、特定少数、かつ観測データ発生時のみのバースト的通信を基本とするため、ハブの役目を果たすステーションの運用はスケジュールベースでの資源予約によることで十分に機能することができる。

そこで、本年度はまず、伝送遅延が桁で異なるような 2 つのポアソンの呼源からのトラフィックがハブステーションで合流する状況において、長遅延側のファイル転送時間の増大を抑制するようなバッファ制御方式について検討を行った。

2.2 具体的成果

地上系と惑星間系での遅延時間の差に関わらず、LFN では従来の TCP の転送制御は十分な性能を発揮できないことは既知であるので、転送制御方式としては現在宇宙ミッションで標準的である CCSDS をベースとし、不特定多数のポアソンの呼源が合流する状況での長遅延側の転送性能改善を主目的としたバッファ制御方式を考案した。

具体的には、長遅延側に高優先度を与え、バッファあふれ時に低遅延側で廃棄によるペナルティの小さなバッファ内容を選択して廃棄することにより、低遅延側の性能低下を抑えつつ長遅延側の性能改善を行う方式を提案し、シミュレータによる評価で一定の改善を示すことができた。

3 成果要覧

その他の発表論文リスト

- [発表 1] 久保淳, 中村文隆, 若原恭: 長遅延ネットワークにおいて効率の良いバッファ管理方式, 電子情報通信学会技術研究報告, Vol. 107, No. 524, pp. 155-160, 2008 年 3 月.
- [発表 2] 久保淳, 中村文隆, 若原恭: 惑星間インターネットにおけるバッファ管理方式, 電子情報通信学会総合大会, B-7-105, 北九州, 2008 年 3 月.

教職員のための情報共有サイトの開発と 高品質なインターネット中継システムの研究

中村 誠

1 概要

従来、メールの転送を主な手法としてきた学内の情報伝達業務を改善するため、ウェブサイトによる、より直接的に情報を伝達できる情報共有サイトを開発している。また、公開講演会のようなイベントを、プレゼンテーション資料を十分な品質とリアルタイム性を損なわずにインターネット中継するシステムを開発している。また、全学的な迷惑メール対策サービス実現のために、迷惑メール対策システムの構築を行っている。

2 教職員のための情報共有サイト

2.1 背景

学内の情報伝達は従来、メールや紙を使用する事が基本的な業務スタイルとして定着しており、情報の発信者から実際の対象者まで、本部や部局、さらに専攻の担当部署を順にメールを転送するためコストと時間がかかっていた。また、電子的に作成され印刷された紙を再度電子化するために読み取られたデータが作成されたり、メール転送の際に通知本文のコピーや関係文書ファイルの添付など情報の複製が多重、大量に生成されたりするなど、メールシステムやストレージなど情報インフラの浪費も問題である。東京大学では数年前から、情報を直接対象者に伝達できることを目標に、ポータルサイトを構築し業務の改善を図ってきた。またポータルサイトは規則など文書ファイルを共有するなど情報集約のツールとしても開発された。しかし、個人認証とそれによって得られた属性に基づいた高度にカスタム化されたインターフェイスを基本機能としていたため、アクセスするためには専用アカウントによる認証が必要であった。情報の提供方法が複雑なため必要な情報に直接、簡単にアクセスする事が困難である、組み込みの検索システムのレスポンスが悪く、検索結果から必要な情報が得にくいなどの理由によって、利用の促進が十分に進んだとは言えなかった。既存ポータルサイトの問題点を改善するため、情報の機密性を十分に検討し、必要以上に認証によるアクセス保護を行わない、必要な情報へ簡単にアクセスできる事を目標とした情報共有システムを開発する事となった。

2.2 内容

既存ポータルサイトの大きな問題として、これまでの業務スタイルからの移行に際して、メールや部局等で利用しているサイトから、ポータルサイトに掲載された情報を案内したくても、通知などの情報固有の URL が存在しないために、一覧表示や検索結果から必要な情報を探す必要があった点が挙げられる。また、一覧表示も本文が折りたたまれた状態で表示され、タイトルをクリックする事で本文が動的に表示される仕様のため、ブラウザのページ内検索が有効に機能しないなど情報を探す事を困難にしていると考えられる。そこで、必要な情報に簡単にアクセスでき、リンクするなど再利用可能な形態を実現するために、通知の大半を占める教職員一般を対象として掲載された通知を、既存ポータルサイトから通知本体と添付ファイルを取得し公開する機能を開発した。さらに、既存ポータル

サイトに格納された事務組織が保有する規則や様式等の文書ファイルを公開する機能を開発した。いずれの機能も各情報が恒久的で直接参照可能な URL を持ち、一次情報源としてメールや他システムからのリンクを通じた情報の再利用が容易に実現できる。また、各通知単独のページ生成など汎用的な検索エンジンでも情報収集ができるなど、既存の情報共有システムが持つ問題を解決している。

2.3 具体的成果

東大ポータルというサイトを構築し、教職員各人が周知文などの通知や事務長会議などの会議資料といった必要な情報を直接アクセスでき、業務に必要な各種システムの利用を開始できる場所を提供した。便利帳と呼ばれる職員各人自らが内容を編集可能なシステムも導入し、各種手続きなどサービスを受ける教職員が必要な情報や担当者を発見したり、職員が業務マニュアルとそれに付随する Q&A 集などを作成したりすることが出来るようになった。今後は、既存ポータルサイトの必要機能を実装し、ポータル機能を担うシステムを東大ポータルに統一するとともに、学内の業務上必要な情報が網羅されたポータルサイトに発展させることを目指している。

3 インターネット中継システム

3.1 背景

公開講演会など広報活動の一環として、会場となる大学から遠い地域のユーザがインターネット上でイベントに参加できるよう、リアルタイムにイベントの様子を中継する事が求められている。その際、プレゼンテーション資料を、可読性を保ったままリアルタイム性を損なわずに提供する事が、ユーザの理解度を高め、満足してもらうために重要となる。

3.2 内容

インターネット上でイベントに参加するユーザに、会場に足を運んだ場合と同等の体験を提供する事を目標に、プラットフォームとして Adobe Flash を採用し、会場の様子を映した動画のライブ配信と共に、講演者の進行と同期したプレゼンテーション資料の表示を行うシステムを開発した。プレゼンテーション資料を撮影して配信すると、解像度が不十分なためユーザが公演内容を理解するのが困難である、資料の内容が更新される頻度が稀であるにもかかわらず常にネットワーク帯域を必要とするといった問題がある。本システムでは、プレゼンテーション資料を Flash の形式で配信するため、資料の細部まで確認することができる。また、資料に埋め込まれた動画やアニメーションなども再現することができる。さらに、イベントの開催時間前後や、休憩時間などライブ配信する動画がない時間帯に、あらかじめ用意したコンテンツを放送する仕組みも有している。

3.3 具体的成果

理学系研究科の公開講演会、東大-JAXA 学際理工学 20 周年記念公開シンポジウム、国立情報学研究所 SINET3 利用説明会において、本システムを用いたインターネット中継を実施した。

4 迷惑メール対策

4.1 背景

迷惑メールがメール配送システムやユーザ個人に及ぼす影響は大きく、迷惑メールによって、正常なメールの配信が不可能となったり、遅延したりする、また、メールサーバのメール保存領域が不要な迷惑メールで浪費される等の問題がある。また、ユーザが迷惑メールを処理するコストが看過できないまでに増加しており、サーバ側で迷惑メールを可能な限り正常なメールと分離出来る事が望まれている。ただし、迷惑メールの判断基準は主観的要素を含むため、大量のユーザに対する解を提

供するのは用意ではない。また、東京大学では、メールアドレスを部局や研究室といった様々な単位で発行しており、メールサーバも同様に大量に存在する。こういった複雑な状況に対応した全学的な迷惑メール対策システムを提供する事が必要とされている。

4.2 内容

学内の現在のメール利用状況に対応するため、以下の要件を満たす迷惑メール対策システムの構築準備を行っている。

1. 研究上使用する専門的な表現や単語が含まれている場合にも対応した判定基準を有する。
2. 大量の複数ドメインに対応している。
3. ドメイン単位およびメールアドレス単位で選択的に迷惑メール対策を施すことができる。
4. 迷惑メールと判定されたメールへの対応として、受信拒否、隔離、削除、ヘッダーの追加や件名の修正などから選択できる。
5. 既存の統一的な認証基盤を必要としない。

4.3 具体的成果

複数の迷惑メール対策ソリューションを機能比較及びテスト運用を行い、総合的な評価を行っている。

DNS の安定運用に向けた分散計測システムの 設計と構築に関する研究

関谷 勇司

1 概要

インターネットは社会生活の情報インフラとして確立し、日常生活ならびに研究活動に不可欠なものとなっている。インターネットの重要性が増すにつれて、その信頼性や耐障害性といったものがますます必要とされている。

Domain Name System(DNS) はインターネットの基盤技術であり、DNS が正常に動作しない環境においては、Web やメール、インスタントメッセージといった、日々日常的に使われているサービスを利用することができなくなる。つまり、DNS を安定運用することは、インターネットにおけるサービスを安定して提供するための必要条件である。

ところが DNS は世界最大規模の分散データベースであり、インターネット上に存在する数万台の DNS サーバが連携することにより、一つのデータベースを形成している。すなわち、DNS の運用状態を把握するためには、単一の DNS サーバに対して、また単一の地点からだけ計測を行うだけでは全体の状況を把握することはできず、系全体を計測するためには多数地点から多数の DNS サーバに対して計測を行う必要がある。

そこで、本研究では分散計測を低コストで継続的に行うことができるための手法ならびにシステム的设计・構築を行った。その結果、DNS の運用状態を単一地点から単一サーバのみならず、多地点から DNS データベースの構造に基づいて計測することができた。また、本研究で構築したシステムは、低コストで展開、運用ができる仕組みとなっており、データ収集における通信の安全性も確保したシステムとなっている。本報告書では、その成果について述べる。

本年度は、具体的に以下の2項目について活動を行った。

- DNS 分散計測システムの設計と構築
- 分散計測システムの展開

2 DNS 分散計測システムの設計と構築

本研究では、DNS がインフラを支える基盤技術として信頼性を有するものであるかを計測することを目的とする。そのための分散計測システムを設計・構築した。その詳細について報告する。

2.1 背景

DNS はインターネットのサービスを支える基盤技術である。DNS に障害が発生すると、インターネットにて動作するほぼ全てのアプリケーションに影響を及ぼす。すなわち、DNS が健全に運用されることはインターネットが健全に運用されることであり、インターネットの信頼性を向上させるためには、DNS の信頼性を向上させる必要がある。

一方、DNS は世界最大規模の分散データベースであり、インターネット上に存在する数万台の DNS サーバが連携することにより、一つのデータベースを形成している。DNS の信頼性を向上させるた

めには、DNS の運用状態を把握する必要がある。しかし、単一の DNS サーバに対して、また単一の地点からだけ計測を行う手法では DNS 全体の運用状況を把握することはできない。DNS の系全体を計測するためには、多数地点から多数の DNS サーバに対して計測を行う必要がある。

そこで、本研究の目的は、低コストかつ容易に運用することのできる DNS 分散計測環境を構築することである。そのためには全世界に容易に展開できる必要があり、かつ低コストで管理・管理できることが必要である。

これらの点をふまえ、システムの構成機器や管理機構の設計を行った。具体的な設計方針を以下に示す。

1. 計測機器が耐故障性に優れていること
2. 計測機器が容易に設置可能であること
3. 計測機器との通信を安全に行うことができること
4. 運用管理が低コストでできること
5. DNS 計測のみならず多種の計測に適用できること

このような分散計測に関する既存研究では、計測機器として PC を利用することが多い。PC は計測ソフトウェアを作成、展開するには便利であり、データを長期的に蓄積できるという点でも優れている。しかし、価格や設置スペース、ならびに海外へ展開する場合の送付の容易性という点では、専用の小型機器を利用する方が望ましい。

これらの条件を満たすために、本研究では、PC ではなくアプライアンスとして市販されている機器を改造して計測機器として利用することとした。本システムの計測機器として利用したのは SEIL¹ と呼ばれる、Internet Initiative Japan Inc. (IJ) が販売している SOHO ルータである。この機器は企業用に販売されている小型ルータであり、耐故障性に優れ、遠隔管理機能有したファームウェアを搭載している。そこで、IJ 技術研究所との共同研究を行い、この遠隔管理機能を改造し、計測ソフトウェアを搭載した特別ファームウェアを作成することによって、計測機器として利用した。

安全性の観点からは、計測機器との通信は暗号化され、証明書等を用いて認証されるべきである。

また、計測機器の運用管理は最小限のコストで行うことが望まれる。PC を計測機器として利用した場合、それぞれの機器に遠隔ログインできるようにしておけば細かい制御ができる。しかし、すべての計測機器に逐次ログインして管理していたのでは、台数が増加するにつれて管理コストも大きくなる。そこで、本研究にて設計したシステムでは、遠隔から一律に制御できる機構を用いて、機器の管理や計測の指示、データ収集までを一元的に制御できる機構を構築した。

2.2 内容

本システムは計測機器、管理サーバ、計測データ収集サーバから構成される。次に述べる遠隔管理機能を利用もしくは改造して、計測のための遠隔管理機構として構築した。

(a) Plug-and-Measure 機能：計測機器を設置するにあたって、個別に事前設定をすることなく、機器をネットワークに接続するだけで自動的に設定を完了し、計測ができる機構を構築した。

(b) 低運用コスト：計測機器の日々の管理、運用が低コストで行えるよう、遠隔管理機構を設計した。XML-RPC を用いて制御メッセージを交換し、管理者やユーザは GUI を用いて各機器の制御や計測を行えるよう構築した。

(c) 安全な通信ならびに制御：X509 証明書を用いてサーバと計測機器間の認証を行い、かつ証明書を用いた SSL 通信を行うことで通信の秘匿性を確保した。もし計測機器が物理的な盗難にあった場合にも、証明書を失効させることでシステムから切り離すことができるため、安全な運用を実現した。

¹<http://www.seil.jp/>

計測機器が設置され、起動する際のシステムの挙動について図 1 に示す。

DHCP が利用できる環境に設置された場合には、計測機器は DHCP にてアドレスを取得し、管理サーバに対して通信を行う。管理サーバは、証明書を使い計測機器を認証し、計測機器にあった設定を送信する。計測機器は、送信された設定にもとづき自分自身を設定した後、計測可能状態に遷移する。

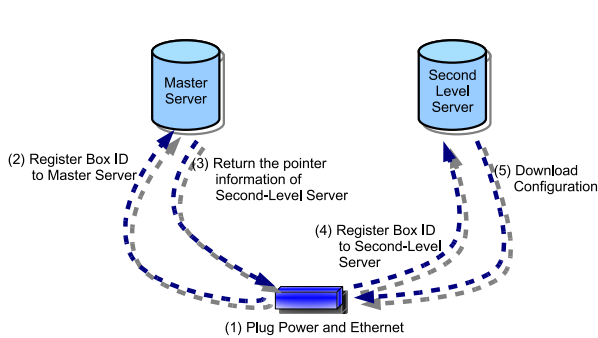


図 1: 計測機器起動時システム図

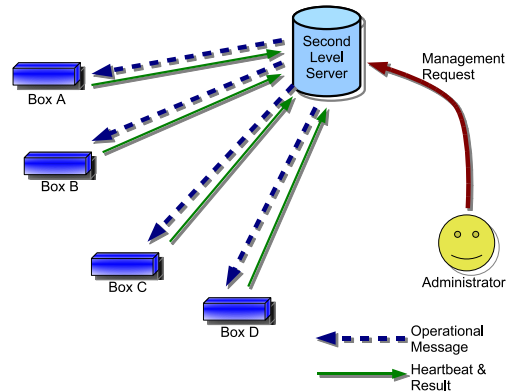


図 2: 計測機器管理システム図

計測機器は、計測可能状態に遷移した後は管理サーバからの指示待ち状態となる。ユーザもしくは管理者は、図 2 に示すように、管理サーバ経由で各機器個別に、もしくは一斉に命令を発行することで、世界中の計測機器を制御することができる。この制御は本システムにて構築した GUI インタフェースを用いて行うことができる。

2.3 具体的成果

本研究にて構築した分散計測システムの制御画面を図 3 に示す。この画面から全ての計測機器に対して、ファームウェアの更新や再起動、状態の監視といった運用を行うことができる。計測機器の状態監視画面例を図 4 に示す。また、計測の指示や計測結果の表示を行うこともできる。Root DNS サーバに対する計測結果表示例を図 5 に示す。

3 分散計測システムの展開

本研究にて構築した分散計測システムにて計測を開始するために、計測機器の展開を行った。その経過について報告する。

3.1 背景

2 節にて述べたシステムを構築し、DNS 分散計測を開始するために、計測機器を世界中に配布し、設置する必要がある。2008 年 3 月現在の計測機器の分布状況について報告する。

3.2 内容

2008 年 3 月現在、20 台の計測機器が世界各地に設置され、計測を行っている。設置状況を表 1 に示す。本研究での計測は、インターネット発展途上国に重点をおいた計測をめざしており、それを考慮した計測機器の配置が行われている。計測機器はまだ配布途中であり、これからさらなる配布を行う予定である。

現在、この計測機器には 2 種類の計測ソフトウェアが組み込まれている。(1) DNS 運用状況計測に関する計測ソフトウェアと、(2) インターネットのトポロジ計測に関する計測ソフトウェアが組み込

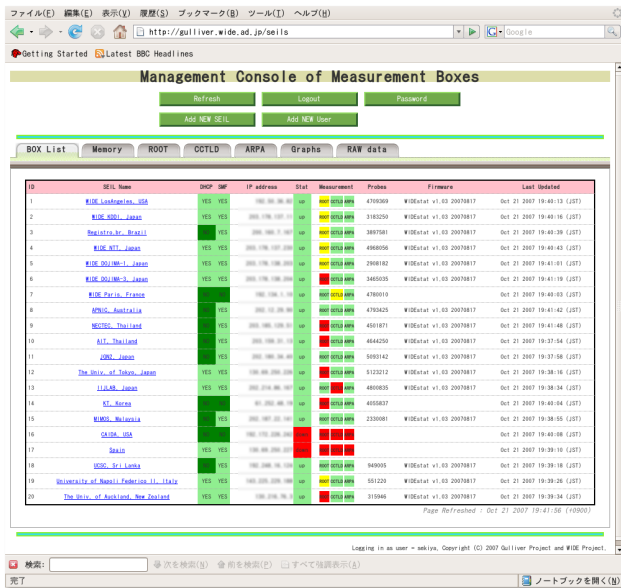


図 3: 計測機器管理画面

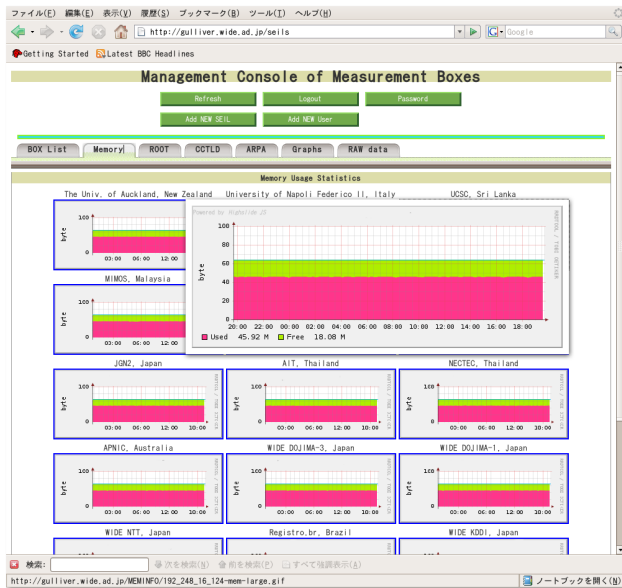


図 4: 計測機器状態監視画面

表 1: 計測機器の配置状況

Box ID	Organization	City	Country
1	WIDE	Los Angeles	U.S.A.
2	WIDE	Tokyo	Japan
3	Registro.br	Sao Paulo	Brazil
4	WIDE	Tokyo	Japan
5	WIDE	Osaka	Japan
6	WIDE	Osaka	Japan
7	WIDE	Paris	France
8	APNIC	Brisbane	Australia
9	NECTEC	Bangkok	Thailand
10	AIT	Pathumthani	Thailand
11	JGN2	Tokyo	Japan
12	The University of Tokyo	Tokyo	Japan
13	IJ Inc.	Tokyo	Japan
14	Korea Telecom	Seoul	Korea
15	MIMOS	Kuala Lumpur	Malaysia
16	CAIDA	San Diego	U.S.A.
17	University of Colombo School of Computing	Colombo	Sri Lanka
18	University of Napoli Federico II	Napoli	Italy
19	The University of Auckland	Auckland	New Zealand
20	DePaul University	Chicago	U.S.A.

まれている。(2) はまだテスト段階であり、実際の計測を開始していないため、本報告書では (1) の計測に関してのみ報告する。(2) の計測は、本研究にて設計・構築した分散計測システムが DNS 以外の計測研究にも用いることができる可能性を示すために取り組んでいる。

3.3 具体的成果

本研究にて構築したシステムを用いて、計測を開始することができた。現在、Root DNS サーバ、ccTLD DNS サーバ、in-addr.arpa DNS サーバのあわせて 800 台程度の DNS サーバに対して、各計測地点から計測を行っている。

本研究では、さらなる計測機器の設置を行うことを目指す。具体的には 100 台以上の計測機器を、特にインターネット発展途上国を中心に設置することを目標とする。さらに、DNS 計測以外にも、インターネットのトポロジ計測といった、DNS の運用状況評価のために必要となるさらなるデータを収集するためにも本システムを利用することを目指す。

本研究の最新状況ならびに計測結果・成果は、<http://gulliver.nc.u-tokyo.ac.jp/> から取得することができる。

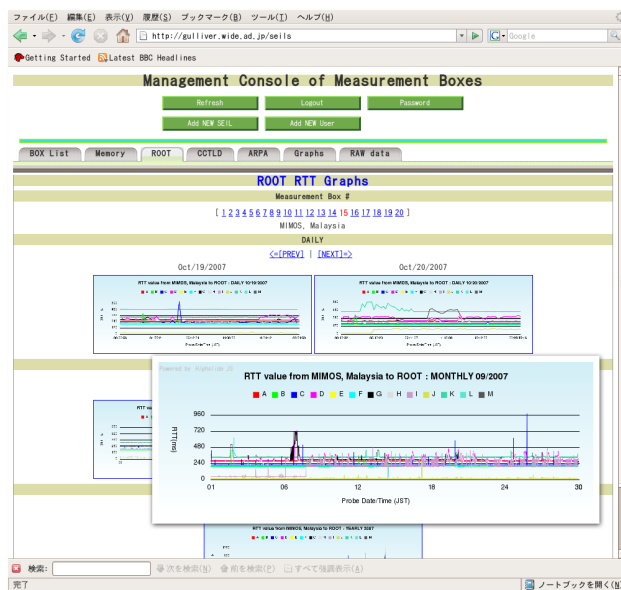


図 5: 計測結果表示画面

4 成果概要

その他の発表論文

- [発表 1] Yuji Sekiya: Gulliver Project Summary on 2007, 9th CAIDA/WIDE Workshop, University of Hawaii, Hawaii, U.S.A., <http://www.caida.org/workshops/wide/0801/>, Jan. 2008.
- [発表 2] 関谷勇司: 「基盤技術の標準化・ディプロイメント技術 (ルーティングエリア・DNS エリア・セキュリティエリア)」, インターネットウィーク 2007 IP Meeting, 秋葉原コンベンションホール, 2007 年 11 月.
- [発表 3] 関谷勇司: 「AS112 Project」, インターネットウィーク 2007 DNS Day, 秋葉原コンベンションホール, 2007 年 11 月.
- [発表 4] Yuji Sekiya: Distributed Active Measurement Framework, 4th JST/CNRS Workshop, Lyon, France, Oct. 2007.
- [発表 5] Yuji Sekiya: Gulliver Project Status, 8th CAIDA/WIDE Wrokshop, Chicago, U.S.A., Jul. 2007.
- [発表 6] Yuji Sekiya and Kenjiro Cho: DNS Remote Measurement Statistics, ICANN 28th RSSAC meeting, Chicago, U.S.A., <http://www.rssac.org/meetings/04-08/RSSAC28.pdf>, Jul. 2007.

公開ソフトウェア

- [公開 1] USAGI Project: Linux における IPv6 プロトコルスタックの開発と公開, linux-2.6-kernel における IPv6 プロトコルスタック, IPsec スタック, Mobile IPv6 スタックの提供, <http://www.kernel.org/>, <http://www.linux-ipv6.org/>, 2007 年 4 月 ~ 2008 年 3 月.
- [公開 2] Gulliver Project: DNS 運用状態分散計測プロジェクト, DNS 分散計測データの公開, <http://gulliver.nc.u-tokyo.ac.jp/>, 2007 年 5 月 ~ 2008 年 3 月.

スーパーコンピューティング研究部門

金田 康正

スーパーコンピューターに基づく大規模数値計算に関する研究

石川 裕

次世代コンピュータシステムの研究開発
—オープンスパコン、グリッド、ディペンダブルシステム—

佐藤 周行

セキュリティを保证する言語処理系の研究
Optimization Verifying Compiler, Document Carrying Authorization, and LoA of Identities

佐藤 文俊

創薬・バイオ新基盤の技術開発へ向けたタンパク質反応全電子シミュレーション・システムの研究開発

黒田 久泰

4倍精度演算ライブラリの高速度および疎行列ライブラリの自動チューニングの効果に関する研究

松葉 浩也

スーパーコンピューターのための高速通信機構

片桐 孝洋

ソフトウェア自動チューニングおよび高性能数値計算ライブラリの研究とHPC教育

吉廣 保

タンパク質量子化学計算支援統合環境の研究開発

スーパーコンピューターに基づく 大規模数値計算に関する研究

金田 康正

1 概要

本研究報告は2007年度中におこなった、以下に示す5種類の研究テーマ（前3種類の研究テーマは主に研究室の大学院学生への研究指導としての研究テーマ）に関する研究概要報告である。まず研究の背景を述べた後、それぞれの研究テーマの解説を以下に行う。

- 4倍精度四則演算の高速化
- $\pi(x)$ 計算の高速化
- リーマンゼータ関数の零点計算の高速化
- pi.super-computing.org 上での π 並びに関連する情報公開
- 並列実行性能の高い数値計算ライブラリー及びプログラム開発環境に関する研究開発

2 背景

以下に示す事を繰り返し毎報告で記述しているが、本センタースーパーコンピューティング研究部門教員の研究における主たる役割として、「利用者サービスに提供している既存のスーパーコンピューターの安定運用並びに、スーパーコンピューターの可用性向上の為の各種問題点の摘出」、また「利用者が将来要求する高性能計算需要をスーパーコンピューターの将来動向を見据えた上で安定的に導入・提供する事柄に関する情報収集や研究活動」がある。

また教育における役割として、「新領域創成科学研究科基盤情報学専攻の情報基盤センターに割り当てられている学生定員に対する兼任教員としての、同専攻修士・博士学生の教育指導」も、前者に劣らず主要な役割となっている。ちなみに2007年度は博士1年1名、修士2年2名、修士1年2名、計5名の教育指導をおこなった。それまでは理学系研究科情報科学専攻の兼担として修士・博士課程の学生指導を引き受けることはあった。ところが毎年一定数の学生教育を引き受けるという状況ではなかった。学生定員が文部科学省定員として本センタースーパーコンピューティング研究部門とネットワーク研究部門に措置されてからは、一定人数の学生が進学してくる様になり、中・長期的な研究テーマに取り組むことが可能となった。なお、情報基盤センターに対する学生定員は、情報基盤センターになって初めて実現したということ、また2008年度からは工学系研究科へ移る予定となっている事を記録としてここに記しておく。

私の研究室の指導方針として、研究テーマは各人の興味に基づき選択させているものであり、2005年度以降の年報はそれまで発行していたセンター全国共同利用部門が発行している年報の記述とは変更になったが、それまでの年報の記載内容もそれぞれの教員の研究・教育活動報告であったことを考え、報告内容の継続性が重要であるということを考えて、これまでとは記載する研究テーマの順

序を変更し、まずは研究指導の一環で研究室学生が得た成果を記載し、その次に学生が関係しない研究テーマの研究成果を記載するという流れで、2007年度の研究・教育活動報告を以下に記す。

3 研究内容

3.1 4倍精度四則演算の高速化

本研究は本研究室の修士2年の永井貴博君が修士課程2年間におこなった研究である。彼がこれまでに得た成果は、すでに印刷公表されている論文[査読付1, 査読付2, 査読付3]また、出版予定の論文[発表1]で発表済みである。この研究の説明は、簡単には彼の修士論文[文献1]の研究背景部分に記してある文章が参考となる。即ち、

「数値計算法や並列計算機の性能向上の進展は、大規模科学技術計算における大きな鍵となっている。特に数値計算分野における浮動小数点数演算においては、計算機上で有限桁に丸められるため、演算ごとに誤差が発生する可能性がある。その有効桁は、広く用いられている倍精度数の場合10進で約16桁である。演算ごとに誤差が発生していると考え、大規模な計算をした場合には最終結果にある程度の誤差が含まれることは容易に想像できる。しかし、その計算誤差がどれくらい含まれるかは問題及び計算手法依存となっており、理論的に求まる値と大きくかけ離れていることもしばしばある。その結果、計算規模が増すに従って多くの計算量を必要とし、さらに計算誤差も増大する。また並列計算機の性能向上では、毎年演算性能は上昇する一方にある。そのために、倍精度数よりも有効桁数が多い4倍精度数を用いた4倍精度演算の必要性が高まってきており、注目されている。4倍精度数の表現には倍精度数を2つ用いて表される128ビットのデータ型があるが、4倍精度演算はソフトウェアによって2つの倍精度数をエミュレートして実現されているため、倍精度演算に比べ多くの計算回数を必要とする。言い換えると、精度を向上させると計算時間が増してしまうという問題点があるが、これは精度と計算時間がトレードオフの関係にあると言える。CPUコア数を増やして並列処理することで演算性能を上げている現状の並列計算機状況を考えると、4倍精度演算はデータアクセスより演算の割合が多く、今ある現状より一般的になっていくと考えられる。そこで本研究では、4倍精度演算アルゴリズムの定量的な解析を行った。高速化手法としては積和演算命令(FMA, Fused Multiply-Add)を用いて4倍精度演算の演算回数を削減し、さらに並列計算機SR11000モデルJ2(以下SR11000/J2)上のCPUであるPOWER5アーキテクチャーに特化した最適化を行った。以下では、本論文に関連する研究内容、またその手法について述べる。」である。

実際浮動小数点数における数値計算法の中には、計算規模が増すに従って計算誤差が増大し、より多くの計算量を必要としてしまうものもある。例えば、線形方程式 $Ax = b$ の解法であり、クリロフ部分空間を用いるCG法などは誤差の影響を大きく受ける。また計算機上における浮動小数点数演算では、実数を一定の有効桁で近似してしまうため、丸め誤差が発生する。そのために、倍精度データ型から仮数部を増やした4倍精度データ型を用いた4倍精度演算の必要性が高まってきており注目されている。

BaileyやHidaらは倍精度浮動小数点数を組み合わせた”double-double”精度のアルゴリズムをFortranやC++などで実装している。またそれらをQD(Quad Double)ライブラリーとして公開している。QDライブラリーには、”double-double”精度の4倍精度数を用いて、4倍精度と倍精度の演算や、倍精度浮動小数点数を4個用いた”quad-double”精度で8倍精度演算を実現している。しかし、倍精度数をさらに上位と下位部分に分割して計算するライブラリーで用いられているアルゴリズムは、積和演算命令を使用しておらず、本研究で用いた4倍精度乗算の約2倍の演算回数を必要としてしまう。積和演算を用いた4倍精度乗算もあるが、演算回数は多くなってしまっているという欠点がある。さらに四則演算だけではなく、三角関数や対数関数、指数関数なども実装されている。またPOWERアーキテクチャーに特化されておらず最適化は施されていない。そしてSSE2などのSIMD命令を用いた高速化も行なわれていない。

一方、演算量の大きい4倍精度演算の専用ハードウェアへの実装も提案されている。4倍精度演算を線形計算や反復解法に応用した研究がいくつか行なわれている。まず4倍精度演算をサポートしているライブラリーとして、GMM++(Generic Matrix Methods)がある。GMM++はC++で記述された密行列と疎行列に対するテンプレートライブラリーで、4倍精度演算についてはQDライブラリーを外部ライブラリーとして呼び出している。Langouらは倍精度演算の直接解法で得られた解を4倍精度演算で反復することで、4倍精度演算の直接解法より高速に解を求められることを示している。さらにKotakemoriらは直接解法ではなく、反復解法を用いた倍精度と4倍精度の混合型反復解法を提案しており、その混合型反復法は反復法ライブラリーLisで実装されている。また一組の倍精度数をSSE2命令を用いて高速化している。

修士論文の2章で、浮動小数点数と本研究で扱う4倍精度数についての概要を説明している。次いで3章で、4倍精度数を用いた四則演算アルゴリズムをいくつか紹介し、これらのアルゴリズムを用いて最適化を行なう。その上で4章で、提案手法である4倍精度数の最適化方法に言及している。そして5章で、POWER 5アーキテクチャであるSR1000/J2上での実験結果を示し、最後に6章で全体のまとめを述べている。実験で使用したマシンは本センターのマシンであり、その上の最適化Cコンパイラー01-03-/Aを用いている。OSはIBM AIXバージョン5.3であり、ラージページの設定は行われてはいない。

本研究で提案する4倍精度数の各種あたりのクロック数性能比較を実際の実験により得ている。この実験でおこなった各種条件とは(1)L1キャッシュサイズのメモリーを使用する演算の場合、(2)L2キャッシュの半分程度までメモリーを使用する演算の場合、(3)L2キャッシュ全体にのる程度のデータサイズの演算の場合、(4)L3キャッシュの半分のデータサイズの演算の場合、(5)L3キャッシュ全体にのる程度のデータサイズの演算の場合、(6)L3キャッシュの1.5倍のデータサイズの演算の場合、の6ケースである。

本実験では、加算、乗算、積和演算、そして積和演算を用いた行列積の4本の4倍精度演算を実装、評価している。また実験は、東京大学情報基盤センターに設置されているHitachi SR11000/J2上の最適化Cコンパイラー01-03-/A、01-03-/Cおよびgcc 4.1.1を用いて行っている。OSはIBM AIX 5.3である。各命令のレイテンシーを考慮する必要があったため、誤差が保証される範囲の演算順序で最適化を行ってアセンブラコードを作成している。そしてC言語で記述したメイン部分と作成したアセンブラコードをリンクして実行している。また、実験はラージページ設定がある場合とない場合の二種類について実験を行っている。なおラージページ設定とはメモリー管理の単位であるページサイズを拡張し、実行に先立って固定的にメモリーを確保するシステムである。この機能は、アドレス変換やページングのオーバーヘッドの削減に役立ち、実効性能の向上に貢献する。またラージページ設定の測定の際にはgcc 4.1.1上でも測定している。測定は実測実効クロックサイクル数、4倍精度演算の性能を示すMQFlops(1秒間に1回の4倍精度演算を1QFlopsとする)を計6パターンのデータサイズで、SR11000/J2最適化コンパイラーの場合と提案した実装とで比較している。時間の測定は、Time Base Registerを用いて行う。これはCPUのクロック数と1対1対応していないため、計測時間は1回のループの最小測定時間とした。また、演算1回あたりの平均をとったことで小数点以下の端数が生じた。行列積における行列は実正方向列であり、その格納には圧縮なしの2次元配列を用いている。行列積はBLASルーチンと同様の3重ループで実装している。本研究の目的は4倍精度演算の定量的な解析であるため、キャッシュサイズに関するブロッキングなどは特に考えない。

これらの各ケースの実験により4倍精度加算では、アンローリング段数を増やすことによってパイプライン処理が効果を出すため、

1. 段数を増やすに従って4倍精度加算演算一回あたりに必要なクロックサイクル数の割合は小さくなっている。
2. 最適化を行うとL1やL2キャッシュなどを用いることで、データ転送の影響を受けない範囲で

は SR11000/J2 最適化コンパイラーより高い性能を示す。

3. 提案手法の6段までは SR11000/J2 最適化コンパイラーを上回ることができておらず、SR11000/J2 最適化コンパイラーも6段から8段アンローリングに値する程度のパイプライン処理を行っていると推測できる。結果的には、4倍精度加算においては SR11000/J2 最適化コンパイラーと同等の性能を示す結果となっているが、レジスタ数の最適化などが限界まで行われていないので、提案手法が勝っている。
4. ラージページ設定に限った gcc4.1.1 との比較では、gcc4.1.1 は提案手法や SR11000/J2 最適化コンパイラーより大きく劣っており、その差は L1 サイズにおいて $257.80 \div 20.41 \simeq 12.6$ となっている。この理由はコンパイラー性能そのものであるが、具体的にはパイプライン処理である。というのは提案手法では8段のアンローリングでパイプライン処理を構築しており、SR11000/J2 最適化コンパイラーにおいても同様のパイプライン処理が施されている。一方、gcc4.1.1 ではソースがインライン化されていないために、毎回の4倍精度演算でライブラリの関数をコールしている。そこで発生するオーバーヘッドも性能低下の一要因として考えられる。これらの理由から gcc4.1.1 における性能低下が大きく見られる。

ことが判明している。

次に本研究で提案する4倍精度数の一乗算あたりのクロック数性能比較について、同様のケースで実験をおこなったところ、乗算においてもアンローリング段数の増加によるクロック数の割合は小さくなっており、4倍精度加算と同様の傾向が見られる。

4倍精度除算では、提案手法が SR11000/J2 最適化コンパイラーに大きく勝っている。浮動小数点乗算命令 fdiv は演算コストが大きいため、SR11000/J2 最適化コンパイラーが4倍精度除算全体をパイプライン処理しなかったことに起因する。他の浮動小数点命令が6クロックで実行するのに対して除算命令 fdiv のみが33クロックと大きなレイテンシーで実現されていた。4倍精度加算や乗算はパイプライン処理されて最適化されていた。加えて、提案手法のアンローリング段数が1のときも SR11000/J2 最適化コンパイラーの性能を上回っており、SR11000/J2 最適化コンパイラーが除算においてはアンローリングしていないことを意味している。

4倍精度数積和演算についても同様に実験を行い、SR11000/J2 最適化コンパイラーと比較すると、提案手法ではデータが L1 キャッシュに乗る場合、 $257.60/167.62 \simeq$ 約 1.5 倍の高い性能が出ていることが分かった。また、4倍精度加算や乗算などそれぞれの演算だけでは性能が劣っている L3 キャッシュのデータサイズでも、二つの演算をまとめた積和演算を最適化することによってデータサイズに依存せず、アンローリング段数が4のときでも SR11000/J2 最適化コンパイラーを上回り、高い性能が出ている。提案手法では L2 キャッシュと同程度のデータサイズで演算性能が最大になった。

4倍精度行列積における性能比較実験について、SR11000/J2 上での行列積と提案手法の行列積では、4倍精度積和演算の性能がほぼ一定の割合を保ったまま推移している。これはデータ転送の割合が少ない4倍精度積和演算性能がそのまま表れた結果と思われる。本研究では議論しなかったキャッシュの考慮についての知見を得るため、ブロックサイズ 200 で提案手法に適用すると、行列サイズが大きくなっても性能の低下は見られなかった。さらなる最適化のためには、データ転送における考慮も必要である。

なお、今回報告していると同様な倍精度を二つ用いたアルゴリズムで Fortran、C++ を用いて実装したライブラリーが公開されている [文献 5]。しかし倍精度数をさらに上位と下位部分に分割して計算する乗算アルゴリズムでは、今回用いた4倍精度乗算の約2倍の演算回数を必要としてしまう。積和演算を用いた乗算もあるが演算回数は多くなっている。Hida らは4倍精度演算のアルゴリズムを応用して倍精度数を4つ用いた8倍精度演算を提案している [文献 6]。本研究では IEEE754 の倍精度実数を2つ用いた4倍精度演算の性能を解析評価し、SR11000/J2 最適化コンパイラーにおいて積和演算

及び行列積を実現、高速化した。SR11000/J2 最適化コンパイラよりも浮動小数点演算数削減、さらに4倍精度における浮動小数点演算の計算量に着目したループアンローリング、浮動小数点命令のレイテンシー隠蔽などによって最大で約1.5倍の演算性能を実現している。

並列計算機の性能向上や数値計算法の進展は、大規模科学技術計算における大きな鍵となっている。特に浮動小数点における演算においては、計算規模が増すに従って多くの計算量を必要とし、計算誤差も増大する。そのために、倍精度数よりも有効桁数が多い4倍精度数を用いた4倍精度演算の必要性が高まってきており、注目されている。

4倍精度数の表現には倍精度数を2つ用いて表される128ビットのデータ型があるが、4倍精度演算はソフトウェアによって2つの倍精度数をエミュレートして実現されているため、倍精度演算に比べ多くの計算回数を必要とする。言い換えると、精度を向上させると計算時間が増してしまうという問題点がある。そこで本研究では、並列計算機 SR11000 モデル J2 上の POWER 5 アーキテクチャーにおける積和演算命令を用いた高速な4倍精度演算の四則演算と積和計算を提案した。

提案手法の1点目として、4倍精度演算アルゴリズムを定量的に解析し四則演算などの演算回数を削減したことがある。POWER 5 アーキテクチャーにおいては、x86 アーキテクチャーに実装されていない積和演算命令を実行できることから、誤差を発生させずに高い有効桁を保ったまま演算できるという特徴を利用した。2点目は POWER 5 アーキテクチャーに特化した浮動小数点演算演算の最適化である。そもそも4倍精度演算は1回の演算に大きな計算コストを必要とするため、データ転送がボトルネックになりやすく、データ転送遅延を考慮せずに演算のみを最適化することができる。主な手法は浮動小数点演算命令のレイテンシー隠蔽と64ビット浮動小数点レジスタ数の考慮という2点で最適化を行った。レジスタ数に収まる範囲内でループアンローリングを行い命令の順序を入れ替えて浮動書数点演算命令のレイテンシーを隠蔽した。レイテンシーを隠蔽することにより CPU はストールせずに実行できるので高速化が可能である。

以上の最適化を行った結果、並列計算機 SR11000 モデル J2 上の gcc4.1.1 に対して提案手法が最大で約5倍の速度向上、また SR11000/J2 最適化コンパイラを上回る性能を得られた。これより、ベクトルデータに対する4倍精度演算の高速化は十分な効果が期待できることが分かった。今後の課題としては、以下のようなことがあげられる。

1. x86 アーキテクチャーにおける高速な4倍精度演算の実現
2. アプリケーションへの適用
3. 4倍精度を超える高精度演算の実現
4. 4倍精度三角関数や指数関数の高速化

積和演算命令が実装されていない Intel、AMD などの x86 アーキテクチャーでは、高速な4倍精度演算が実現されていない。4倍精度乗算において積和演算命令が存在しない環境では計算回数を多く必要とするアルゴリズムが適用されていた。積和演算命令を使用したアルゴリズムで4倍精度演算を実行でき、さらに SSE 命令などを用いてデータ並列化を実現できれば、さらなる高速化が期待できると考えられる。

また、本研究で解析、最適化した4倍精度演算を用いたアプリケーションの開発がある。数値計算分野における様々なライブラリーには4倍精度演算を組み込んでいるものが少なくない。その一例に高速フーリエ変換 (FFT) ライブラリーである FFTW があり、FFTW には4倍精度数での演算も組み込まれている。それらライブラリーの本来の目的は数値解法の改善やその高速化など演算精度以外であるため、4倍精度に代表される高精度演算に最適化されていないと思われる。したがって、様々なアプリケーションでの4倍精度演算高速化が実現できれば有用性がより高まると考えられる。

そして4倍精度にとどまらず、より高精度である8倍精度や16倍精度演算の実現である。関連研究でも述べたように、Baileyらが8倍精度演算ライブラリーを公開している。さらに高精度演算には多倍長計算という分野が存在する。多倍長計算とは限られた桁数のみによる任意の桁数の数の計算を行うものであり、4倍精度演算などとの性能比較については未知数である。しかし、CPUアーキテクチャーにおけるハードウェアレベルで128ビット演算(106ビットの有効桁)がサポートされると、容易に高速な高精度演算が実現できるかもしれない。最後に四則演算にとどまらず、三角関数や指数関数の高速化である。4倍精度数の演算結果を得ようとする、誤差を保証するアルゴリズムを用いて演算する必要がある。

3.2 $\pi(x)$ 計算の高速化

本研究は本研究室の博士2年の吉田仁君が修士課程の2年の間に行い、2007年度博士課程学生として私の研究室に進学した際におこなった研究である。彼の修士論文[文献3]の日本語アブストラクトは、

「本論文では、まず初めに素数個数関数 $\pi(x)$ についての解説と、それに関連するいくつかの関数、理論について述べている。次に、 $\pi(x)$ の具体的な計算方法やアルゴリズムについても述べ、プログラムを実装するために十分な情報を記述している。また、その中の1つのアルゴリズムである Gourdon の手法に基づいたプログラムを作成し、その計算時間に関連するパラメーターの挙動と並列性について実験をおこなった。その結果、 $c=8.096 \times 10^5$ 、 $d=1.864$ とするパラメーターに従うと、 $O(x^{0.805}/\log^{6.64}x)$ で計算できることが分かった。ここで、パラメーター c と d は計算の範囲や時間に関わる別のパラメーターを調整する係数である。さらに、並列計算を行うことで単独のプロセッサでの計算の11.4倍まで計算速度が上昇することが判明した。以上の結果から、これまでの実験と同様に8プロセッサを使うと、40日前後の計算時間で未到達の値である $\pi(10^{23})$ を求めることができると予測される。」である。

具体的手法等についての報告は、2007年度中に投稿論文が採録となり印刷公表されてはいるが[査読付4]、修士論文そのものが英語で記載されていることもあり、ここでは詳しくは述べないこととするが、 $\pi(x)$ は、一般に知られている円周率の値のことでは無く、 x までの素数の数を与える $\pi(x)$ 関数であり、この関数の計算の高速化の観点から行っている、スーパーコンピューターの利用が必須となるこの研究である。これは引き続き博士課程で実施している大きな数の高速素因数分解の研究に向けての事前研究となっている。ちなみに大きな数の高速素因数分解は、 $\pi(x)$ の計算とは比較にならない計算量の計算を必要とする計算であるが、現在までに行われている研究は、計算科学の専門家が行ってはならず、数学研究を主とする研究者が行っており、計算科学の専門家が、計算機を使いこなす観点から、新しい計算アルゴリズムを発見し、実際にそのアルゴリズムに基づく高速計算プログラムを書くことが出来れば、大きな数の素因数分解の分野の研究が進展すると期待している。この修士課程での研究は、その為の事前トレーニングを兼ねた研究であった。

さて2007年度までに実施したこの研究でやり残した点は、今だ正確な値が求められていない $\pi(10^{23})$ を求める計算を実際に行うことであり、これは2008年度前半で完了させたいと考えている。もしこの計算が成功したならば、世界で初めて得た結果となるであろうことに間違いは無いが、問題は実際にこの計算を行う計算環境の準備である。この点に関しては、本学情報基盤センターの前進である大型計算機センターの時代において、教官やその教官の指導を受ける研究室の学生が研究で必要とする計算機リソースは潤沢に利用可能となっていた。しかし、法人後はその様な状況には無く、意欲を持つ学生にとっては実際に利用可能な計算機環境が身近にあるにも係わらず、研究経費がなければ計算が行えないという、実になさけない状況になってしまっている。この問題の改善無くして、スーパーコンピューターを利用する研究の進展、あるいは若い研究者の育成が順調に進むとは思えないということも昨年度に引き続き本年度の報告の中でも記しておくこととする。

3.3 リーマンゼータ関数の零点計算の高速化

本研究は本研究室の博士1年の清野義裕君が修士課程の2年の間に行い、2008年度博士課程学生として私の研究室に進学した際におこなっていた研究である[文献2]。彼の修士論文は英語で記述されており、詳しい内容に関しては原著を参照されたい。

本研究で取り扱う Riemann zeta 関数は素数の性質を通じて多岐の数学的対象と関わりのある重要な関数である。特に Riemann 予想に代表される零点の性質を調べる事は主要な対象である。本研究では Riemann zeta 関数の零点の実際の計算に用いられるアルゴリズムを見直し、そこに存在する周期性に注目する事で、二つの主要な部分において新しい方法を導入し高速化する事が出来る事を明らかにした。具体的な計算において無駄な計算を省く事で、片方では少なくとも3.5倍の高速化が可能である事を確かめており、もう片方ではおよそ3倍の高速化が見積もられている。これらの結果により現在の零点探索記録である 10^{13} 個を越える際に大きく貢献する見込みである。実際の計算は2008年度前半を目処に実施予定であり、来年度の報告では具体的成果を示す事が可能と考えている。その際の報告で使用する言語は英語となる見込みである。

3.4 pi.super-computing.org 上での π 並びに関連する情報公開

80年代に入ってからこれまで4半世紀以上に渡り実施してきている円周率計算に関する学問的意義他については、一般向けの啓蒙・報告書[文献4]並びに、その報告書中で引用している文献、また一般聴取者向けの講演で述べており繰り返すことになるので、本報告では省略することとする。

さて1995年6月末から、<http://www.super-computing.org/>上で円周率の値、並びに関連する(anonymous ftpによる)情報の公開を継続しているが、情報公開開始以降これまでに約13年が経過したことになる。公開当初、一般公開中の物は2億桁までの π と $1/\pi$ 、並びにそれらに関連する統計値であり、特別な使用同意契約をFAXでやりとりして締結することで、アクセス可能な期間は最大1ヶ月であったが、40億桁の π 並びに $1/\pi$ がアクセス出来る様になっていた。しかしそれまで公開していたwww.cc.u-tokyo.ac.jp (sun ws) マシンの不調が度重なった為、2000年9月中旬にサーバーマシンを変更したついでに、40億桁の π 並びに $1/\pi$ が自由にアクセス出来る様にした。

2007年度中に転送された総データ量は、2006年度実績値とほぼ同等の数値であった。しかしアクセスサイト数、アクセス回数、アクセス元のアドレスのそれぞれは増加している。アクセス元のトップアドレスは2006年度に過去最大数を更新し2005年度の10%増の126種類となったが、2007年度は、ほぼ同数の125種類となった。トップアドレス数は減少したが、新しいドメイン名の追加が行われている事から、来年度以降も同様の傾向を示す物と予想している。各種マスメディアでの報道のとおり、2002年度に新しい円周率計算の世界記録を樹立することが出来たので、その時に得られた数値を利用し自由にアクセスできるデータ量及びデータ種の増加させる価値は十分にある。しかしこの種の計算桁数をWebで公開する為のハードウェア整備の為の競争的研究資金の獲得が困難という事情もあり、公開桁数の大幅な増加は2007年度も実施できなかった。しかし近年ハードディスクの容量あたり単価の下落は激しく、2008年度中には大容量ハードディスクが安価に購入できる見込みなので、2008年度中における計算データの公開を実現させたいものである。ただしその際の最大の問題は、計算結果が正しくハードディスクに転送できるかどうかという事と、転送先に正しくデータが転送できたかどうかの確認をどのようにして行うのかという事にある。

さて2007年度中の利用統計は81,664サイト(1995年度は748サイト、1996年度は2,052サイト、1997年度は5,604サイト、1998年度は9,615サイト、1999年度は16,396サイト、2000年度は15,619サイト、2001年度は15,354サイト、2002年度は25,529サイト、2003年度は35,652サイト、2004年度は60,728サイト、2005年度は64,312サイト、2006年度は73,469サイト)(DNSに名前が載っているサイトとしては71,500サイト(1995年度は641サイト、1996年度は1,627サイト、1997年度は4,654サイト、1998年度は8,302サイト、1999年度は14,092サイト、2000年度は13,500サイト、2001

年度は 13,518 サイト、2002 年度は 21,795 サイト、2003 年度は 29,503 サイト、2004 年度は 46,434 サイト、2005 年度は 54,459 サイト、2006 年度は 62,896 サイト)) から総計 147,086 回 (1995 年度は 5,012 回、1996 年度は 9,296 回、1997 年度は 16,127 回、1998 年度は 24,495 回、1999 年度は 40,457 回、2000 年度は 41,191 回、2001 年度は 37,188 回、2002 年度は 58,243 回、2003 年度は 67,851 回、2004 年度は 108,379 回、2005 年度は 116,351 回、2006 年度は 129,497 回) ファイルがアクセスされ、転送された総データ量は 124,604.4 MB (1996 年度は 6,918MB、1997 年度は 13,880MB、1998 年度は 14,455MB、1999 年度は 22,353MB、2000 年度は 55,392.6MB、2001 年度は 59,866.2MB、2002 年度は 111,082.7MB、2003 年度は 86,500.9MB、2004 年度は 120,464.2 MB、2005 年度は 167,199.9 MB、2006 年度は 126,747.6 MB) と一昨年度に引き続き下回っている。なお前年度を下回った理由は不明である。

これらのアクセス元のアドレス (u-tokyo.ac.jp の jp に対応する部分) 種別は総計で 125 種類 (ad ae al am ar arpa as at au aw az ba be bg biz bm bo br bw by ca cc ch cl cn co com cr cx cy cz de dk do edu ee eg es eu fi fj fo fr ge gl gov gr gt gy hk hn hr hu id ie il in info int is it jp ke kg kr kw ky kz la lb lt lu lv ma md mil mk mu mx my mz na net ni nl no nu nz org pe pf ph pk pl pt py ro ru sa se sg si sk su sy th tm to tr tt tv tw ua uk us uy uz ve vi vn ws yu za zm zw。なお 1996 年度は 50 種類、1997 年度は 59 種類、1998 年度は 65 種類、1999 年度は 78 種類、2000 年度は 68 種類、2001 年度は 76 種類、2002 年度は 82 種類、2003 年度は 89 種類、2004 年度は 104 種類、2005 年度は 115 種類、2006 年度も 126 種類) と、初めて前年度を下回った。

2007 年度末までの円周率計算の公表されている世界記録は 2002 年 11 月末に SR8000/MPP を利用して樹立した 10 進で 1 兆 2411 億桁である。なお、この時同時に 16 進で 1 兆強計算しており、この桁数も当然ながら新記録である。計算した全ての桁数の公開は近年の磁気ディスク記憶装置一台当たりの記憶容量の観点からは何ら問題は無いが、転送するデータ量を考えると、ネットワーク経路による最新の計算結果の公開はますます困難となっている。しかし Giga ビット回線が国内外で充実しつつあることと、高信頼性ディスク装置の容量当たりの単価が下がって行っており本章の初めの部分にも述べてあるように、公開用のハードウェアの準備が行えたなら、2008 年度中には公開する予定である。

また 2006 年度末には 2004 年度末に導入した新しいスーパーコンピューター HITACHI SR11000/J1 の中間レベルアップが実施され、現在サービスしているスーパーコンピューターの 3 倍強の主記憶容量・総合演算性能を持つ新しいスーパーコンピューター HITACHI SR11000/J2 のサービスが開始された。新しいマシンはこれまでの 3 倍弱のノード数と総主記憶容量、3 倍強の性能と高信頼性・高速磁気ディスクを有しており、本センターでこれまで新機種の導入当初に行い、各種問題点を的確に検出してきた実績を有する、信頼性・可用性確認用の大規模計算を実施することになる。システムに高負荷をかけるこの計算の実施により、結果として計算桁数記録の更新が実現される可能性があった。しかし残念ながら 10 進結果による新記録樹立には至らなかった。その理由は、プログラム中の問題が発現した事によるものであるが、その問題点の確定および対策確認に必要な計算リソースの確保が困難だった事による。いずれ、機会を見て問題点の確定および対策確認を行いたいと考えているが、それがいつになるかの見通しは本報告書作成時点では全くたっていない。

3.5 並列実行性能の高い数値計算ライブラリー及びプログラム開発環境に関する研究開発

分散並列型計算機の性能を引き出せるプログラムを書くことは、職人芸的要素が多くまたなかなか評価を受けにくいという現状がある。従ってプログラミングを生業としている、あるいは計算科学を専門としない多くの研究者にとって、高性能プログラムの作成を困難にしているし、プログラム作成効率を高めるための努力や、プログラムそのものの実行効率を極限まで高めようとすることに興味を持ち、長期にわたり没頭することに喜びを感じる学生が生まれることは多くは無いという現状がある。しかし研究の最前線の進展速度を高める為には、多くの研究者にとって使いやすく、高い実行性能が

出せ、ノード数に関するスケーラビリティが高く、広範囲の機能を有する数値計算ライブラリーの提供が少なくとも必要である。その様な目的を持った数値計算ライブラリーが、一部ベンダーあるいは独立ベンダーから提供・販売されてはいるが、その導入コストが高かったり、機能的に不十分だったり、実現している機能間の性能にアンバランスがあったり、機種間の互換性に問題があるのが現状となっている。また将来的には、計算機を利用した研究・開発における研究環境のインフラ整備の一環として、高い実行効率を容易に出せたり、プログラムの保守管理が容易となる様なプログラム開発環境の開発や提供が今後必要となるであろう。

この様な理由により本研究に係わる目的として、分散メモリー型計算機および共有メモリー型計算機、また PC クラスターから、スカラー並列計算機あるいは（ハード及びソフト）ベクトル並列計算機までの広範囲のアーキテクチャーのどれに対しても高い並列実行性能を発揮でき、広範囲の機能をカバーし、自動チューニング機能等による使いやすい数値計算ライブラリーの研究開発、並びに簡単なプログラミングで実行効率が高いプログラム開発環境の研究開発、それぞれを目的としている。具体的には、東京大学新領域創成科学研究科基盤情報学専攻の学生諸君達とこの研究テーマで研究を行う予定である。2004 年度までにおこなった研究として、特に連立一次方程式の反復解法における階の直交性と収束特性に関する研究と、大きな素数の基底に対する高速フーリエ変換の高速化に関する研究を実施したが、2005・2006・2007 の各年度はこの種の数値計算に興味を持つ学生が全く存在しなかった為、本研究は一時停止状態になっている。加えて 2008 年度に数値計算に興味を持つ修士課程の学生は進学して来ない事となったので、2009 年度後半以降に研究活動を開始することになるであろう。

なおこの研究テーマに関連して、「次世代スーパーコンピューティング応用言語研究会（Next Generation Computing Application Language Consortium（略称 NSCAL）」を 2005 年 11 月に立ち上げて、研究室外のメンバーと活動を開始し、2007 年度は合計で 6 回の研究会を開催した。この研究会で発表した、仕様を検討した事のある言語で記述した数値計算プログラムの一例を最後に示す。この言語については今後地道な改善が必要であるが、過去のソフトウェア財産を生かす一つの方法でもあるので、必要機能や仕様の検討を 2008 年度以降も継続することになる。

```
section "HIMENOBMTXP"
$inline sendp
let jacobi(nn, gosa) be {
  IMPLICIT REAL*4[a-h, o-z]
  get 'mpif.h'
  include 'param-unix.h'
  do loopll=1, nn {
    gosa:=0.0;
    wgos:=0.0;
    do K=2, kmax-1 {
      do J=2, jmax-1 {
        do I=2, imax-1 {
          S0:=a(I, J, K, 1)*p(I+1, J, K)+a(I, J, K, 2)*p(I, J+1, K)+
            a(I, J, K, 3)*p(I, J, K+1)+b(I, J, K, 1)*(p(I+1, J+1, K)-
              p(I+1, J-1, K)-p(I-1, J+1, K)+p(I-1, J-1, K))+b(I, J, K, 2)*
              (p(I, J+1, K+1)-p(I, J-1, K+1)-p(I, J+1, K-1)+p(I, J-1, K-1))+
              b(I, J, K, 3)*(p(I+1, J, K+1)-p(I-1, J, K+1)-p(I+1, J, K-1)+
                p(I-1, J, K-1))+c(I, J, K, 1)*p(I-1, J, K)+c(I, J, K, 2)*
                p(I, J-1, K)+c(I, J, K, 3)*p(I, J, K-1)+wrk1(I, J, K)
          SS:=(S0*a(I, J, K, 4)-p(I, J, K))*bnd(I, J, K)
```

```

        WGOSA:=WGOSA+SS*SS
        wrk2(I,J,K):=p(I,J,K)+OMEGA*SS
    }
}
};
do K=2,kmax-1 {
    do J=2,jmax-1 {
        do I=2,imax-1 {
            p(I,J,K):=wrk2(I,J,K)
        }
    }
}
sendp(ndx,ndy,ndz);
mpi_allreduce(wgosa,gosa,1,mpi_real4,mpi_sum,mpi_comm_world,ierr)
}
// End of iteration
}

```

4 具体的成果

2007年度は博士1年1名、修士2年2名、修士1年2名の学生が研究室に存在したため、修士2年の修士論文を投稿論文とした論文[査読付1, 査読付2, 査読付3, 査読付4]を印刷・公表することができたので、これを具体的成果とする。もう一人の修士学生は博士に進学しており、2007年度中に英語で国外の国際会議に投稿したが、2007年度中に採録報告が届かなかったため、2007年度中の具体的成果とはしないこととする[発表1]。来年度以降の修士進学生は存在しないが、博士課程学生が2名となるので、世界で初めての計算を実施する等の、対外的なアピール度の高い成果を得ることとなる。また本報告では詳しくは述べなかったが、対外発表として圧電トランス関係の発表[発表2, 発表3]と、ワークショップでの発表[発表4]がある。

査読付論文リスト

- [査読付1] 永井貴博, 吉田仁, 黒田久泰, 金田康正: SR11000 モデル J2 における 4 倍精度積和演算の高速化, 情報処理学会第 5 回先進的基盤システムシンポジウム SACSIS2007, 2007 年 5 月 23 日-25 日, 学術総合センター講堂・会議室, pp. 129-136, 2007.
- [査読付2] T. Nagai, H. Yoshida, H. Kuroda and Y. Kanada: Fast Quadruple Precision Arithmetic for Multiply/Add Operations on SR11000/J2, The 2007 Internal Conference on Scientific Computing 07 (CSC'07), pp. 151-157, 2007.
- [査読付3] 永井貴博, 吉田仁, 黒田久泰, 金田康正: SR11000 モデル J2 における 4 倍精度積和演算の高速化, 情報処理学会論文誌: コンピューティングシステム, Vol. 45, No. SIG13(ACS19), pp. 214-222, 2007.
- [査読付4] H. Yoshida, H. Kuroda and Y. Kanada: Parameter Selection for Fast Computation of $\pi(x)$, The 2007 International Conference on Foundation of Computer Science (FCS'07), pp. 177-182, 2007.

その他の発表論文リスト

- [発表 1] T. Nagai, H. Yoshida, H. Kuraoda and Y. Kanada : Fast Multiple Precision Arithmetic Library on Parallel Computer SR11000/J2, (accepted), ICCS 2008, 23–25 June, Kurakov.
- [発表 2] M. Imori and Y. Kanada: Low Voltage Power Supply Incorporating Ceramic Transformer, Topic Workshop on Electronics for Particle Physics (TWEPP-07), 03–07, Sept. 2007, Hotel Olympik Artemis, Prague.
- [発表 3] M. Imori, Y. Unno and Y. Kanada: Low Voltage Power Supply Incorporating Piezoelectric Transformer: Plan for Development, ATLAS Upgrade Tracker Power Preview held on 04 Feb. 2008 at CERN.
- [発表 4] Y. Kanada: Ultra High Speed Computer 「京速計算機」 Project in Japan : Success or Failure?, The 10th International Workshop on Next Generation Climate Models for Advanced High Performance Computing Facilities, 28 Feb. 2008 – 1 Mar. 2008, Honolulu Convention Center.

引用文献関連

- [文献 1] 永井貴博 : A Study on the Fast Quadruple Precision Arithmetic Using Fused Multiply-Add Instruction on POWER 5 Architecture (Power 5 アーキテクチャにおける積和演算命令に基づく高速 4 倍精度演算に関する研究) , 東京大学新領域創成科学研究科基盤情報学専攻修士課程修士論文, 2008 年 3 月.
- [文献 2] 清野善裕 : Computation of zeros of the Riemann zeta function (Riemann zeta 関数の零点計算) , 東京大学新領域創成科学研究科基盤情報学専攻修士課程修士論文, 2008 年 3 月.
- [文献 3] 吉田仁 : Computation of $\pi(x)$ 東京大学新領域創成科学研究科基盤情報学専攻修士課程修士論文, 2007 年 3 月.
- [文献 4] 金田康正 : 計算機による円周率計算, 数学文化, pp. 72–83, 日本評論社, Vol. 1, No. 1, 2003.
- [文献 5] D.H. Bailey: A fortran-90 double-double library, <http://crd.lbl.gov/dhbailey/mpdist/mpdist.html>.
- [文献 6] Yozo Hida, Xiaoye S. Li, and D. H. Bailey: Algorithms for quad-double precision floating point arithmetic, Proceedings of the 15th Symposium on Computer Arithmetic, pp. 155 – 162, 2001.

次世代コンピュータシステムの研究開発

— オープンスパコン、グリッド、ディペンダブルシステム —

石川 裕

1 概要

平成19年度、次世代コンピュータシステムを実現するために、オープンスパコンアーキテクチャの仕様決定、グリッド環境下での高性能データ転送の提案と評価、ディペンダブルシステムを支援するオペレーティングシステム的设计とプロトタイプシステムを開発した。オープンスパコンアーキテクチャの仕様では、京都大学 学術情報メディアセンター、筑波大学 計算科学研究センターと共に共通基本仕様を決定した。当基盤センターには、平成 20 年 6 月1日から理論総性能値 140 Tflops を有するクラスターが稼働する。

グリッド環境では、これまで、広域ネットワークの通信バンド幅が小さいと仮定されていた。しかし近年、通信バンド幅は急速に向上し従来の仮定とミスマッチしてきた。そこで、高遅延かつ高バンド幅のネットワーク環境を仮定する。そのような広域ネットワーク上に設置されている複数クラスター上で一つの並列アプリケーションを実行することを仮定する。広域ネットワーク上では、TCP/IP プロトコルが利用される。TCP/IP は連続したデータを転送し続ける時には、ネットワークバンド幅を有効利用するが、間欠的なデータ転送では、ネットワークバンド幅が有効利用されない。TCP/IP プロトコルの仕様を修正することなく、ネットワークバンド幅を有効利用できる枠組みを提案し評価した。本研究は産業技術総合研究所グリッドセンターとの共同研究で行われた。

ディペンダブルシステムに関する研究ではネットワークで接続された組込みコンピュータによるデータベースサーバや高性能計算システムを実現するための高信頼高性能並列分散オペレーティングシステムを実現する。処理能力に応じて、ディスクを搭載したコンピュータ群、マルチコア CPU とメモリだけを搭載したコンピュータ群が適時ネットワークでつながれるクラスターシステムを想定し、Linux を基に、高信頼単一システムイメージを提供する並列分散 OS を実現する。このために、まず、オペレーティングシステムの拡張性と安全性を提供する枠組み P-Bus を設計し、プロトタイプシステムを開発した。P-Bus 上で Linux OS を安全に拡張し、オペレーティングシステムレベルでハードウェアの耐故障機能および実時間性を実現する。

2 オープンスパコン

2.1 背景

科学技術の更なる発展のために、スーパーコンピュータによる問題解決やサービス提供が世界レベルで大きな関心になっていることは、言うまでもない。このような状況で、東京大学、京都大学、筑波大学の計算センターはそのサービスにおいて、応用領域の多様性に答え、広範なユーザーニーズ資すべきであるという理念で一致した。また、3 センターのスーパーコンピュータのリプレース時期が平成 20 年と一致したこと、そして、3 大学のセンターに所属する教授陣が長年培ってきたコンピュータサイエンス分野における先端研究のノウハウ (PACS-CS, SCore, MegaProto 等) のシナジー効果により、より良いスーパーコンピュータ調達が可能になると考え、共同で調達の基本仕様を策定した。

通常のスパコン調達の手続きを経て、平成 19 年 12 月、日立製作所製 HA8000-tc/RS425 のノード群を 10Gbyte/sec の高速ネットワークでつながるシステムの導入が決定した。平成 20 年 6 月 1 日から理論総性能値 140TFlops のスパコンが稼働する。本稿では、本システムを T2K オープンスパコン東大版と呼ぶことにする。T2K は、北から筑波大学、東京大学、京都大学の頭文字をとったものである。

2.2 内容

T2K オープンスパコン東大版に使われている計算ノードおよびネットワークスイッチを図1に、システム全体の諸元およびノードの諸元を表 1、表 2 にそれぞれ示す。

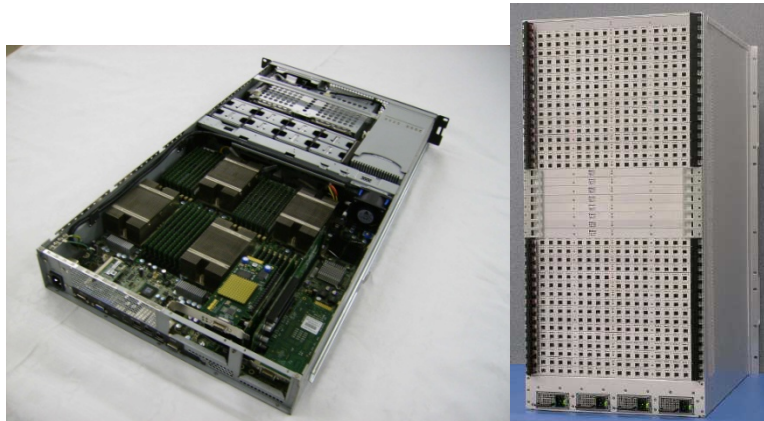


図 1 計算ノード日立製作所製 HA8000-tc/RS425(左)と Myrinet-10G スイッチ

表 1 システム全体の諸元

総理論演算性能	140.1344 TFlops
総主記憶容量	31.25 TByte
ノード数	952
ネットワーク性能 (両方向物理データ転送性能)	タイプ A(512, 128 ノード) : 10 GByte/sec (Myrinet-10G x 4) タイプ B(256, 36, 16) : 5 GByte/sec (Myrinet-10G x 2)
ファイルシステム容量	1 PByte
OS	RedHat Enterprise Linux 5
バッチジョブシステム	現有バッチジョブシステムと同様の機能が提供されます
コンパイラ	日立製作所社製最適化 Fortran, C/C++ (OpenMP 有)
通信ライブラリ	MPI 通信ライブラリ (MPICH-MX)
数値計算ライブラリ	MSL2 MATRIX/MPP MATRIX/MPP/SSS BLAS, LAPACK, ScaLAPACK
アプリケーション	Gaussian03

表 2 ノードの諸元

CPU	AMD 社製 Quad Core Opteron 2.3 GHz
ソケット数(CPU 数)	4 (16)
理論演算性能	147.2 GFlops
主記憶容量	32 GByte (936 ノード)、128GByte(16 ノード)
ディスク	250 GByte

また、図2に示す通り、T2K オープンスパコン東大版は計算ノードとログインノード合わせて 952 台から構成(管理ノードは省略)され、大きく4つのグループに分かれる。タイプ A、タイプ B の違いはネットワーク性能であり、両方向物理データ転送性能は、それぞれ、10GByte/sec、5GByte/sec で、フルバイセクションバンド幅を提供する。このように 2 つのタイプを設けたのは、並列アプリケーションでもネットワーク性能を比較的必要としないアプリケーションも想定してのことである。また、タイプ B の中で 16 ノードは 128GByte の主記憶を有する。

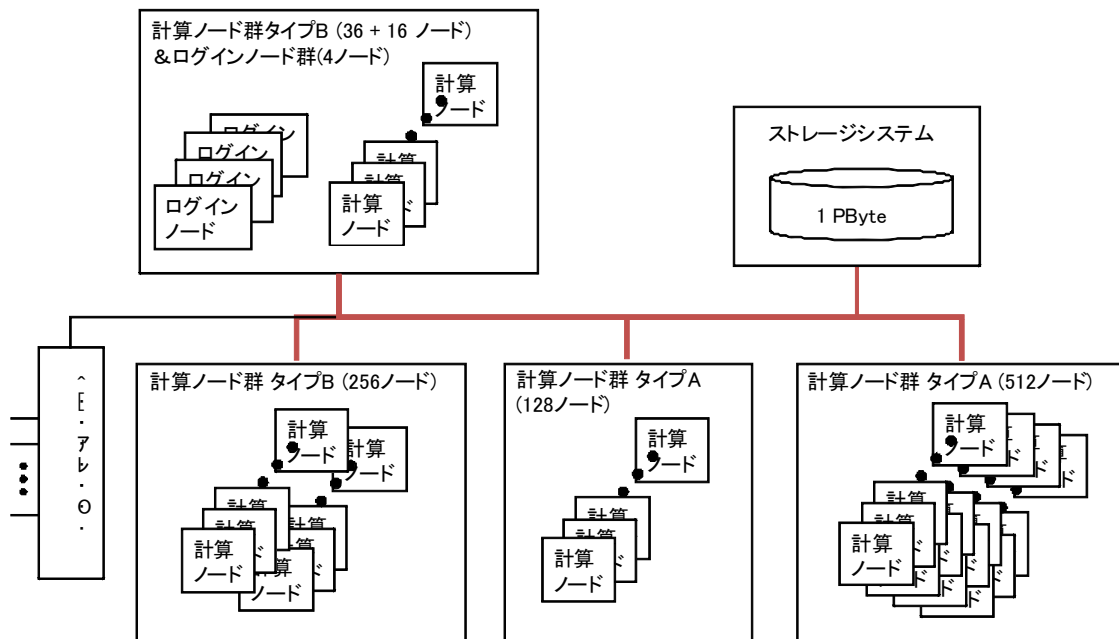


図 2 T2K オープンスパコン東大版の概略

表 3 現有スパコンとの比較

	T2K オープンスパコン東大版	SR11000/J2
総理論演算性能	140.1344 TFlops	18.841 TFlops
総主記憶容量	32000 GByte	16384 GByte
ノード数	952	128
単体ノード演算性能	147.2 GFlops	147.2 GFlops
単体ノード主記憶容量	32 GByte(128 GByte)	128 GByte

表 3 に、現有の SR11000/J2 とシステム全体の比較を示す。総理論演算性能値では、現有スパコンの約7倍、総主記憶は 2 倍となっている。ノード単体での性能を見ると、理論演算性能値は同じで、主記憶容量は 1/4 となっている。表には掲載していないが、CPU とメモリ間のデータ転送幅は、SR11000/J2 が約 5 倍良い。SR11000/J2 は、昨年春に SR11000/J1 から増強されたばかりで、後、3 年間運用される。SR11000/J2 を利用しているユーザが、次期スパコンに移行するかどうかの大きな目安は次のとおりとなる。

- 1) CPU と主記憶間のデータ転送バンド幅が必要なアプリケーションユーザは、SR11000/J2 を引き続き使うのが良い。たとえば、主記憶転送性能がアプリケーション性能にインパクトを与える流体系のアプリケーション実行では、SR11000/J2 の方が 2~3 倍速い。

- 2) 並列化されていて台数効果が期待できるアプリケーションユーザは、T2K オープンスパコン東大版への移行を期待したい。ノード単体での性能が劣っていても、並列化されたアプリケーションであれば、使用するノード数を増やすことにより、性能及び利用可能となるメモリ容量も増やすことができる。T2K オープンスパコン東大版では、32 ノード(512CPU)や 64 ノード(1024CPU)が自由に使える運用を行っていく予定でいる。また、ニーズに応じて最大 512 ノード(8192CPU)を利用できる運用も行うことを考えている。
- 3) PC クラスタ上のアプリケーションを実行したい方は、T2K オープンスパコン東大版を勧めている。T2K オープンスパコン東大版の OS は RedHat Enterprise 5 であり、多くの PC クラスタ上のフリーソフトウェアは再コンパイルするだけで稼働すると期待している。

2.3 具体的成果

「オープンスパコン仕様」は、従来の調達で行われてきた単なる市場調査の結果として定めたものではない。石川研究室で開発してきた MPI 通信ライブラリの一実装である YAMPI[公開 1]を使用してコモディティ部品(一般に市販されている部品)によるスーパーコンピュータの実現可能性をさぐった結果を反映している。T2K オープンスパコン東大版は、6 月 1 日より全国共同利用施設として利用される。また、平成 20 年度秋からは、イノベーションに結びつくような民間利用に対して有償支援する予定でいる[報道 1]。

筑波大学、東京大学、京都大学で進めたスパコン調達は、単なるスパコン調達として終わるのではなく、今後 T2K オープンスパコンアライアンスとして連携を深めていく。運用面では、グリッド技術による協力関係を築く作業を開始している。また、ペタフロップス級スパコン時代を迎えるにあたって、並列プログラミング支援での協力、さらに、より使い勝手の良い並列プログラミング環境実現に向けて共同研究開発を行っていく。

3 グリッド

3.1 背景

グリッド環境で複数のクラスタを接続して利用するための MPI システムとして MPICH-G2、MagPle、PACX-MPI などが知られている。これらグリッド用 MPI システムでは、通信遅延が大きいことと同時に通信バンド幅が小さいことが想定されている。

しかし近年、広域ネットワークの通信バンド幅は広がり、10Gbps から 40Gbps といったバンド幅が提供されるようになってきた。このバンド幅は典型的なクラスタノードの NIC の持つ 1Gbps というバンド幅を越えている。この傾向は、クラスタノードの NIC にかけるコストを勘案すると、しばらくは続くと考えられる。このようなネットワーク環境は、これまでのグリッド用 MPI における仮定とマッチしていない。

広域ネットワーク上の複数の並列システム(クラスタシステム)を使って並列アプリケーションが稼働していると仮定する。並列アプリケーションでは、各ノードがローカルな計算を行い、その結果を他のノードに配布するために全ノードが一斉に通信を行う、ということを繰り返すことが多い。全ノードが一斉に通信するとき、ネットワークが輻輳する可能性がある。一般に、広域ネットワークでは通信に TCP/IP プロトコルが用いられる。TCP では輻輳により致命的に性能が低下することが知られている。このため、バンド幅を有効利用するとともに、輻輳を避ける必要がある。平成 18 年度は、コレクティブ通信のアルゴリズムを高遅延かつ高バンド幅なネットワークに最適化した。平成 19 年度は、並列アプリケーションの通信パターンに応じて、利用するバンド幅を調整する機構の研究開発を行った。本研究は、産業技術総合研究所グリッドセンターとの共同研究で行われた。

3.2 内容

広域ネットワークでつながっている2つのクラスタ上で並列ベンチマークプログラムである NAS Parallel Benchmarks の BT, SP, CG を動かすことを考える。図3に通信パターンを示す。BT, SP は、 $N \times N$ のノード上で、隣接同士の通信を行う。32 x 32 のノードを図3(a)のように Cluster A と Cluster B にマッピングすると、クラスタ間の通信は赤線で示した通りとなる。図3(b)には、CG ベンチマークの通信パターンを示す。

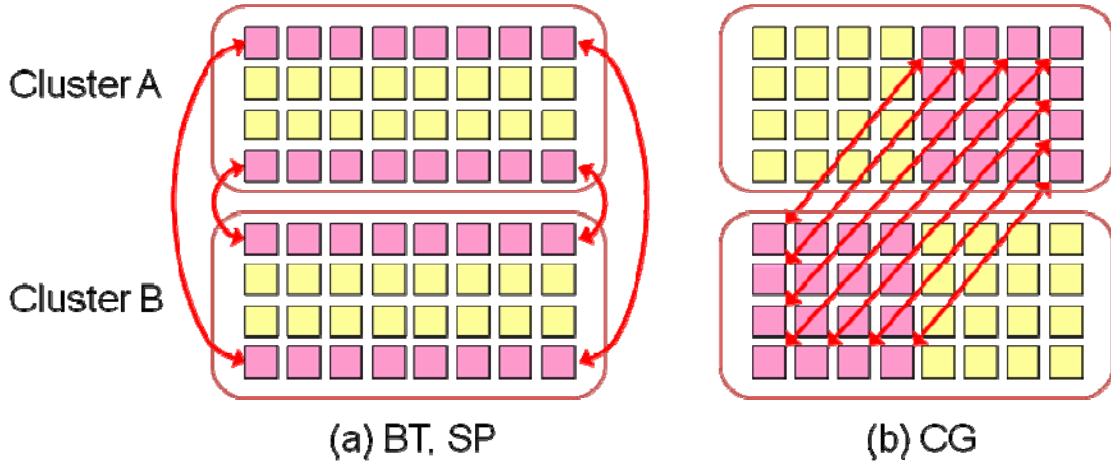


図3 NAS Parallel Benchmarks BT, SP, CG の通信パターン

Cluster A と Cluster B の間は広域ネットワークなので、通信プロトコルには TCP/IP が使われる。何ノードが同時に広域ネットワークを利用するのがあらかじめ分かるので、各ノードが広域ネットワーク経由で通信する場合には、利用するネットワークバンド幅を指定することにより、輻輳が生じないようにする、というのが本研究の主アイデアである。輻輳を起こさずにノードが利用できる最大のネットワークバンド幅を MATB (Maximum Allowable Transmission Bandwidth)と呼んでいる。表4に、NAS Parallel Benchmarks の各ベンチマークにおける、10 Gbps 広域ネットワーク利用時の利用可能ネットワークバンド幅を示す。

表4 プログラムと各ノードが利用できるネットワークバンド幅

Benchmarks	MATB	(B=10 G, N=32)
BT	$B/(2\sqrt{2N})$	625 Mbps
CG	$B/(N/2)$	625 Mbps
LU, MG	B/N	312.5 Mbps
IS, FT (all-to-all)	B/N	312.5 Mbps

あらかじめ計算した MATB を使って、ノードが送信する通信量を調整する。通信量調整には、すでに産業技術総合研究所と開発した PSPacer を使う。YAMPI 通信ライブラリから PSPacer の通信量を設定できる機能を実装した。図4に示す通り、通信量に関する値の取り出し・設定は、MPI の属性値として与えるようにした。

```

int *rate, *matb, flag;
MPI_Attr_get(comm, YAMPI_PSP_MAXRATE, &rate, &flag);
MPI_Attr_get(comm, YAMPI_PSP_MATB, &matb, &flag);

*matb = *rate / n;
MPI_Attr_put(comm, YAMPI_PSP_MATB, (void *)matb);
    
```

図4 MATB の設定方法

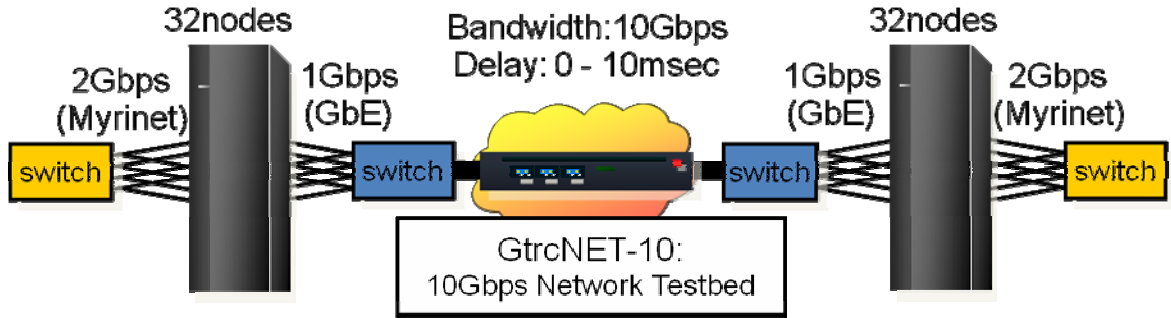
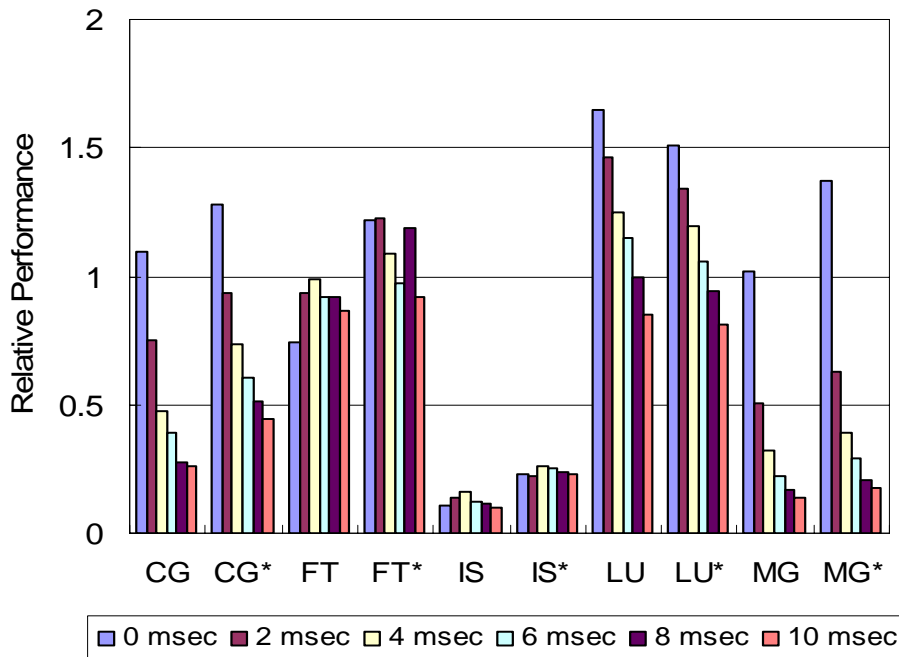


図5 実験環境

本システムを図5に示す実験環境で評価した。2つの32ノードのクラスタを GtrcNET-10 10 Gbps ネットワークテストベッドでつなげた。GtrcNET-10 は、通信遅延やパケット消失を人工的に加えられる装置である。実験では、遅延を 0 msec から 10 msec まで加えたときの性能を計測している。



CG*, FT*, IS*, LU*, MG*は MATB で通信量を調整した場合
 図6 NAS Parallel Benchmarks CG, FT, IS, LU, MG の実行結果

結果を図6に示す。64 ノードを Myrinet でつなげた場合の性能を 1 とした時の相対性能で示している。おおむね、MATB による通信量調整により高性能化につながっている。詳細は、論文[査読付 2]を参照のこと。

3.3 具体的成果

本稿で紹介した研究成果は、IEEE Cluster 国際会議で発表した[査読付 2]。また、平成18年度に研究したグリッド向けコレクティブ通信アルゴリズムは、Cluster 論文誌に採録された[査読付 3]。

4 ディペンダブルシステム

4.1 背景

科学技術計算や基幹業務系で使われているスーパーコンピュータや大規模サーバシステムの故障により、数か月におよぶ計算結果や会計情報などの重要なデータを喪失し、社会に対して甚大な損失を与える。このため、このようなコンピュータシステム向けに、ハードウェア、ソフトウェアの双方で耐故障性を実現するための研究開発がおこなわれてきた。コンピュータ技術の進展に伴い、基幹業務系コンピュータに限らず、ホームサーバ、デジタル TV、組込み型高性能サーバ、車載制御装置、通信制御装置、ロボット、携帯機器、モバイル・ウェアブルコンピュータなど、社会の隅々にコンピュータが使われ、これらコンピュータがネットワークにつながり、相互連携しながら豊かな環境が構築されつつある。このようなシステムの故障は、社会に対して致命的な損失を与える。我々は、このような組み込み機器向けに、組込みディペンダブルシステムを支援するオペレーティングシステムの研究開発を科学技術振興機構が実施している戦略的創造研究推進事業 CREST研究領域「実用化を目指した組込みシステム用ディペンダブル・オペレーティングシステム」において実施している。

4.2 内容

平成19年度までにCREST研究領域内で決定したディペンダビリティを向上するためのソフトウェアアーキテクチャを図 7 に示す。本研究領域では、LinuxOS機能の信頼性、安全性を主に扱っている。ディペンダビリティを提供するために新しいOS機能を研究開発するとき、Linux カーネルを直接変更するのではなく、安全かつ容易に拡張できる枠組み P-Bus を研究開発している。

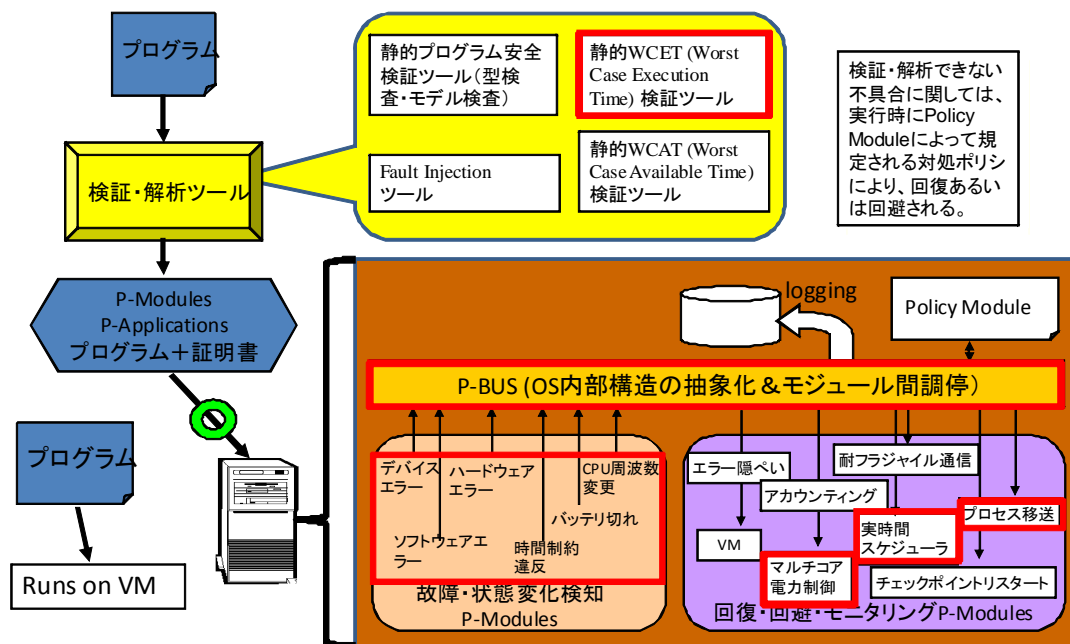


図 7 DEOS アーキテクチャ

平成19年度は、以下の項目の研究開発を進めた。

1) P-Bus 設計およびプロトタイプ

安全かつ低オーバーヘッドで拡張できる Linux カーネルを実現するために、Linux のローダブルカーネルモジュール機構を発展させた P-Bus と呼ばれる枠組みを研究開発してきている。P-Bus は、Linux ロードブルカーネルモジュールの OS 依存性、曖昧性、閉鎖性などの問題を解決する。図 8 に示す通り、P-Bus は、Core と Components に分かれる。

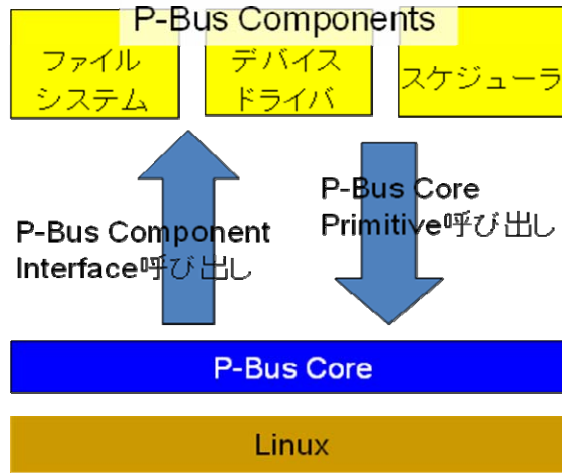


図 8 P-Bus アーキテクチャ

P-Bus Core

P-Bus Core は、カーネルの基本機能を抽象化し、カーネル拡張に必要なプリミティブを P-Bus Component に提供する。図 9 に示す通り、P-Bus Core は、デバイス、プロセス・スレッド、スケジューラ、メモリ管理など、カーネル機能をオブジェクトとして抽象化する。

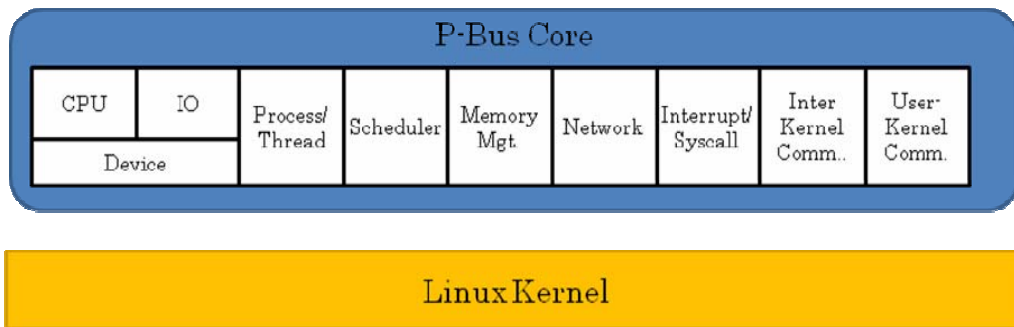


図 9 P-Bus Core オブジェクト

P-Bus Component

カーネル機能の変更あるいは拡張機能は P-Bus Component によって実現される。P-Bus Component は、P-Bus Core オブジェクトから呼び出される関数(インターフェイス)を定義する。たとえば、スケジューラのアルゴリズムを変更する P-Bus Component を記述する場合には、P-Bus Core スケジューラオブジェクトが呼び出す add_thread、switch_thread 関数を定義する。P-Bus Core スケジューラオブジェクトのプリミティブには、pbus_sched_add、pbus_sched_switch などがある。

スケジューラおよびメモリ管理、リモートカーネル間通信機構の P-Bus インターフェイスを設計し、プロトタイプを開発した。

- 2) 高性能マルチコアマルチプロセッサ管理機構
マルチコア上で複数 Linux カーネルを稼働させる機構を実現した。
- 3) シングルシステム機能を実現するための機構
シングル IP アドレスを提供する分散システムを実現するという観点から、Linux カーネルを直接変更して実現した。この経験をもとに、P-Bus に必要な機能を平成 20 年度設計・実装する。
- 4) 時間制約保証・検証ツール RETAS の SH プロセッサへの移植
我々が開発中の時間制約保証・検証ツール RETAS は、基本ブロック毎に実機上での実行時間を計測する。測定精度を上げるためにタイムスタンプカウンタを用いる。SH プロセッサはタイムスタンプカウンタは特権モードでしかアクセスできないため、タイムスタンプカウンタを制御するカーネルモジュールを開発した。
- 5) 障害管理システム設計&プロトタイプ
故障→異常→障害関係を表現する error relation tree と機器データベースから、障害が発生した時に、故障個所を自動判別するシステムの設計とプロトタイプを実装した。

4.3 具体的成果

P-Bus の設計に関しては、[発表 2]において口頭発表している。P-Bus の設計と並行してシングルシステム機能を実現する試みを行ってきている。論文[査読付 1]および口頭発表[発表 1]は、その成果である。また、障害管理システムに関する成果は、論文[査読付 4]および口頭発表[発表 3]で発表している。その他、本テーマに関連して 2 件の招待講演を行った[発表 4][発表 5]。

5 成果要覧

査読付論文リスト

- [査読付 1] Hiroya Matsuba, Yutaka Ishikawa: Single IP Address Cluster for Internet Servers, IEEE International Parallel and Distributed Processing Symposium (IPDPS 2007), 2007 (CD-ROM).
- [査読付 2] Ryousei Takano, Motohiko Matsuda, Tomohiro Kudoh, Yuetsu Kodama, Fumihiro Okazaki, Yutaka Ishikawa: Effects of Packet Pacing for MPI Programs in a Grid Environment, IEEE Cluster 2007 (CD-ROM).
- [査読付 3] Motohiko Matsuda, Tomohiro Kudoh, Yuetsu Kodama, Ryousei Takano and Yutaka Ishikawa: The design and implementation of MPI collective operations for clusters in long-and-fast networks, Cluster Computing, Springer pp. 45 – 55, Netherlands, 2007
- [査読付 4] Masato Sakai, Hiroya Matsuba and Yutaka Ishikawa: Fault Detection System Activated by Failure Information, Proceedings of the 13th Pacific Rim International Symposium on Dependable Computing (PRDC'07), pp. 19 - 26, December 2007

公開ソフトウェア

- [公開 1] 石川裕, 他: YAMPI, MPI 通信ライブラリ, <http://www.il.is.s.u-tokyo.ac.jp/>

その他の発表論文リスト

- [発表 1] 藤田肇, 松葉浩也, 石川裕: 柔軟な負荷分散を可能にする分散型シングル IP クラスタ, 情報処理学会 研究報告 2007-OS-106, pp. 1 - 8, 2007 年
- [発表 2] 平野貴仁, 藤田肇, 松葉浩也, 石川裕: PBus: 柔軟なカーネル機能拡張のためのインタフェース, 情報処理学会 研究報告 2007-OS-106, pp. 63 - 70, 2007 年

[発表 3]酒井将人, 松葉浩也, 石川裕: エラー情報から原因を特定する障害検知システム, 第 5 回
ディペンダブルシステムワークショップ, 日本ソフトウェア科学会ディペンダブルソフトウェア研究会,
函館, 2007 年

[発表 4]石川裕: ディペンダブルシステムソフトウェアにおける挑戦, 組込みシステムシンポジウム
2007, 情報処理学会, 東京, 2007 年

[発表 5] 石川裕: システムソフトウェアにおける検証技術への期待, 第 4 回システム検証の科学技
術シンポジウム, 日本ソフトウェア科学会ディペンダブルソフトウェア研究会, 名古屋, 2007 年

報道関連

[報道 1] 日経コンピュータ 2008.2.1 号: 東大, 最速スパコンを格安で民間開放, p23, 2008.

セキュリティを保証する言語処理系の研究

Optimization Verifying Compiler, Document Carrying Authorization, and LoA of Identities

佐藤周行

1 概要

われわれは、「検証付最適化コンパイラ」のもと、最適化に際してセマンティクスを保証するための検証の枠組と、最適化の結果、それが性能を確実に上げていることを検証する枠組を研究してきた。

2007年度は特に以下の研究を行った。

1. プログラムの性能改善のためのコンパイラ最適化を超えたアルゴリズム変換の理論。特に性能最適化の選択肢を広げるためにイディオム認識を拡張する研究をおこなった (Optimization Verifying Compiler)。
2. 分散環境における知識抽出。フォークソノミーのタグ付け行動を対象として知識抽出とその応用、評価を行った。
3. 認証におけるセキュリティ表現。セキュアな認証を用いて、認可を表現する枠組みの研究を行った。プラットフォームは言語処理系に限らず、一般に言語で表現されているもの一般である。特にワークフローを証明するための理論的な枠組みを提唱した。
4. ID の LoA の研究。ID の LoA は、強固かつ合理的な認証基盤を構築する上で必須の研究対象である。東大で運用中のサーバ証明書審査体制において認証強度の criteria を定めるための評価軸を定めることを行った。

2 Optimization Verifying Compiler

2.1 背景

コンパイラの研究の中心である最適化において、従来、研究の展開はアドホックなノウハウの蓄積によってなされてきた。しかし、副作用を許容する最適化、性能的にトレードオフを持つ最適化群の登場など、最適化が複雑・高度になった現在では、理論的基盤を持たなければ最適化適用の正しさの証明さえ危うくなってきている。事実 Gnu-C では、いまだに大量のバグレポートがなされている。

検証付きコンパイラ (Verifying Compiler) は、もともと Hoare の提唱したプログラムの正当性を併せて証明するコンパイラシステムとして 21 世紀の課題として提唱した概念である。本研究で目指すのは、コンパイラ最適化における上のような危機的状況に対応すべく最適化をその正しさの証明を込めて適用するコンパイラの構築であり、最適化検証付きコンパイラ (**Optimization Verifying Compiler**) と呼ぶべきものである。

従来の国内外における構成要素としての研究成果にはアセンブラの型付けによる動作正当性の証明 (Necula)、最適化の適用前後での意味保存の証明 (Necula)、最適化適用条件の時相論理等での形式化 (D Jones) を含む。さらに広義にはモデル検査の応用として動作正当性を証明することも含む。

現在、最適化検証付きコンパイラが検証することのひとつの柱は最適化適用の前後でプログラムの意味が保存されることである (Translation Validation)。証明を最適化の適用と同時に行うことを、アセンブラ言語に対する型理論として展開する。最適化適用はデータフロー解析を本質的に用いることから、データフロー理論を包含した型理論の構築につなげることができる。この最大の意義はデータフロー型の最適化の正当性を形式的に証明する体系を提示することである。また研究の基礎として、動作同一性証明に適したプログラム実行の適切な抽象化ができることになる。

また、最適化検証付きコンパイラはプログラム実行の抽象化を通して、最適化適用の最大の目的である性能向上の理論的解析のベースになる。現実的に妥当な数理的な性能モデルを仮定すると、その中で適用最適化が「真に」最適なコードを出すかどうかの検証ができる。(従来の最適化は、もとのコードと比較して、性能が良くなるかもしれないということを定性的にしか示せなかった。自動チューニング (Atlas 他) は、没理論でノウハウを適用することを目指しているものである。また、アーキテクチャ研究ではエミュレータが実質的なモデルの役割を果たしてきた。しかし、アンローリングなどのパラメトリックな最適化を解析するには数理的なモデルが必須である。) 性能モデルと、各種整数計画法の適用により、複数の最適化の組み合わせのどれが真に最適かを決定することができる。これはまさに性能面からみた検証であり、検証というプロセスが従来のアドホックな最適化研究に与えるインパクトの大きさを証明するものである。理論的な検証付きコンパイラが具体的な性能モデルの概念と結合して具体的な性能保証をすることは、理論と実践の両面での貢献が大きいものである。

2.2 内容

目的としてあげたゴールに対して 2007 年度、本研究では具体的に以下のようにテーマを設定した。

プログラムの性能改善のために、コンパイラ最適化を超えたアルゴリズム変換の理論を研究した。性能最適化は、従来から変数の R/W を変更しない意味的な制約の元で行われるのが通常であった。これを超えて、特定のパターンが出現したときに、あらかじめ用意された高速ルーチンで置き換える手法は最適化と呼ばず、「イディオム認識」と呼ばれ、細々と使われてきた。高速化が強く求められ、しかも超高級言語でのプログラム開発が普通のものになった今、今一度イディオム認識の効果を認識することが必要である。具体的には関数型言語における高階関数のパターンの認識を通じてアルゴリズムの変換を可能にして、性能改善のためのルーチンの探索空間を本質的に広げる研究を行った。

2.3 具体的成果

[発表 1]において、イディオム認識を高階関数まで拡張することでプログラムのテンプレートまでを含めたアルゴリズム変換を可能にした。従来の関数型言語におけるプログラム変換と、イディオム認識の橋渡しをする研究になった。

従来の枠組みでは一階の TRS (の逆) を用いて Fortran の組み込み関数の組み合わせ全体を探索していた。今回は高階の TRS を用いてアルゴリズムのテンプレートをイディオムとして認識し、探索空間を飛躍的に拡大した。

ただし、拡大した空間から最適解をどう探索するかはユーザのノウハウに依存する。具体的なプログラミングの経験から、高次のノウハウを蓄積する技術ははまだ課題として残っている。

3 分散環境における知識抽出

3.1 背景

分散環境下でのプログラミングモデルには、スケジューリング、セキュリティなどの重要な視点が Web-service の中に次々と取り込まれている。これらの表現が今後重要になる。現在、Semantic Web の動きが停滞し、オントロジーのサイトの閉鎖が相次ぐ中で、それとは逆にボトムアップに知識を積み上げるアプローチが提案されている。Web マイニングとして、エンドユーザが集団的に情報にアクセスしてタグ付けや感想の付加などのアクションをとる行動が対象になってきている。特に Folksonomy を解析することで、Web から有用な知識が抽出できるのでないかと考えられている。分散計算環境の振る舞いを記述する上でこの方法論の調査と展開が必要であった。

3.2 内容

Web マイニングとして、エンドユーザが集団的に情報にアクセスしてタグ付けや感想の付加などのアクションをとる行動が対象になってきている。これらはフォークソノミーとよばれ、現在の Web のアクティビティからいって注目すべき現象である。われわれはフォークソノミーのタグ付け行動を対象として知識抽出とその応用、評価を行っている。具体的には (1) タグの階層構造の構築とその評価 (2) SVM を利用したタグ付けの機械学習を行った。

3.3 具体的成果

[査読付 1]において、Folksonomy を解析することで Folksonomy のタグ付け行動を対象として知識抽出とその応用、評価を行った。階層構造の結果としては、直感的には有用な構造を得ることができた。また、評価に関しては、有意な結果を出すことは出来なかったものの、階層構造の新たな評価手法を提示することができた。機械学習の結果としては、正解率がマイクロ平均で 92% の有用な分類器を得た。また、オブジェクト当たりのタグ付け数を指標とすることで性能の良い分類器を選択することが可能になることが分かった。

また、時系列的な概念の生成過程を抽出することを試みている。この方向の研究はまだ実験的であるが、2008 年度の成果として以下の発表を予定している。

- Kawanaka, Sho., Sato, Hiroyuki.: Analysis of Chronological Tag Dependency in Folksonomy, to appear in Proc. SWWS'08.

4 認証におけるセキュリティ表現

4.1 背景

コンピュータが情報システムの中核をなすようになると、データへのアクセス権限の証明と、認証（アクセス元の本人性証明）が重要になっている。それは一つとしてアクセス権設定のためのロールモデルの研究を生み出し、もうひとつはデジタルデータを対象にした厳密な証明をもとめるPKIを生み出した。ロールモデルとしてRBAC（NIST Standard CS1.1）が実用化され、また組織のロールの研究からORBAC(<http://www.orbac.org/>)が研究されるなどこの分野の成長は著しい。一方、PKIは、現在電子署名や暗号化メール（RFC2633等）、さらに認証に使われるようになった。しかし、これらの学術的な成果が一般的な情報システムに組み込まれるには、まだ時間がかかるというのが共通の理解である。

他方において、サービス志向アーキテクチャにおけるサービス構築技術においてシステムどうしが通信を行なうワークフローの定式化がWS3C(<http://www.w3c.org/>)によって進行している。とくにWS-*の枠内においてWS-BPELによりビジネスプロセスを記述する流れが決定的なものになっている。ワークフローの定式化はGridその他、複数のプロセスのインタラクションにより計算が進行していくものにおいても重要である。ここではWS-BPELにとどまらない言語が提案、研究、実装されている。しかし、ここでのワークフローはロールモデル中の複雑な意思決定機構を反映できるようなものではなかった。

組織における意思決定をワークフローに反映させるためには、従来の貧弱なワークフロー表現では対応できないことはもはや明らかである。

4.2 内容

現在のサービス構築技術、またサービス志向アーキテクチャにおいて、ワークフローの表現が本質的になっている。現実のワークフローによるプロセスの進行においては、進行をアクティベートさせるものが重要である。とくに人間による「承認」が本質的に重要である。これが署名や印鑑で担保されることにより現実の世界は動いている。情報システム、およびその結合体は、ワークフローと進行のアクティベータの実現により、はじめてある程度の複雑度を持ったサービスを構築できる。ところが、現状は、要素的なサービス（DBアクセス、計算等）の記述は進んでいても、それらを結合させるワークフローの記述力は十分でない。

しかし、要素技術は存在する。それがロールモデルであり、PKIの特に電子署名技術による認証（証明）機能の提供である。不足しているのは、一定の動作がある言明を証明していることの形式化とその表現である。本研究では、ワークフロー、特にプロセスのアクティベーションに関する記述を定式化し、加えてPKIの電子署名技術を使って一連のプロセスに（実世界上での）セマンティクスを与えるためのフレームワークを確立する手法を研究した。

情報化以前においては、これら「承認」に基づくアクティベーションはドキュメントの流通によって行なわれてきた。ここではドキュメントが組織の意思を体現し、署名や押印がそれを承認するものとして扱われてきた。本研究は、特にこれに着目する。すなわち、ドキュメントが持つ

「意思をセマンティクスとして表現し、PKIによる電子署名で承認を表現する。これによって電子化されたドキュメントをワークフローの中に組み込むことが可能になる。

4.3 具体的成果

現実的な環境でのセキュリティモデルとその表現についての考察を行った。特に、データそのものが持つ認証情報と権限情報を計算環境下でどのように利用するかについて検討を行っている。

特に分散環境中にワークフローが検証可能なドキュメントをやりとりする枠組みを提唱した [発表 2, 発表 4]。これは検証可能なプログラムコードが分散環境を飛び交う PCC (Proof Carrying Code) 等と並列に論じられるべきものである。

PCC では検証器がコードの証明を受け取る。われわれの枠組みでは検証器はある要求を満たすドキュメントをモデルとして受け取る。証明の生成が困難であるのと同じように、「ある要求を満たす」(上司の許可を得ていることを証明している等) ことを充足するドキュメントの作成は業務におけるワークフローの完成と同等の困難さを持つ。したがってモデルとしてのドキュメントを受け渡すことが重要である。

5 LoA of Identity

5.1 背景

認証基盤で、ID のライフサイクル管理は基本的なものである。その構築において NIST の基準にも見られるように、「ID の保証基準」を定め、それによってコンテンツへのアクセスを制御する技術が一般的になってきた。この背景には、合理的なコストでセキュリティを担保することへの強い要請がある。

PKI プロジェクトで実施しているクライアント証明書発行のための UTCA と、サーバ証明書発行審査のための TRA は、強い認証強度をもつように CP/CPS を作成し、実際に運用しているにもかかわらずその認知度は一般的ではない。これはひとつには監査その他標準的な認証強度の認定プロセスを経ていないことが原因であるが、もうひとつには認証強度の criteria が定められていないことも原因としてあげられる。

5.2 内容

この研究では、認証強度の criteria を定めるための評価軸を定めることを行っている。特にサーバ証明書において FQDN のすべてのドメインについてまで保証を行うための審査体制の構築の研究を行った。また、これをもとに、EV 証明書と WTCA 証明書の中間の強度を持つ証明書を合理的なコストで発行するための基準についても検討を行った。

5.3 具体的成果

強い認証のための基盤の構築については認証の LoA の研究とともに精力的に行っている [発表 3]。

なお、研究の途中で Open source software のアップグレードのコストに注目せざるを得ない状況に至った。ソフトウェアの信頼性を高めるためにバグフィックスや機能拡張の成果を誠実に反映

することで、実は製品購入以外の人的コストが跳ね上がっていることが観察できるようになった。この観察の結果は以下の論文で 2008 年度に成果を公表する予定である。

- Sato, Hiroyuki: Delay Front and Skip Interval for Assessing Open Source Software Maintenance of System, to appear in Proc. IEEE 2008 COMPSAC.

6 成果要覧

査読付論文リスト

[査読付 1] Ogino, Ken, Sato, Hiroyuki: Extracting Tag Hierarchy from Folksonomy, Semantic Web and Web Services 2007, 118–123, 2007.

その他の発表論文リスト

[発表 1] Sato, Hiroyuki: Program Transformation for Performance Tuning – Beyond Compiler Optimizations, Proc. 11th Workshop on Innovative Architecture for Future Generation High-Performance Processors and Systems, 2008.

[発表 2] Sato, Hiroyuki: Documents in Workflow – Extending Authorization Scheme by Documents, Proc. Int’l Symposium on Secure-Life Electronics, 229–234, 2008.

[発表 3] 西村健, 佐藤周行: 東京大学におけるサーバ証明書発行体制の構築と課題, IPSJ 研究報告 2008-DSM-48, 79–84, 2008 年.

[発表 4] Sato, Hiroyuki: Document Carrying Authorization, Workshop on Computer Science and Category Theory (oral), 2008.

創薬・バイオ新基盤の技術開発へ向けた タンパク質反応全電子シミュレーション・システムの研究開発

佐藤 文俊

1 概要

当グループは、タンパク質をターゲットに量子論に基づく精密なシミュレーションシステムを開発している。ベースとなるソフトウェアは、大規模な系を取り扱うことができるよう独自に開発した、ガウス型関数を用いた密度汎関数法プログラム ProteinDF である。これを用いてこれまで 2000 年に 104 残基の金属タンパク質シクロム c、2006 年には 306 残基のインスリン 6 量体の全電子波動関数計算に世界で初めて成功した。本システムの研究開発の目的は ProteinDF を基に、タンパク質の新しい研究法、ならびに創薬やバイオ素子の設計などへの応用に必要な機能を追加・拡張し、インフラを整備し、創薬・バイオの新基盤技術開発に向けたタンパク質反応全電子シミュレーションシステムへと発展させ、これを公開することである。本システムは以下の 4 つの研究開発項目(サブグループ)、(1)密度汎関数法による超大規模タンパク質全電子計算エンジン、(2)タンパク質の構造最適化、分子動力学、自由エネルギー計算機能、(3)タンパク質波動関数 DB の大幅更新、(4)シミュレーションを支援し、物性を評価・解析できる高品位 GUI、に関する諸成果を統括したシステムである。本システムはオブジェクト指向言語 C++で構築されるため、各研究開発項目は独立かつ安全に開発することができた。今年度は、RSS21 最終リリース版の作成を行った。(1)ハイブリッド汎関数、新高速化手法、新初期波動関数作成法および新収束法の各開発とリファクタリング、(2)自由エネルギー計算のための高速分子動力学、(3)100 種類のタンパク質波動関数の完成、(4)タンパク質構造モデリング機能の強化、反応シミュレーション解析のための新しい表現方法の開発を組み込んで、ProteinDF システムを完成させた。また、密度汎関数法-配置間相互作用法による電子励起・電子移動計算、新しい高速多重極展開計算法、何種類かのインスリン 2 量体解離シミュレーションの研究も行った。

2 ProteinDF システムの開発

2.1 背景

創薬技術、バイオテクノロジーの発展は世界各国がしのぎを削る戦略目標である。その定番の手段となったゲノム情報の比較研究は、病原や薬剤に関与するタンパク質の同定などに威力を発揮するが、設計の決め手となる薬剤の効率やタンパク質の性能を調べることができない。そのため、タンパク質の反応シミュレーションは本来創薬に欠かすことができない基盤技術である。しかし、これまで実用的で信用に足る計算方法やそれに基づくソフトウェアは確立していなかった。タンパク質は電子移動、光励起、化学反応などの反応を制御して機能を発現するため、電子の働きが最も重要である。したがって、薬剤の設計には薬剤、基質、タンパク質を対等に取り扱う精密で大規模な量子化学計算によるシミュレーションが必須である。また、タンパク質は反応のプロセスで変形運動をする。この部分はニュートン方程式を用いて追跡するのが定石である。タンパク質どうしあるいはタンパク質と溶媒の相互作用には統計力学が必要である。このようにタンパク質の反応シミュレーションは必然的にマルチスケール・マルチフィジックスであり、その技術開発そのものも世界の戦略目標である。

我が国は文部科学省 IT プログラム「戦略的基盤ソフトウェアの開発」プロジェクト(前プロジェクト)などを通して、この分野のソフトウェア開発において世界をリードしている。当グループは ProteinDF を開発し、大規模タンパク質(306 残基のインスリン 6 量体)の全電子計算に成功した。ベースとする密度汎関数法はタンパク質のような金属とペプチド鎖を持つ分子の計算で電子相関を取り込んだ精密な計算ができるうえ、最近では、反応解析に必要な励起状態の計算も実験値を良く再現することが明らかになっている。間もなくペタフロップス超級の汎用スーパーコンピュータが登場する 21 世紀初頭の今、タンパク質の反応を全電子シミュレーションによって追跡できる環境が揃いつつある。

2.2 内容

本研究の目的は、ProteinDF をベースに、タンパク質反応の全電子シミュレーションシステムを世界に先駆けて開発することである。この新しい手法は創薬の信頼性の高い基盤技術となる。実際に、超々速効性のインスリン設計などのシミュレーションを通して、その実用性も検証した。基礎研究のみならず、医薬品開発の高品位化、高効率化ならびに次世代の医薬品研究開発モデルの創成といった応用貢献に本システムの開発はまさに急務の課題である。方法特許やビジネスモデル特許などを含め欧米に対するリードに寄与することも期待できる。触媒、分子素子、環境物質などへの応用も大いに期待してよいだろう。Fig. 1 に本プロジェクトで研究開発したタンパク質反応全電子シミュレーション・システム(ProteinDF システム)の概念図を示す。本システムは以下の(1)~(4)の研究開発項目を統括したシステムである。さらに、今年度、密度汎関数法-配置間相互作用法(ProteinCI)も設計が終了し、常に最先端のシミュレーション技術が取り込めるシステム構造となった。

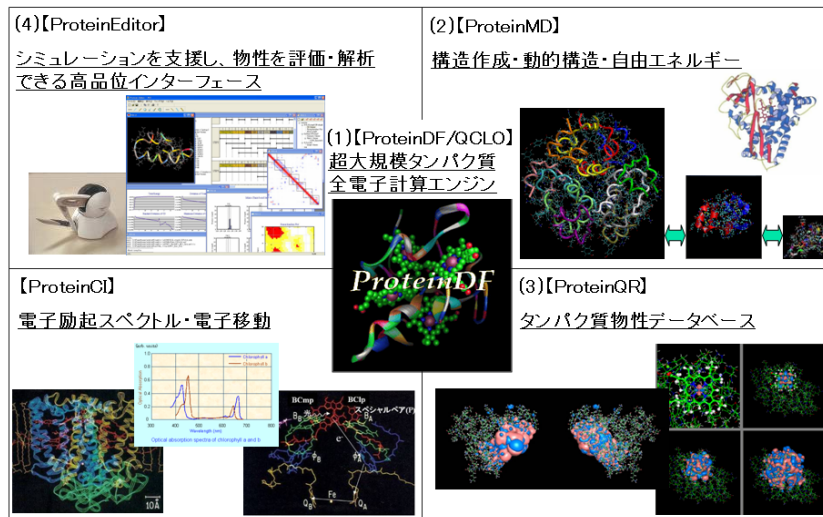


Fig. 1: ProteinDF システムの概念図

2.3 具体的成果

(1) 密度汎関数法による超大規模タンパク質全電子計算エンジンの機能拡充 プログラム名: ProteinDF, QCLO

a) 交換相関汎関数の機能拡充

あらたな交換汎関数として、密度勾配近似である Becke88 および LYP ならびに B3LYP をサポートした。近年、一般化学分子の密度汎関数計算では、ハートリー・フォック(HF)交換項や非局所密度近似補正も組み合わせたハイブリッド汎関数が使用されており、なかでも B3LYP ハイブリッド汎関数は数多く実験結果と比較が行われ、その高精度の結果から業界標準となっている。ハイブリッド汎関数の一つである Becke3 パラメータによるハイブリッド汎関数は、以下の式により定義される。

$$E_{XC}^{B3LYP} = AE_X^S + (1-A)E_X^{HF} + BE_X^{B88} + CE_C^{LYP} + (1-C)E_C^{VWN} \quad (1)$$

ここで E_X^{HF} は Fock の交換項であり、一般に以下の 4 中心 2 電子積分から求められる。

$$\langle pq|rs\rangle \equiv \iint g_p(\mathbf{r})g_q(\mathbf{r})\frac{1}{|\mathbf{r}-\mathbf{r}'|}g_r(\mathbf{r}')g_s(\mathbf{r}')d\mathbf{r}d\mathbf{r}' \quad (2)$$

(2)式の 4 中心積分計算を Obara-Saika の積分公式を用いて実装した。この計算は理論上軌道数 N に対し $O(N^4)$ の依存性があるため、タンパク質のような大規模分子計算では膨大な計算量になる。効率よく 4 中心 2 電子積分を求めることが要求されるため、いくつかの高速化手法を組み込んだ。

(i) Fock 交換項生成アルゴリズム

Fock の交換項を求めるには、多くの場合 integral-driven 法と呼ばれる、4 中心積分の対称性を利用して積分計算量を削減する方法が用いられる。ProteinDF にもこの手法を取り入れ、効率よく Fock の交換項を求められるようになった。一方、並列計算機を用いて Fock の交換項を求める場合、integral-driven 法では多くの通信が発生し、通信による遅延が問題になる場合がある。そこで ProteinDF では、integral-driven 法に加えて RT 法も実装した。RT 法は integral-driven 法に比べ 2 倍の積分計算が必要となるが、通信量は大幅に減少するため、大規模分子を並列計算する場合に適している。計算サイズによりユーザが選択できるように実装した。

(ii) スクリーニング

計算すべき 4 中心積の数を少なくする手段として、積分値の極めて小さな積分計算を省略する方法(スクリーニング)は極めて効果的である。これにより、4 中心積分の計算量を $O(N^4)$ から $O(N^2)$ 程度にまで減らすことができる。具体的には、シュワルツの不等式によるカットオフ、ダイレクト SCF 法と相性の良い密度行列によるカットオフ、長距離相互作用によるカットオフといった方法を組み合わせ、積分すべき個数を絞り込む。ProteinDF では、これらのスクリーニング手法も実装した。タンパク質のような大規模分子系において、精度良くパフォーマンスを発揮できるスクリーニング手法の組み合わせ・チューニングを行った。

b) RI-K エンジンの追加実装

ProteinDF では resolution of identity (RI) 法を採用しており、

$$\langle pq|\alpha\rangle = \iint g_p(\mathbf{r})g_q(\mathbf{r})\frac{1}{|\mathbf{r}-\mathbf{r}'|}g_\alpha^p(\mathbf{r}')d\mathbf{r}d\mathbf{r}' \quad (3)$$

という 3 中心 2 電子積分を高速チューニングしている。(2)式で表わされる 4 中心積分は、以下に示す 3 中心積分を利用する式に変形することが可能である。

$$\langle pq|rs\rangle = \sum_{tu} \langle pq|\alpha\rangle V_{\alpha\beta}^{-1} \langle \beta|rs\rangle \quad (4)$$

ここで $V_{\alpha\beta}^{-1}$ は $\langle \alpha|\beta\rangle^{-1}$ である。なお、この積分は $\langle pq|v\rangle = \sum_t V_{vt}^{-1/2} \langle pq|t\rangle$ とすることにより対称性の良い形に変形することができるため、3 中心積分を 2 度行う必要はない。したがって 3 中心積分を利用した Fock の交換項は次のように求められる。

$$K_{pq} = \sum_{vrs} \langle pr|v\rangle \langle v|qs\rangle P_{rs} \quad (5)$$

これにより、高速な 3 中心積分ルーチンを用いて Fock の交換項を求めることが可能となった。3 中心積分は 4 中心積分に比べて少ないメモリで計算可能なため、積分値をメモリ上に残したまま SCF 計算が可能な大規模計算機では有利に働くようになった。

c) 大規模タンパク質用高品位ソルバプログラムの改良

大規模タンパク質の計算が達成できるためのソルバプログラムの開発を行った。大規模計算の成否は、初期値の良さにかかっている。精度の高い初期値を得るための計算を、簡単にしかも効率よく安全に行なうことができるソルバプログラムには、タンパク質の立体構造を考慮した計算手法が有効であると考え、Structure-based QCLO 法を開発した。これまで本研究では、高精度の初期値の作成するために、独自に擬カノニカル局在化軌道 (Quasi-Canonical Localized Orbital : QCLO) を提案した。QCLO は一種の局在化軌道であり、その軌道は任意の単位では局在化しているものの、その単位内ではカノニカルな性質を持つ。従って、QCLO を用いることにより、各単位で、その化学的性質

が表現された、全分子の局在化軌道を得ることができる。これまでタンパク質の全電子計算では、タンパク質の一次構造に沿って各アミノ酸残基の QCLO を繋ぎ合わせるにより、大きなペプチド鎖分子の高精度の初期値を作成していた。本研究では、タンパク質の静電相互作用や水素結合などの相互作用に注目し、QCLO を用いた高精度の初期値を作成すると共に、計算化学的手法を用いてそれらの相互作用を評価することができた。データは示さないが、従来の QCLO 法と本研究で開発した Structure-based QCLO 法による SCF 計算初期値と収束値の Mulliken 電荷の差を比較したところ、新手法を採用した領域では、ほとんど差がなくなった。立体構造を考慮した QCLO 法を利用すれば、特に巨大なタンパク質の計算におけるトライ&エラーにかかる時間、設備の節約に有用であり、1000 残基クラスの巨大なサイズのタンパク質の計算においても非常に安全な計算が行えるだろう。

(2) タンパク質の構造最適化、分子動力学、自由エネルギー計算機能の拡充

プログラム名: ProteinMD

a) 新 FMM 法の開発

タンパク質のような大規模分子の量子化学計算では、クーロン相互作用の計算に膨大な時間を要する。この問題を解決する1手法として、高速多重極展開法(FMM)を量子化学計算に応用する試みがある。Gauss型基底関数系を用いた量子化学計算では、広がりを持った原子軌道間のクーロン相互作用を計算する必要があるが、十分遠く離れた2つの原子軌道間の相互作用は、半古典的な2粒子間の相互作用に置き換えることができる。本研究では、この性質を積極的に利用することにより、従来の古典FMMの計算方法にわずかな変更を加えるだけで、量子化学計算のクーロン項を高速高精度に計算できることを示した。FMMを用いた電子間クーロン相互作用計算の高速化には、McMurchie-Davidsonの手法をベースにした。この手法を用いて、カーテンアンGauss型関数の積をエルミートGauss型関数で展開し、その2電子積分をBoys関数で書き表した。また、エラー関数の性質により妥当性のある極限が設定できる(この極限を半古典極限と名付けた)。すると、S波のみからなるとき、通常古典クーロン相互作用の形になるので、GreengardらによるFMMの手法を適用することができる。本研究では、それ以外の時にも同様に多重極展開を適用することが可能であるように新たに立式することに成功した。本手法の有効性を確認するために、水分子系(16, 33, 59, 94, 152, 207, 277, 365個)についてクーロンエネルギーの計算を行った。Fig. 2で示すように、本手法におけるクーロン相互作用エネルギーの計算誤差は 10^{-7} 以下であった。本手法は古典FMMの自然な拡張になっているため、従来からのFMMに対する様々な改良法を容易に取り入れることが可能となった。

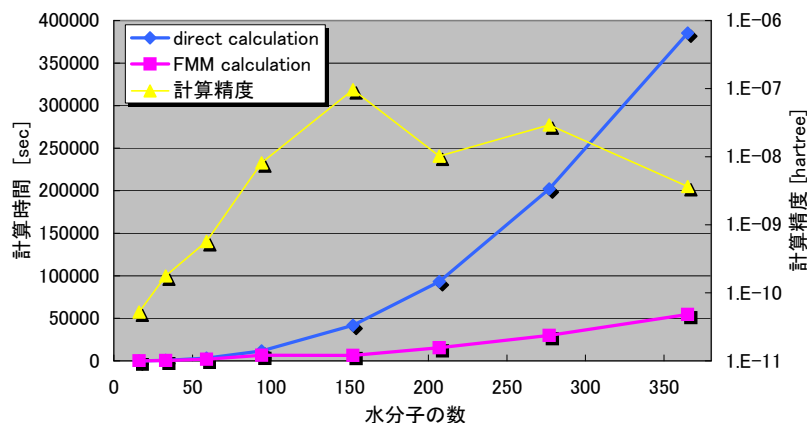


Fig. 2: ダイレクト法と新しい FMM 法による計算時間と精度の比較

b) 分子動力学シミュレーション

タンパク質の立体構造の取り扱いには非常に重要でまた多岐にわたる。構造決定やダイナミクスの再現などが代表例である。ProteinMD はそれらの実現を目的としたソフトウェアである。

(i) タンパク質計算構造の構築法

電子状態計算にはタンパク質の立体構造の決定は必須で、通常 Protein Data Bank (PDB)などのデータを参考にする。しかし、これらの立体構造が必ずしも化学的に適切な構造を保障しているわけ

ではない。この問題に対して、現在 ProteinDF 計算で標準となった手法を開発した。PDB に登録されている立体構造はある程度の構造最適化(リファインメント)が為されていることが多いが、それでも主鎖の 2 面角や原子間距離に異常がみられる。これらは電子状態計算結果に大きな影響を与えてしまう。また、たとえ歪みを取るにしても、全体のタンパク質構造は参照する PDB 構造からなるべく変化させたくない。そこで、この異常を適切に取り除く古典分子力場による構造最適化法によるタンパク質計算構造作成法を提案した。また、密度汎関数法では、真空中のタンパク質の全電子状態を安定に計算できないことも明らかとなった。そこで、全電子状態計算が安定に収束するような溶媒の与え方と計算シナリオとのカップリング法もこれらに追加した。

(ii) MD 計算高速化と ProteinMD 特別版

タンパク質の自由エネルギー計算法のための高速化手法を開発した。タンパク質の自由エネルギー計算法は様々提案されているが、どのような方法でも、取り扱う系はタンパク質のみならず溶媒も含めた大規模なものであり、さらに、自由エネルギー計算の精度を上げるには、サンプル数を増やすことが必須となる。その結果、タンパク質の自由エネルギー計算は、大規模系の長時間シミュレーションとなる。その根本的な解決法は、近似テクニックではなく、シミュレーションの高速化である。本研究では、ProteinMD の高速化を最優先課題とし、ソースコードを一から作り直した。また、長距離相互作用計算のアクセラレータである MD-Grape3 を利用できるよう改良するとともに、無駄なオーバーヘッドを極力減らすことに成功した。この MD-Grape3 が利用できる ProteinMD 特別版を RSS21 最終リリース版に公開した。

(3) タンパク質波動関数データベースの大幅更新

プログラム名: ProteinQR

本研究では、目標であった 100 種程度の DB を完成させた (Table 1)。この成功にとどまらず、タンパク質波動関数データベースの標準化のために、今後とも逐次 DB の更新を続ける。

(4) シミュレーションを支援し、物性を評価・解析できる高品位 GUI の拡充

プログラム名: ProteinEditor

タンパク質の密度汎関数法による全電子計算をサポートするまったく新しい GUI、ProteinEditor の開発に成功した。本タンパク質反応全電子シミュレーションシステムでは解析する対象は動的となり、データ量も物理量の種類も格段に増える。独自の経験と技術をつぎ込んできた ProteinEditor をベースに、本システムの GUI へと発展させることが最適な開発方法である。薬剤設計のための構造モデリング機能を強化し、データベースとのリンクや新しい物性量を表現する手段を開発することによって、より高度な GUI へと成長させた。

a) モジュールの連携操作

これまでに ProteinEditor は 100 残基規模の一点全電子計算をサポートできるように開発してきた。今年度は計算規模を増すと同時に、新しい自動計算法を導入し、ProteinEditor と連携して動作させた。これにより、ProteinEditor で作成した計算構造と自動計算シナリオを直接計算サーバーへジョブ投入することが可能となった。また、その計算結果も逐次計算サーバーにアクセスして取り出すことも可能となった。さらに ProteinDF システムでは、タンパク質全電子計算を行うための計算構造として PDB データを用いているが、wwPDB のデータ修正プロジェクトにより IUPAC(国際純正・応用化学連合)命名法に従うフォーマットに修正されたため、ProteinEditor ではこの IUPAC 命名法に基づく PDB データ読み込みにも対応した。

b) タンパク質構造モデリング機能を強化

(2)b(i)で述べたとおり、タンパク質のシミュレーションにおいてタンパク質立体構造の決定は重要で、計算結果に大きく影響を与える場合が多い。また、タンパク質はやわらかく、他の分子と相互作用する場合には変形する。精細でフレキシブルな立体構造を作成するためには、細部にわたって構造の編集ができ、しかもその操作は分かりやすいものである必要がある。そこで今年度は、これまで ProteinEditor に実装されていた水素付加やアミノ酸置換といった機能の高品位化とともに、タンパク質構造のためのモデリング機能 ModelingEditor の機能を強化した。Fig. 3 に ModelingEditor の操作画面を示す。プロテアーゼインヒビターの 35 残基目のアミノ酸残基の編集を行っている。

Table 1: タンパク質波動関数 DB のリスト

タンパク質	残基数	原子数	軌道数	タンパク質	残基数	原子数	軌道数	タンパク質	残基数	原子数	軌道数
アミロイドβペプチド1-28番残基 [簡易中性化](IAMC)	28	441	2506	ミニプロテインリガンド(I18T)	15	220	1274	カンナビノイド受容体 338-346 番残基 (ILVR)	9	158	820
グルカゴン類似体(BHO)	29	470	2650	アミロイドβペプチド1-28番残基 [ダミーチャージによる中性化](IAMC)	28	437	2498	低コクシジウムペプチド(IMO2)	12	229	1270
ユビキチン(IC3T)	76	1231	6685	クラベン類似体(ICRN)	46	642	3627	模倣機元ペプチド(IC8Y)	14	213	1186
クラベン(IEJG)	46	642	3627	ヒストン H3 の C 末端類似ペプチド (ICT6)	9	124	691	アポリポプロテイン E263-286 番残基 (IOEF)	19	390	2182
インターロキニン1(BBN)	133	2157	11909	両親媒性塩基性ペプチド(DJF)	15	238	1267	DMPD/HPC バイセル下のメチオニン-エンケファリン(IPLW)	5	75	439
サイトカニン受容体(IC8F)	102	1660	9329	両親媒性酸性ペプチド(DNG)	15	217	1225	DMPG バイセル下のメチオニン-エンケファリン(IPLX)	5	75	439
コクリン LCCLドメイン(IJB)	100	1475	8196	アズリン7 残基モデル[真空中](IDZ)	7+Cu*	97	539	電位依存性 Na ⁺ チャネルタンパク質 α サブユニット(IQ39)	21	356	1944
RGD ペプチド異性体 A, α 銅(I1UV)	11	135	838	出芽酵母より仮定された WW ドメイン (IE0N)	27	452	2528	C 型肺炎ウイルス非構造タンパクの膜結合ドメイン(IR7E)	31	532	2944
RGD ペプチド異性体 B, α 銅(I1VU)	11	135	838	アズリン7 残基モデル[真空中](IDZ)	7+Cu*	97	539	60S リボソームの C 末端ドメイン h13(IS4)	13	182	1094
グルカゴン(ICGN)	29	471	2673	水分子を含む RGD ペプチド[異性体 A, α 銅(I1UV)]	11+H ₂ O ×136	543	2606	ヘビ毒 SRTX 前駆体(SRB)	21	333	1943
オキシトシン(INPO)	9	134	769	グアニン(IGNA)	13	163	971	SeaA ATP 合成酵素の C 末端ドメイン (ISX)	22	336	1868
オキシトシン(IEQK)	102	1599	8844	ペプチド MBH12(IJ4M)	14	236	1319	トロンボモジュリンの表皮成長因子ドメイン(ITMR)	19	281	1595
ヒストン H3 ペプチド(IC8S)	9	127	704	バクテリオン[ダミーチャージによる中性化](IJK4)	6	95	579	ラバマイシン標的タンパク質 2438-2470 番残基(IW1N)	33	542	3055
バクテリオン[簡易中性化](IJK4)	6	95	579	HIV-1 中の Gp41 の 659-671 番残基ペプチド(ILCZ)	13	236	1257	サンリの毒より検出されたペプチド OmTx3(IWQ5)	23	320	1880
心スルチムタント(IHLS)	51	782	4439	被覆タンパク複合体 p23 の近位細胞内領域(IM23)	13	260	1374	ヘマグルチン糖結合ペプチド(IXOP)	20	287	1619
インスリンミュータント(Aspart)	51	784	4464	トランスデューシンのγサブユニット 60-71 番残基(IMP6)	12	188	1036	クランチュラの毒より検出されたペプチド HWTX-X(IY29)	28	400	2254
インスリンミュータント(Laspro)	51	786	4461	シクロピラチン O(I1NB)	30	419	2390	スタフィロコッカススクレアーゼ 97-109 番残基(2FNZ)	13	209	1120
インスリンミュータント(Glamine)	52	813	4261	ニューロキニン B(I1P9)	10	160	926	ヒトプリオンタンパク質 180-195 番残基 (2V4)	16	256	1366
インスリン単量体(IAI)	51	786	4461	アセチルコリン受容体の免疫領域(ITOR)	10	153	866	アズリン7 残基モデル[水+マリケン電荷](IDZ)	7+Cu*	97	539
インスリン6量体(IAI)	306	4716	26706	甲殻類シメクトンペプチド(IV46)	9	122	724	アズリン15 残基モデル[真空中](IDZ)	15+Cu*	234	1283
ペプチドトリプシン阻害剤(IJBN)	14	213	1186	ウシラクフェリン(IY5C)	11	227	1226	アズリン15 残基モデル[水+マリケン電荷](IDZ)	15+Cu*	234	1283
鎌倉アミロイドβペプチド(IKUW)	10	139	775	シムレリン類似ペプチド(IYT6)	10	134	790	アズリン24 残基モデル[真空中](IDZ)	24+Cu*	364	2091
ラスタリン A(I1J3)	12	168	954	アポリポプロテイン C 前駆体 35-53 番残基 (IALF)	19	339	1856	アズリン29 残基モデル[水+マリケン電荷](IDZ)	29+Cu*	441	2305
タキキニンペプチド(IMYU)	12	186	1032	副甲状腺ホルモン受容体の結合ドメイン (IBL1)	31	502	2783	グリシン(GLY) B3LYP	1	10	55
グアニンスクレオチド結合タンパク質 (I1VZ)	11	185	1007	セロピオピドラーゼの C 末端部(ICBH)	36	495	2846	アスパラギン酸(ASP) B3LYP	1	15	93
ミトリン J29(I1STP)	21	295	1661	βアドレナリン受容体 345-359 番残基[ダミーチャージによる中性化](IDEP)	15	275	1490	メチオニン(MET) B3LYP	1	20	112
Blood Depressing Substance-I(I1BDS)	43	629	3615	インテグリン α 細胞膜内ドメイン(IDPK)	20	325	1833	GLY-HIS シベプチド結合 B3LYP	2	27	159
キモトシプシンシヒター(IQH2)	46	671	3822	ユビキチン 1-17 番残基ミュータントペプチド(IEQG)	17	285	1524	シトクロム c の 20-22 番残基(IHRC)	3	56	294
抗腫瘍ペプチド(ID6X)	13	274	1514	ビスコトキシン(IEQ0)	46	667	3719	B3LYP			
Bowman-birk 阻害剤のトリプシン活性ループ(I1GM2)	11	166	938	グアニン前駆体(IGNB)	13	163	971				
アンジオテンシン I(I1N9U)	10	182	1015	サルコピリン(IJDM)	31	549	2971				
アミロイドβペプチド 25-35 番残基 (IQCM)	11	155	830	エキセンディン4(IJRL)	39	573	3220				
トランスサイチン(IRVS)	11	172	939	HIV-1 gp41 659-671 番残基(ILB0)	13	236	1257				
アセチルコリン受容体の細胞外領域 (ITOS)	10	141	814	インスリン前駆成長因子1の拮抗剤ペプチド(ILB7)	16	255	1440				
シニオン(I1UAC)	10	138	815								
RGD ペプチド異性体 B, α 銅(I1FUL, 塩橋の相方を取の込む計算)	11	135	838								
リボソームタンパク質 P2(1S4H)	13	181	1085								
さざりペプチド P01(IACW)	29	410	2386								
CFTR タンパク質(I1CKW)	25	413	2263								

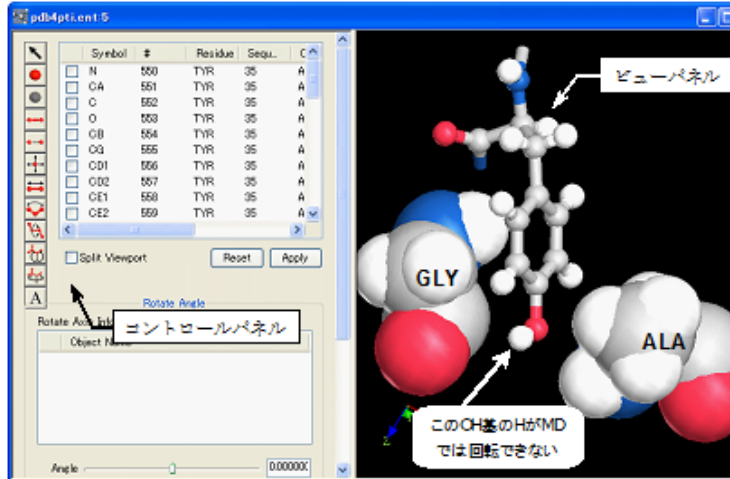


Fig. 3: ModelingEditor のスナップショット (Proteinase/Inhibitor: 4TPI の TYR35 周辺)

左窓(コントロールパネル): 構造編集のための作業機能一覧.

右窓(ビューパネル): 3次元構造と編集作業状態の表示.

ModelingEditor は、分子グラフィックスで選択された構造に対してダミーチャージの追加、原子の追加と削除や原子位置、結合距離、結合角、二面角などの変更、また選択された部分構造に対して点中心の回転、軸中心の回転など分子グラフィックスからインタラクティブに原子(単独および特定原子種集団)、基、側鎖、アミノ酸残基(群)、基質、配位子といった様々な単位で構造を編集することが

できる。これに伴い、MD 法や構造最適化法で局所解に落ち込み除去できなかった Fig. 3 のような構造ひずみを対話的、かつ視覚的に編集することが可能になった。手作業による構造編集後には、構造最適化が必要である。このとき、ユーザが独自に力場を指定して局所部分構造最適化を行う手法が簡便かつ有効である。そこで構造妥当性評価機能を用いて評価された構造歪を除去するために、簡易部分構造最適化機能も実装した。これらは連携して動作させることができ、ユーザが直感的に構造編集を行えるようになった。

c) 反応シミュレーション解析のための新しい表現方法

反応シミュレーション解析では、全く新しいポスト処理が必要である。今年度はまず反応過程における分子軌道の変化をストレスなく観察できる表現手法を開発した。また、タンパク質の重要な反応の一つに酵素反応がある。タンパク質(酵素)の表面や内部にはポケットと呼ばれるくぼみや空洞(鍵穴)があり、作用を受ける基質分子(鍵)はポケットに結合して、適切な化学変化を受ける。一般に、この活性部位の立体構造は、鍵穴と鍵の関係のように特定の基質とぴったり合うようになっており、酵素は基質の立体構造を認識することができる分子といえる。このドッキングのシミュレーション(Virtual Screening)を行う上で、ポケットの認識は非常に重要なステップである。ProteinEditor では、ポケットの検出方法としてよく使われている POCKET の派生である LIGSITE のアルゴリズムを改良し、ポケットの検出、及び表示を非常に高速に行うことができるようにした。

d) タンパク質モデリング VR システムの拡張(本研究の一部は東大生研・選定研究による)

モデリング機能をベースに Virtual Reality (VR) 技術を取り入れ、複雑なタンパク質立体構造の 3 次元空間における原子位置関係と原子間に生じる力の場を直感的に把握し、対話的な構造(原子、アミノ酸残基、基質、配位子)の編集を支援する機能として、タンパク質モデリング VR システムの開発を行った。従来の分子モデリングでは、奥行き感があまり得られない 2 次元的な表示と平面的な動きしか表すことの出来ないマウスを用いることが多かったため、編集する構造の位置を直感的に決定することは困難であった。また数十原子以下の一般分子ではその必要性も低かった。数千原子からなる巨大で複雑なタンパク質の密構造を編集する場合には、構造の 3 次元空間的位置関係や力を直感的に把握し、最適な状態へと構造を変形させる操作の機能開発が必要である。本システムは反力デバイス PHANTOM による触感インターフェース、および裸眼立体視ディスプレイによる視覚インターフェースを取り入れた。モデリングは反力デバイスを使用してタンパク質の異常構造を変異・修正させるが、その時に修正部位にかかる力を MD で計算し、デバイスにフィードバックさせることによって触感を通じてタンパク質が作る場を認識することを目指し開発を進めた。対話的なモデリングを実現するには、タンパク質の力計算はリアルタイム性が重要となるため、MD-Grape3 の組み込みも行った。裸眼立体視の表示は、SHARP 製 3D 液晶を搭載したノート PC もしくは液晶ディスプレイ、および SGI 製の立体視ライブラリ ISL に対応したハードウェアに対応した。

(5) ProteinDF システム最終バージョンの完成

a) ソースコード

ISO/IEC 14882:1998 準拠の C++ で記述した。標準ライブラリを使用するように変更し、規格に沿った C++ コンパイラにてビルドが可能である。多様な計算機環境でも自動的にライブラリを検索し、ビルドできるスクリプトを用意した。行列の線形代数演算については、LAPACK・ScaLAPACK ライブラリを利用し、各計算機に適したライブラリとリンクすることにより高速な行列演算が可能となった。

b) 並列化

行列演算以外の計算コストが高価であるルーチンは、分子積分・数値積分である。これらを独自に並列化し、並列化率は 99% を超えるプログラムが完成した。256CPU クラスの並列化はすでに十分となり、1,000CPU クラスの並列化が手に届くところとなった。このような高い並列化率をもつカノニカル分子軌道プログラムは大変稀であることを強調したい。

c) ハイブリッド汎関数 B3LYP によるタンパク質電子状態計算

RGD ペプチドをモデル分子として、SVWN・B88LYP(BLYP)・B3LYP の交換相関汎関数を用い、電子状態計算を行った。それら計算によって得られた Mulliken 電荷において、3 つの交換相関汎関数で際立った違いは見られなかったが、エネルギー準位において明確な違いが見られた(Fig. 4)。

B3LYP は低分子の電子状態計算で標準となっている計算手法であり、このハイブリッド汎関数がタンパク質全電子計算でも利用可能になったことは大変意義のあることである。

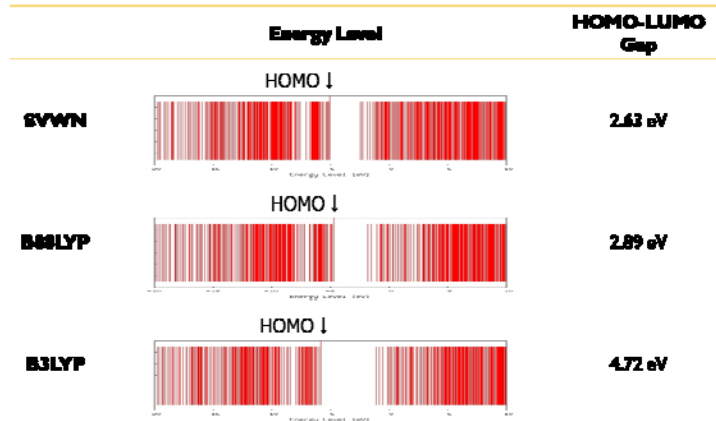


Fig. 4: SVWN, B88LYP, B3LYP 汎関数による RGD ペプチドの軌道エネルギーの比較

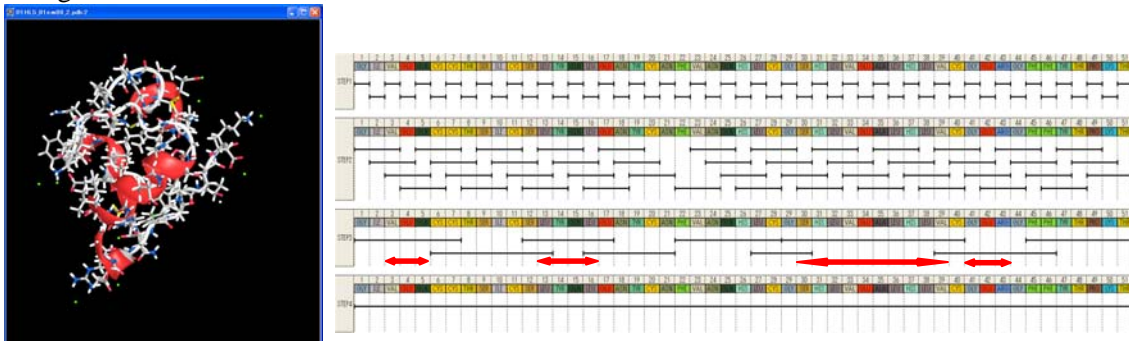


Fig. 5: インスリンの 3 次元構造と構造に基づく QCLO 法によるインスリンの全電子計算収束課程のシナリオ。矢印は α ヘリックス部位を表している。

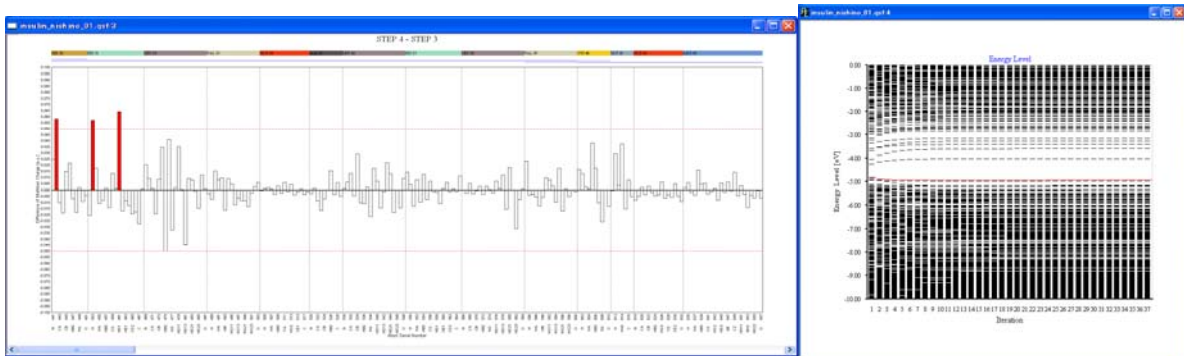


Fig. 6: (左)インスリン SER30-ARG43 部分における最終ステップの Mulliken 電荷初期値と収束値の差。左側強調している線は絶対値が 0.05 を超えている箇所。
(右)SCF 計算における軌道エネルギーの変遷。

d) ProteinDF システムを用いたタンパク質電子状態計算

最終版の ProteinDF システムを使用すれば、タンパク質の全電子計算が計算構造構築から計算制御、解析まで、統一的に達成することができる。一例として、インスリン単量体の計算の流れを示す。図は全て ProteinEditor で表示されたグラフィックスである。インスリン単量体は、4 つの α ヘリックスを持つ 2 本鎖のたんぱく質である(Fig. 5)。立体構造を考慮した QCLO 法を使用した場合のインスリン計算のシナリオを右に示した。ProteinEditor から、このシナリオを計算機サーバーに投入すれば、安全にインスリンの全電子計算が達成される。Fig. 6 には、最終ステップにおけるマリケン電荷の初期値と収束値の差のグラフを一部抜粋したものと SCF 計算における軌道エネルギーの遍歴を示した。最終ステップにおけるインスリン全体の計算では、ほぼ全ての原子におけるマリケン電荷の差は、

0.05 以下で、精度の高い初期値が作成でき、37 回の繰り返し計算でスムーズに収束した。ちなみに、この SCF 計算の収束法には、遅いが最も確実に計算できるダンピング法を採用した。 α ヘリックス部分で 0.05 以上のマリケン電荷差が見られる部分もあるが、それらは、 α ヘリックス部分以外との水素結合を形成しており、本ケースで1つのフレームにまとめて計算する水素結合の扱いからはずれたことが原因であることも、ProteinDF の連携機能を用いて解析できる。Fig. 7 に HOMO, LUMO の軌道と等電子密度面上の静電ポテンシャルの分布を示した。ProteinDF システムを利用すれば、このような形で、全電子計算の一連の流れを簡単に行なうことが可能である。

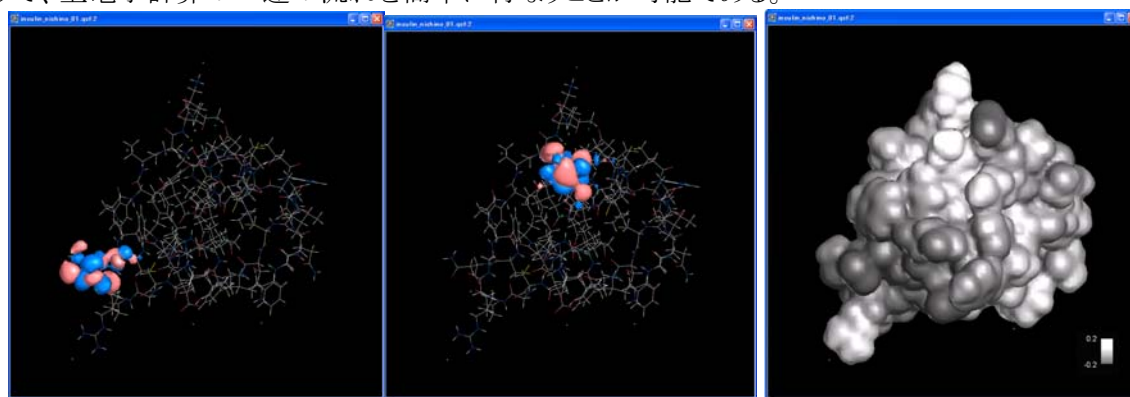


Fig. 7: インスリンの HOMO, LUMO, 静電ポテンシャルの 3 次元構造図。静電ポテンシャルは電子密度等値面(0.001)にマッピングした。

(6) 実証解析によるソフトウェアの有効性確認

a) ProteinDF の性能

RI 法に基づく局所密度汎関数法およびハイブリッド汎関数法の並列エンジンが完成した。律速となる計算ルーチンは、分子積分演算、交換相関数値積分、そして行列演算である。また、これらの計算ルーチンは行列要素の保持に大量のメモリ領域を要する。分子積分演算は、全ての行列要素を各 CPU に分散保存させながら、積分カットオフによる計算プレスクリーニングを組み合わせたダイレクト法によって、計算に要するワークエリアを圧縮して SPMD を達成するという独自のアルゴリズムで並列化した。分子積分演算で計算できる Fock 交換項以外の交換相関積分は、被積分関数が有理関数であるため数値計算で求める。これは原子中心の数値積分の和に変換して計算するため並列化は容易で、原子毎にタスクを分割すればよい。行列演算は ScaLAPACK の密行列用の関数を用いた。本ライブラリも行列要素を分散保持するため、全ての演算は 1CPU 毎に超巨大メモリを必要としないプログラム構造となった。局所密度汎関数法では、すでに 306 残基、原子数 4,728、電子数 18,552、基底関数の数 26,790 のインスリン 6 量体の全電子状態実証計算に成功している。これは世界最大のクラスタ密度汎関数法計算である。局所密度汎関数法よりも 1~2 桁精度が良くなるものの、計算量が断然多くなる B3LYP ハイブリッド汎関数法の計算時間をインスリン単量体実証計算で測定した。計算機は 2.8GHz Opteron 4 コアを使用した。その結果、局所密度汎関数法では 2 日、B3LYP では 1 週間で計算できた。一般化学分子で認められた超高品位の計算法が、廉価な PC クラスタで達成できることを実証したことは、汎用化への大変重要な成果となった。

b) インスリン MD シミュレーション

インスリンの 2 量体の MD シミュレーションを行った。インスリンは血糖値を低下させるホルモンタンパク質であり、様々なアナログ製剤が開発されている。これらはインスリン多量体から機能発現する単量体への解離速度が異なり、速効型や基礎代謝型といった用途に使用される。そこで、天然のヒト型インスリンとそのアナログであるリスプロ、アスパルト、グラルギンの 4 種類のインスリンそれぞれの 2 量体の MD シミュレーションを実行して、これら 2 量体のダイナミクスを解析した。この結果、2 量体間で構成される β シート部位の構造に安定性や構造変化の違いが解析された (Fig. 8)。特に、インスリン 2 量体が速く解離する新しい塩橋スイッチング機構を発見し、分子科学討論会優秀ポスター賞を受賞した。今後の超々速効型インスリン設計に重要な指針を与えるだろう。

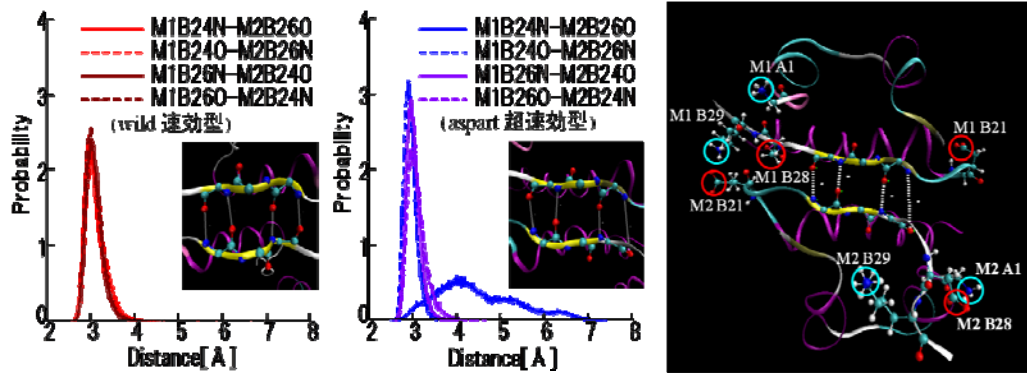


Fig. 8: インスリン単量体間の4つの水素結合

(左)天然、(中)aspartの水素結合間距離。(右)発見された aspartの動的塩橋スイッチング機構。

c) 新しいソフトウェア普及に対する著作活動

これまでも、シンポジウム、ワークショップやセミナー、国内外の雑誌・学会発表以外にも、書籍の執筆を通じて当グループのソフトウェアの普及に努めてきた。今年度は、最終版をリリースしたタイミングで、この新しいソフトウェアの使い方や新しい科学技術の教科書を執筆し、啓蒙に努めた。

3 成果要覧

招待講演／招待論文

- [招待 1] 佐藤文俊, 稲葉亨, 井原直樹, 恒川直樹, 西野典子, 西村康幸, 平野敏行, 吉廣保, 柏木浩: "ProteinDFシステムによるタンパク質電子構造解析", 第10回理論化学討論会, 2007.
- [招待 2] 佐藤文俊, 平野敏行: "ハイブリッド汎関数によるタンパク質全電子密度汎関数計算", CBI学会 2007年大会, 2007.
- [招待 3] 佐藤文俊: "タンパク質の量子シミュレーション:現状と将来", 秋田大学バイオサイエンス教育・研究センター「福祉医療工学セミナー」, 2008.
- [招待 4] 佐藤文俊, 平野敏行: "タンパク質密度汎関数計算:現状と将来展望", 2008 農芸化学シンポジウム, 2008.

受賞関連

- [受賞 1] 伊藤宏比古, 恒川直樹, 佐藤文俊: 第1回分子科学討論会優秀ポスター賞, 2007

著書／編集

- [著書 1] 佐藤文俊, 恒川直樹, 吉廣保, 平野敏行, 井原直樹, 柏木浩: 「タンパク質密度汎関数法」, 柏木浩, 佐藤文俊編, 森北出版, 東京, in press.
- [著書 2] 佐藤文俊他: 「生体分子量子化学計算 第1部」, 森北出版, 東京, in press.

査読付論文

- [査読付 1] 佐藤文俊, 稲葉亨: "タンパク質の全電子計算シミュレーションの発展", 化学工業, 58, pp.19-23, 2007.
- [査読付 2] 西村康幸, 吉廣保, 佐藤文俊: "タンパク質全電子計算のための GUI・ProteinEditor", 生産研究, 59, pp.93-100, 2007.

- [査読付 3] Noriko Nishino-Uemura, Toshiyuki Hirano, Fumitoshi Sato: “Study of the Quasi-Canonical Localized Orbital Method Based on Protein Structures”, *J. Chem. Phys.*, 127, pp.184106:1-10, 2007.
- [査読付 4] Toru Inaba, Fumitoshi Sato: “Development of Parallel Density Functional Program Using Distributed Matrix to Calculate All-Electron Canonical Wavefunction of Large Molecules”, *J. Comp. Chem.*, 28, pp.984-995, 2007.
- [査読付 5] Toru Inaba, Naoki Tsunekawa, Toshiyuki Hirano, Tamotsu Yoshihiro, Hiroshi Kashiwagi, Fumitoshi Sato: “Density functional calculation of the electronic structure on insulin hexamer”, *Chem. Phys. Lett.*, 434, pp.331-335, 2007.
- [査読付 6] 佐藤文俊: “生体分子量子シミュレーションと今後の展開”, 計算工学, 13, pp.1749-1752, 2008.

公開ソフトウェア

- [公開 1] 佐藤文俊他: ProteinDF システム,
<http://www.rss21.iis.u-tokyo.ac.jp/theme/life/reaction/index.html>

その他の発表論文

- [発表 1] 恒川直樹, 伊藤宏比古, 佐藤文俊: “分子動力学法によるヒトインスリン二量体の安定性の解析”, 日本物理学会 2007 年春季大会, 2007.
- [発表 2] 伊藤宏比古, 恒川直樹, 佐藤文俊: “ヒトインスリンのワイルドタイプとアスパルトのダイナミクス比較”, 日本物理学会 2007 年春季大会, 2007.
- [発表 3] 伊藤宏比古, 恒川直樹, 佐藤文俊: “ヒトインスリン二量体のダイナミクス解析”, 第 10 回理論化学討論会, 2007.
- [発表 4] 石川寛人, 西村康幸, 吉廣保, 佐藤文俊: “タンパク質構造変動に伴うポケット形状変化の動的表示”, 第 10 回理論化学討論会, 2007.
- [発表 5] 平岡克章, 井原直樹, 平野敏行, 佐藤文俊: “交換相関ポテンシャル計算高精度化に向けた単中心積分値計算法の研究”, 第 10 回理論化学討論会, 2007.
- [発表 6] Toshiyuki Hirano, Toru Inaba, Fumitoshi Sato: “Evaluation of Hybrid Functional for All-Electron Calculations on Proteins”, International Conference on the Applications of Density Functional Theory, 2007.
- [発表 7] Noriko Nishino, Toshiyuki Hirano, Fumitoshi Sato: “Improvement of the Quasi-Canonical Localized Orbital Method Based on Protein Structures”, International Conference on the Applications of Density Functional Theory, 2007.
- [発表 8] Naoki Ihara, Fumitoshi Sato: “The Fast Multipole Method in Quantum Chemistry : A Semi-Classical Approach”, International Conference on the Applications of Density Functional Theory, 2007.
- [発表 9] 井原直樹, 佐藤文俊: “Gauss 型基底関数を用いた量子化学計算における高速多重極展開法の改良”, 第 1 回分子科学討論会 2007, 2007.
- [発表 10] 平野敏行, 佐藤文俊: “ハイブリッド汎関数を用いたタンパク質全電子密度汎関数計算”, 第 1 回分子科学討論会 2007, 2007.
- [発表 11] 恒川直樹, 伊藤宏比古, 佐藤文俊: “ヒトインスリン二量体の安定性に対する水溶媒の効果”, 第 1 回分子科学討論会 2007, 2007.

- [発表 12] 伊藤宏比古, 恒川直樹, 佐藤文俊: “ヒトインスリンミュテーションによる二量体の安定性解析”, 第 1 回分子科学討論会 2007, 2007.
- [発表 13] 西村康幸, 吉廣保, 井原直樹, 佐藤文俊: “タンパク質量子化学計算のための分子モデリング環境の構築”, 第 1 回分子科学討論会 2007, 2007.
- [発表 14] 恒川直樹, 伊藤宏比古, 佐藤文俊: “ヒトインスリン二量体の安定性の自由エネルギー解析”, 日本物理学会年次大会・秋季(春季)大会, 2007.
- [発表 15] 伊藤宏比古, 恒川直樹, 佐藤文俊: “ヒトインスリンのミュテーションに対する立体構造の安定性変化の解析”, 日本物理学会年次大会・秋季(春季)大会, 2007.
- [発表 16] 井原直樹, 佐藤文俊: “A New Implementation of a Fast Multipole Method in Quantum Chemistry”, CBI 学会 2007 年大会, 2007.
- [発表 17] 石川寛人, 西村康幸, 吉廣保, 佐藤文俊: “Study on Dynamic Representation of Pocket Shape According to the Protein Structural Change”, CBI 学会 2007 年大会, 2007.
- [発表 18] 西村康幸, 吉廣保, 井原直樹, 佐藤文俊: “Development of the molecular modeling environment for quantum chemical calculations of protein”, CBI 学会 2007 年大会, 2007.
- [発表 19] 井原直樹, 佐藤文俊: “量子化学計算における高速多重極展開法: 半古典的アプローチ”, 第 21 回分子シミュレーション討論会, 2007.
- [発表 20] Nishino-Uemura Noriko, Fumitoshi Sato: “ProteinDF system for quantum chemical calculation of Proteins”, SC07, 2007.
- [発表 21] 佐藤文俊: “巨大タンパク質の励起状態・電子移動計算”, 次世代生命体統合分子スケール, 2007.
- [発表 22] 平野敏行, 佐藤文俊: “RI 法に基づくハイブリッド密度汎関数計算の超並列化法”, 次世代生命体統合分子スケール, 2007.
- [発表 23] 吉廣保, 西村康幸, 佐藤文俊: “タンパク質量子化学計算支援統合環境 ProteinEditor の開発”, 日本コンピュータ化学会 2007 秋季年会, 2007.
- [発表 24] Fumitoshi Sato, Toru Inaba, Noriko Nishino-Uemura, Yasuyuki Nishimura, Tamotsu Yoshihiro, Naoki Tsunekawa, Naoki Ihara, Toshiyuki Hirano: “Development of Quantum Chemical Simulation System for Proteins; ProteinDF”, GCOE: International Symposium on Secure-Life Electronics -Advanced Electronics for Quality Life and Society-, 2008.

特記事項

- [特記 1] 文部科学省: 「革新的シミュレーションソフトウェアの研究開発」および「次世代生命体統合シミュレーションソフトウェアの研究開発」に参画.

報道関連

- [報道 1] 日経 TRENDY: 「すぐその未来(6)そこにはない物に”触る”技術」, 2007 年 7 月 25 日.
- [報道 2] 日刊工業新聞: 「東大 生命科学などのシミュレーションソフト 26 本完成」, 2008 年 1 月 18 日.
- [報道 3] 科学新聞: 「革新的シミュレーションソフトウェア」, 2008 年 1 月 25 日.
- [報道 4] 化学工業日報: 「26 本の最終版を完成」, 2008 年 1 月 28 日.

4 倍精度演算ライブラリ的高速化および

疎行列ライブラリの自動チューニングの効果に関する研究

黒田 久泰

1 概要

コンピュータを使うあらゆる分野で高性能かつ高機能なアプリケーションが求められている。その中でも計算速度の高速化の要求は、コンピュータがいかに進歩しても留まるところを知らない。そこで、コンピュータを無駄なく最大限に利用するための高性能なアプリケーションを実現するための方法について研究している。2007年度に取り組んだ研究として主に下記の2点が挙げられる。

- 4倍精度演算ライブラリ的高速化

計算機内部の数値表現として一般に整数型と浮動小数点型が使われる。浮動小数点型の場合、単精度型あるいは倍精度型がほとんどであり、多くのプロセッサは倍精度演算をハードウェアでサポートしており高速に処理が行える。一方、多くのプロセッサは4倍精度演算についてはハードウェアでサポートしていないためソフトウェアで実装する必要がある。現在はコンピュータで処理する問題規模が大きくなり4倍精度の重要性も高まってきている。本研究では、高速な4倍精度演算ライブラリの構築を目的としている。

- 疎行列ライブラリの自動チューニングの効果に関する研究

大規模計算に安価なPCクラスタを用いるケースも最近では増えてきている。また、PCクラスタではOSがUNIXではなく、Windowsを使うというケースも徐々に増え始めてきている。そこで、これまで開発してきたILIB (Intelligent Library) の一部である一般化最小残差法 (GMRES法) による解法ライブラリをWindows CCSクラスタ上に移植し、その性能評価を行った。

業務報告としては、2007年度に新規に開始したものとして下記のものが挙げられる。

- スーパーコンピュータの民間開放に関する業務

文部科学省「先端研究施設供用イノベーション創出事業」【産業戦略利用】「先端的大規模計算シミュレーションプログラム利用サービス」の支援業務を行った。その中でも特にホームページの開設と保守管理業務を行った。

先に述べた2つの研究内容について以下、報告する。

2 4倍精度演算ライブラリの開発

2.1 背景

ハードウェアの進歩により、解くべき問題の規模が大きくなる傾向にある。数値計算解法の中には、計算規模が増すに従って計算誤差が増大し、より多くの計算量を必要としてしまうものもある。例えば、線形方程式 $Ax=b$ の解法であり、クリロフ部分空間を用いる CG 法などは誤差の影響を大きく受ける。また計算機上における浮動小数点演算では、実数を一定の有限桁で近似してしまうため丸め誤差が発生する。そこで、倍精度データ型よりも精度の高い 4 倍精度データ型の必要性が高まってきている。POWER 系アーキテクチャーにおいては、4 倍精度データ型は、倍精度浮動小数点数を 2 つ用いて表される。IEEE 規格の倍精度データ型の有効ビット数は 53 ビット(仮数部の 52 ビットにケチ表現で省いた 1 ビットを加えた値)であるため、POWER 系アーキテクチャーにおいては 106 ビットが有効ビット数となる。しかしながら、4 倍精度演算はハードウェアで実現されているわけではないため倍精度データ型に比べて効率は格段に落ちることが弱点となっている。

2.2 内容

4 倍精度演算の実現には、倍精度演算の誤差を考慮した演算アルゴリズムが基本となる。通常、浮動小数点の演算では丸め誤差が発生する。以下のアルゴリズムの説明において、演算 $\in \{+, -, \times\}$ に対して丸め誤差の発生する可能性がある浮動小数点演算をそれぞれ $\{\oplus, \ominus, \otimes\}$ を用いて記述することにする。また、2 つの浮動小数点数 a, b の加算は $a+b=\text{fl}(a+b)+\text{err}(a+b)$ と表すことができ、 $\text{fl}(a+b)$ は $a+b$ の浮動小数点演算、 $\text{err}(a+b)$ は $a+b$ の演算で生じる誤差を表す。

倍精度型の誤差付き加算アルゴリズム

Dekker による加算アルゴリズム $\text{Quick-TwoAdd}(a, b)$ は引数 a, b の大小関係に $|a| \geq |b|$ が成り立つ場合に適用できる。逆に、引数 a, b の大小関係がわかっていない場合には利用できない。また、下記の場合、 $a+b$ の結果は $s+e$ で表わされる。演算回数は 3Flop (Floating Operation) で実現できる。

```
Quick-TwoAdd(a, b){
  s ← a ⊕ b
  e ← b ⊖ (s ⊖ a)
  return (s, e)
}
```

Knuth の加算アルゴリズム $\text{TwoAdd}(a, b)$ は引数 a, b の大小関係に関係なく利用できる。通常はこちらを利用する。演算回数は 6Flop である。

```
TwoAdd(a, b){
  s ← a ⊕ b
  v ← s ⊖ a
  e ← (a ⊖ (s ⊖ v)) ⊕ (b ⊖ v)
  return (s, e)
}
```

倍精度型の誤差付き減算アルゴリズム

減算アルゴリズムも基本は加算アルゴリズムの場合と同様である。Quick-TwoSub(a, b) は $|a| \geq |b|$ が成り立つ場合に利用できる。TwoSub(a, b) は a, b の大小関係がわからない場合に利用できる。それぞれ $s = \text{fl}(a-b)$ と $e = \text{err}(a-b)$ を計算する。演算回数は加算の場合と同じである。

```
Quick-TwoSub( $a, b$ ){
   $s \leftarrow a \ominus b$ 
   $e \leftarrow (a \ominus s) \ominus b$ 
  return ( $s, e$ )
}
```

```
TwoSub( $a, b$ ){
   $s \leftarrow a \ominus b$ 
   $v \leftarrow s \ominus a$ 
   $e \leftarrow (a \ominus (s \ominus v)) \oplus (b \ominus v)$ 
  return ( $s, e$ )
}
```

倍精度型の誤差付き乗算アルゴリズム

乗算アルゴリズムでは、分割アルゴリズム Split(a)を利用する。Split(a) は $a \in \mathbb{R}$ を上位数 a_H と下位数 a_L に分割するアルゴリズムである。そして、 $a = a_H + a_L$ が成り立つ。ここで、 a を倍精度データ型とすると、Split(a)は次のようになる。演算回数は 4Flop である。

```
Split( $a$ ){
   $c \leftarrow (2^{27}+1) \oplus a$ 
   $d \leftarrow c \ominus a$ 
   $a_H \leftarrow c \ominus d$ 
   $a_L \leftarrow a \ominus a_H$ 
  return ( $a_H, a_L$ )
}
```

乗算アルゴリズム TwoMul(a, b) は Split を 2 回用いる。ここでは、 $s = \text{fl}(a \times b)$ と $e = \text{err}(a \times b)$ を計算する。演算回数は 17Flop となる。通常、乗算の場合、加算に比べて多くの計算を必要とする。

```
TwoMul( $a, b$ ){
   $s \leftarrow a \otimes b$ 
  ( $a_H, a_L$ )  $\leftarrow$  Split( $a$ )
  ( $b_H, b_L$ )  $\leftarrow$  Split( $b$ )
   $e_1 \leftarrow s \ominus a_H \otimes b_H$ 
   $e_2 \leftarrow e_1 \ominus a_L \otimes b_H$ 
   $e_3 \leftarrow e_2 \ominus a_H \otimes b_L$ 
   $e \leftarrow a_L \otimes b_L \ominus e_3$ 
  return ( $s, e$ )
}
```

ここまでは一般的な話であるが、IBM の POWER シリーズもしくは PowerPC シリーズには積和演算命令(Fused Multiply-Add Instruction)及び積差演算命令(Fused Multiply-Subtract Instruction)が備わっている。そして、この積和(積差)演算命令の特徴として、途中の乗算結果が 53 ビットに丸められず 106 ビットの精度を保ったまま加算あるいは減算が行われるので、誤差の発生が低く抑えられる。この特徴を活かした乗算アルゴリズム TwoMul_FMA(a, b) は次のようになる。ここで、 $\text{fl}(a \times b - p)$ の部分が積差演算である。このアルゴリズムでは、演算回数はたった 3Flop で済む。

```
TwoMul_FMA( $a, b$ ){
   $p \leftarrow a \otimes b$ 
   $e \leftarrow \text{fl}(a \times b - p)$ 
  return ( $p, e$ )
}
```

4 倍精度型の加算アルゴリズム

4 倍精度型の加算アルゴリズム Quad_TwoAdd(a, b) は $(s_H, s_L) = \text{fl}(a+b)$ を計算する。これまでの話とは異なり、ここでの a, b, s は 4 倍精度の数である。また、 (s_H, s_L) は $s = s_H + s_L$ を表している。

```
Quad_TwoAdd( $a, b$ ){
  ( $t, r$ )  $\leftarrow$  TwoSum( $a_H, b_H$ )
   $e \leftarrow r \oplus a_L \oplus b_L$ 
   $s_H \leftarrow t \oplus e$ 
   $s_L \leftarrow t \ominus s_H \oplus e$ 
  return ( $s_H, s_L$ )
}
```

演算回数は TwoSum の 6Flop と残りの 5Flop を加えて 11Flop となる。ただし、このアルゴリズムでは、上位数部分で発生した誤差を下位数部分に加算するとき発生する誤差を考慮していないため 1 ビット程度の誤差が発生するケースがある。この誤差を無くす場合には、より多くの計算が必要となる。

4 倍精度型の乗算アルゴリズム

4 倍精度型の乗算アルゴリズム Quad_TwoMul(a, b) は $(p_H, p_L) = \text{fl}(a \times b)$ を計算する。ただし、 a, b, p は 4 倍精度の数である、また、 (p_H, p_L) は $p = p_H + p_L$ を表している。

```
Quad_TwoMul( $a, b$ ){
  ( $t, r$ )  $\leftarrow$  TwoMul_FMA( $a_H, b_H$ )
   $v \leftarrow a_H \otimes b_L$ 
   $w \leftarrow a_L \otimes b_L$ 
   $r \leftarrow r \oplus v \oplus w$ 
   $p_H \leftarrow t \oplus r$ 
   $p_L \leftarrow t \ominus p_H \oplus r$ 
  return ( $p_H, p_L$ )
}
```

演算回数は TwoMul_FMA の 3Flop と残りの 7Flop を加えて 10Flop となる。

ところで、4 倍精度型の乗算アルゴリズムでは積和及び積差演算を用いた別の実装方法も考えられる。次の乗算アルゴリズム `Quad_TwoMul_Modified(a, b)` では、演算回数は 8Flop で実現できる。

```
Quad_TwoMul_Modified(a, b){
  m1 ← aH ⊗ bL
  m2 ← aL ⊗ bH
  t ← m1 ⊕ m2
  pH ← fl(aH × bH + t)
  e ← fl(aH × bH - pH)
  pL ← e ⊕ t
  return (pH, pL)
}
```

`Quad_TwoMul_Modified(a, b)` の最初の 3 行で $t \leftarrow (a_H \otimes b_L) \oplus (a_L \otimes b_H)$ を計算している。ここでも積和演算を用いると若干効率が良くなるが、引数 a と b の順番が入れ替わった場合に結果が変わってくるがあるので注意が必要である。

4 倍精度の演算は単独で使われることはほとんどなく、通常はまとまった計算で利用される。POWER5 シリーズの倍精度浮動小数点演算のレイテンシーは 6 クロックと比較的大きな値である。そのため 4 倍精度の演算を 1 回行うような場合には大変効率が悪いが、複数個の演算をまとめて行う場合には、レイテンシーが隠蔽され、性能が向上することが期待できる。直感的には 6×2 (2 は演算器の数) 個の 4 倍精度演算をまとめて行えばいいことになるが、実際には物理レジスタや論理レジスタの数が限られていることを考慮しなくてはならない。例えば、多くの 4 倍精度演算をまとめて行うようにすると一時的なメモリ退避などが発生してしまう。本研究では、これまでに述べたアルゴリズムを用いて、4 倍精度演算の高速化を実現している。

2.3 具体的成果

実験環境として東京大学情報基盤センターにある SR11000/J2 を用いた。各ノードは 16 個の CPU から構成されている。CPU は POWER 5+ 2.3GHz が使われており、1CPU の理論ピーク性能は 9.2GFlops (加算あるいは乗算だけの場合は 4.6GFlops) である。本実験は、1CPU のみを用いて行った。

ここで、1 回の 4 倍精度数同士の演算を 1QFlop (Quadruple Floating Point Operation) と呼ぶことにする。また、1 秒間に計算した 4 倍精度数同士の演算回数を QFlops と表記することにする。

4 倍精度同士の加算

2 つの 4 倍精度数の配列の同じインデックス同士の和を別の 4 倍精度数の配列に格納するプログラムを用いて評価した。結果を表 1 に示す。Datasize 欄の L1 は L1 キャッシュに入力配列及び出力配列の全てが入るサイズ、L2(half) は L2 キャッシュの半分に入るサイズ、L2(full) は L2 キャッシュ全体に入るサイズ、L3(half) は L3 キャッシュの半分に入るサイズ、L3(full) は L3 キャッシュ全体に入るサイズ、OUT は L3 キャッシュサイズには入りきれない大きなサイズを表す。Unrolling Size 欄はループのアンローリング段数で丁度この数に等しい 4 倍精度演算をまとめて計算している。表 1 からキャッシュにデータが収まっているかどうかはあまり影響していないことがわかる。それは、1 回

の 4 倍精度演算の際、メモリ転送に関して 8 バイト×4=32 バイトのメモリからの読み込みと 16 バイトのメモリへの書き込みが発生するが、それらのメモリへの負荷よりも 4 倍精度演算の負荷の方が大きいからである。8 個の 4 倍精度数の和を同時に行うとき、L1 キャッシュに収まるデータサイズで 257.80MQFlops となった。1CPU の倍精度型の加算の場合の理論ピーク性能は 4.6GFlops であるため 4 倍精度数の加算の理論ピーク性能は、418MQFlops であり、その 62%を達成した。

表 1: 4 倍精度数の加算における MQFlops 値

Datasize	Unrolling Size				Hitachi Compiler	gcc Compiler
	1	2	4	8		
L1	102.01	139.17	190.98	257.80	247.63	20.41
L2(half)	104.44	139.33	183.22	256.17	245.17	20.40
L2(full)	104.37	139.42	182.53	255.45	246.23	20.89
L3(half)	95.91	124.52	154.56	216.74	208.78	20.65
L3(full)	95.88	124.61	154.38	220.09	210.45	21.07
OUT	104.28	138.69	182.08	255.56	245.94	20.90

4 倍精度同士の乗算

2 つの 4 倍精度数の配列の同じインデックス同士の積の値を別の 4 倍精度数の配列に格納するプログラムを用いて評価した。結果を表 2 に示す。8 個の 4 倍精度数の積を同時に行うとき、L1 キャッシュの収まるデータサイズで 357.71MQFlops となった。1CPU の倍精度型の積の場合の理論ピーク性能は 4.6GFlops であるため 4 倍精度数の乗算の理論ピーク性能は、575MQFlops であり、加算の場合と同じくその 62%を達成した。

表 2: 4 倍精度の乗算における MQFlops 値

Datasize	Unrolling Size				Hitachi Compiler	gcc Compiler
	1	2	4	8		
L1	171.48	261.75	278.83	357.71	247.86	20.67
L2(half)	171.47	257.52	288.12	309.25	247.09	20.73
L2(full)	171.78	257.40	288.24	310.02	244.55	21.11
L3(half)	160.80	202.87	225.94	264.62	232.92	20.95
L3(full)	160.91	214.64	231.21	263.91	232.12	20.91
OUT	171.62	256.61	286.33	303.59	246.04	21.07

3 疎行列ライブラリの自動チューニングの効果に関する研究

3.1 背景

東京大学情報基盤センターのような共同利用機関のスーパーコンピュータは多くのユーザーが利用している。そのため、ログインしたノード上で対話的に並列プログラムの実行ができなかったり、並列プログラムが実行されるまでに長時間待たなければならなかったりする。そこで、通常はプログラム開発のために自前の並列プログラム開発環境を構築するといったことが必要である。最近では、Windows CCS (Windows Compute Cluster Server) を使って PC クラスタを構築することが主流となってきている。UNIX のコマンドを覚えなくてもよいこと、普段使っている Windows パソコンそれ自身も計算ノードとして利用できること、古くなったパソコンなども計算ノードとして利用できることが利点として挙げられる。

同じ並列計算機環境といっても、スーパーコンピュータと Windows クラスタではプロセッサの特性、メモリバンド幅、ノード間通信性能など大きな違いがある。そのため同じプログラムであっても性能が大きく異なってくることもある。そこで今回はこれまで開発を行ってきた ILIB (Intelligent Library) の一部である一般化最小残差法 (GMRES 法) を Windows CCS 向けに移植を行い、性能評価を行った。

3.2 内容

高性能な数値計算ライブラリを開発する上で重要なことは、プログラム実行環境に適した計算手法や通信手法を利用することである。そのため並列計算機の通信性能やプロセッサの性能特性を把握しておくとは特に重要となる。

Windows CCS の MPI 実装としてマイクロソフトの開発した MS-MPI がある。TCP/IP によるソケット通信と MS-MPI による 1 対 1 ノード間通信性能は図 1 のようになった。各ノードには CPU として Dual Core Xeon 3040 1.86GHz、ネットワークとして HP NC320i PCIe Gigabit と Intel PRO/1000GT の 2 つの 1Gbps の LAN ボードが搭載されており、Intel PRO/1000GT を MS-MPI 専用で使う設定とした。また、ノード間にはスイッチング HUB を設置している。図 1 からまとめて 32KB 以上のデータを送ると 50MB/s 以上の性能になっていることがわかる。

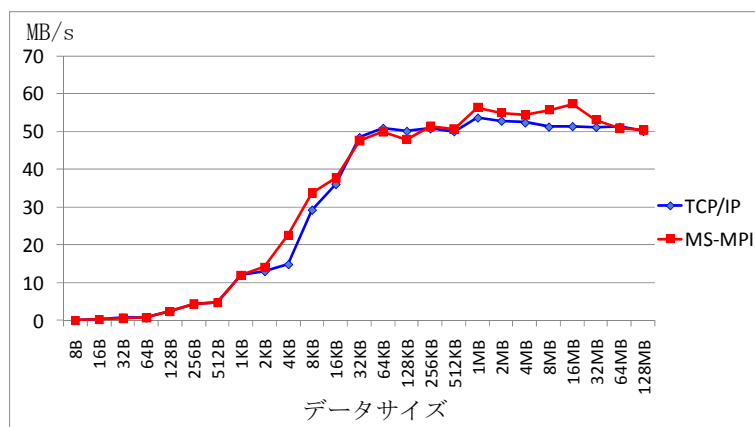


図 1: 1 対 1 ノード間のデータ転送性能

PC クラスタのような環境では、各ノードに搭載されている CPU の周波数が違っていたり CPU の数も異なっていたりすることも多い。そのため、OpenMP やコンパイラの自動並列化を利用せずに、MPI のみを用いて並列化を実現することがよくある。図 1 の測定と同じ環境で、1 ノード内に 2 つの MPI プロセスを立ち上げ、MS-MPI 及び TCP/IP を用いて送受信を行った場合は図 2 のようになった。ここでは、各プロセスが実行される CPU コアを SetProcessAffinityMask 関数で固定している。図 2 では、同じサイズに対し複数回測定しその平均を取ったため、送受信メッセージサイズの合計が L2 キャッシュ (2CPU コア共有 2MB) あたりになったところで通信性能が落ちていることがわかる。

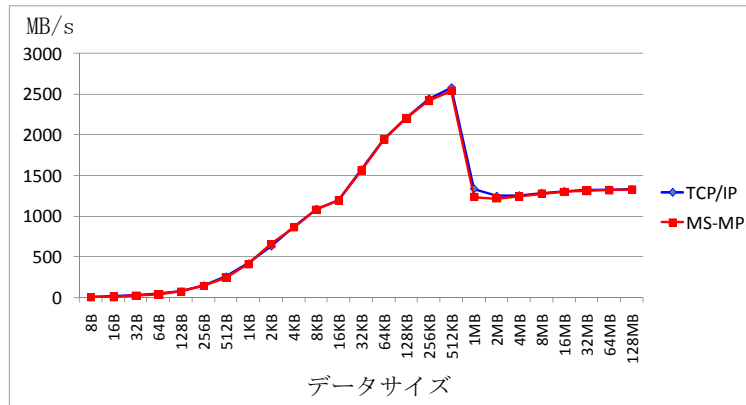


図 2: ノード内 2 プロセス間のデータ転送性能

MPI 集団通信関数は基本的には 1 対 1 ノード間通信の組み合わせで実装されている。しかしながら、MPI 集団通信関数の通信性能は実装方法に大きく依存する。例えば、ノード数が 2 のべき乗の形となっていれば通信の効率が良くなるが、そうでない場合、こういった順番で通信するかによって性能が大きく変わってくる。また、疎行列を対象とするような場合には、ノード間の計算負荷にもバラつきが生じるため通信開始のタイミングが全てのノードで一致するわけではない。さらに、通信量についても 2 つのノードの組毎に異なってくる。そのためライブラリ実行時に最適な通信方式を見つけるといったことが重要となってくる。

GMRES(m)法については、現在、下記の項目が自動チューニングの対象となっている。

1. 行列格納形式
2. ループプアンローリング方法
3. 行列ベクトル積における通信方式
4. リスタート周期
5. ベクトル直交化アルゴリズム
6. 前処理方式

上記の 1、2、3 は主に疎行列の非零要素の分布と CPU の特性によって決定される。4 は疎行列の特性及び他の項目で何が選択されたかに大きく依存する。5 は実行ノード数や通信性能に大きく依存する。6 は実行ノード数、通信性能、疎行列の特性に大きく依存する。これら 6 項目はそれぞれ独立しているわけではなく、相互に関係しているので、ライブラリ実行時において最適な選択を行うのは難しい問題である。現状では、ある程度の制約を置いた上で最適と思われる方式を選択している。

3.3 具体的成果

典型的な例として、SR11000/J2 の 2 ノード 32CPU の内 30CPU を使った実行時間の結果を図 3、Windows CCS クラスタを用いた場合の結果を図 4 に示す。ここで、「AT Off」ではソースプログラム上はループアンローリング無し(コンパイラの最適化は有効)、通信方式としては MPI_Allgather を用いた単純な実装、リスタート周期は 30 で固定、直交化アルゴリズムは IR-CGS (古典的グラムシュミット直交化法を 2 回適用)、前処理方式はスケーリング前処理 (対角成分を 1 に変換する方法) のみを適用した。「Only MPI On」では、「AT Off」に対して、通信方式だけを自動的に最適な方式を選択させた場合である。「All On」は全ての項目について自動チューニングを適用させた。

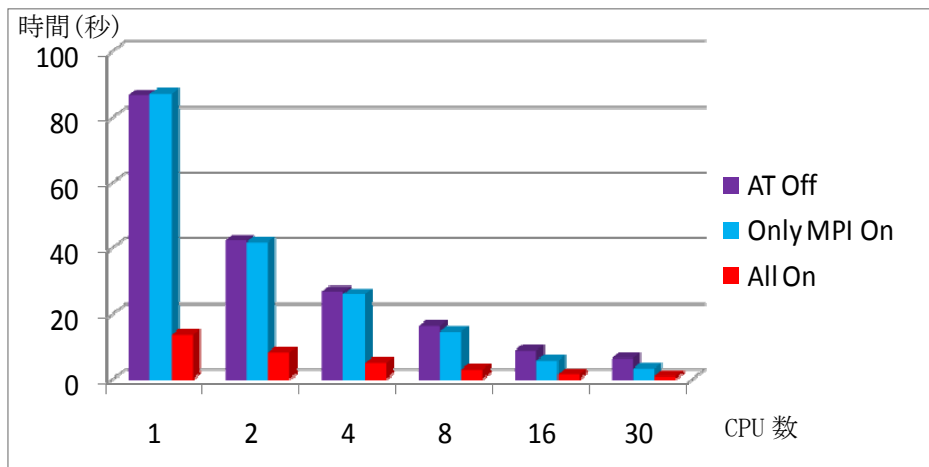


図 3: 実行時間 (SR11000/J2)

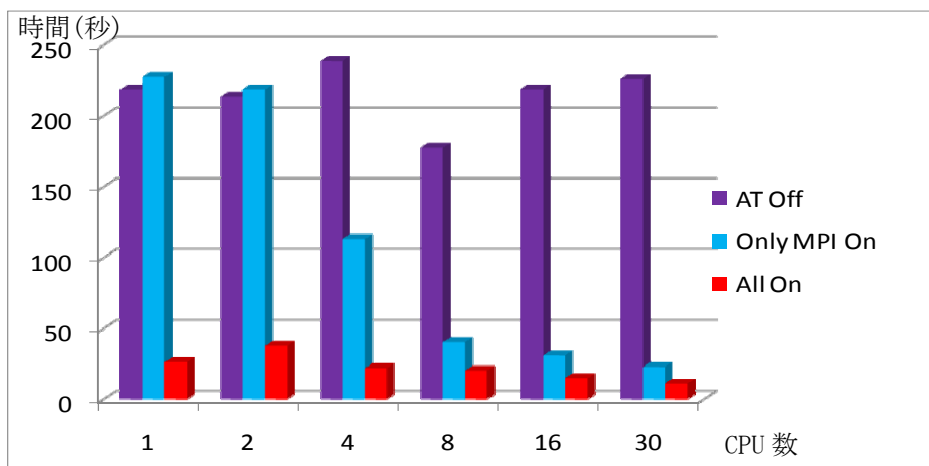


図 4: 実行時間 (Windows CCS クラスタ)

Windows CCS クラスタでは通信に関する自動チューニングの適用は非常に効果が出ている。これは単純に、通信に要する時間の割合が大きかったためにその部分を最適化することで大幅な速度向上となった。また、Windows CCS クラスタで 15 ノード(30CPU)使用時に自動チューニングを適用して最適な方法を選択させることで適用しなかった場合に比べて 21.1 倍の速度向上となった。

4 成果要覧

査読付論文リスト

- [査読付 1] Ryosuke Yano, Kojiro Suzuki, and Hisayasu Kuroda: Numerical analysis of relativistic shock layer problem by using relativistic Boltzmann-kinetic equations, *Physica A: Statistical Mechanics and its Applications*, Vol.381, pp.8-21, 2007.
- [査読付 2] 永井貴博, 吉田仁, 黒田久泰, 金田康正: SR11000 モデル J2 における 4 倍精度積和演算の高速化, *情報処理学会論文誌: コンピューティングシステム*, Vol.48, No.SIG 13(ACS 19), pp.214-222. 2007.

国際会議論文リスト(査読付)

- [国際会議 1] Yusuke Hiromatsu and Hisayasu Kuroda: Vectorizing Method for Complex Control Flow with SIMD Instruction Set, *The 2007 International Conference on Software Engineering Research & Practice (SERP'07) in WORLDCOMP'07*, pp.203-209, 2007.
- [国際会議 2] Hitoshi Yoshida, Hisayasu Kuroda, and Yasumasa Kanada: Parameter Selection for Fast Computation of Pi(x), *The 2007 International Conference on Foundations of Computer Science (FCS'07) in WORLDCOMP'07*, pp.177-182, 2007.
- [国際会議 3] Takahiro Nagai, Hitoshi Yoshida, Hisayasu Kuroda, and Yasumasa Kanada: Fast Quadruple Precision Arithmetic for Multiply/Add Operations on SR11000/J2, *The 2007 International Conference on Scientific Computing (CSC'07) in WORLDCOMP'07*, pp.151-157, 2007.

その他の発表論文リスト

- [発表 1] Hisayasu Kuroda and Takahiro Katagiri: Auto-tuning Effect of Iterative Method Library on Windows CCS, *13th SIAM Conference on Parallel Processing for Scientific Computing (PP08), Mini Symposium, "MS25 Auto-tuning on Numerical Libraries and Advanced Computer Systems - Part II of II"*, Atlanta, Georgia, Mar. 12-14, 2008.
- [発表 2] 黒田久泰, 片桐孝洋: Windows CCS 上における MS-MPI の実行時自動チューニング, *2008 年ハイパフォーマンスコンピューティングと計算科学シンポジウム(HPCS2008)*, 情報処理学会シンポジウムシリーズ, Vol.2008, No.2, p.54, 2008.

特記事項

- [特記 1] 黒田久泰: C 言語による OpenMP 入門, *東京大学情報基盤センタースーパーコンピューティングニュース特集号*, Vol.9, No.Special Issue 1, pp.149-168, 2008.
- [特記 2] 片桐孝洋, 黒田久泰: HPC サーバーにおけるソフトウェア自動チューニングのインパクト: MS-MPI の実行時自動チューニング機構の開発を例にして, *Windows HPC アップデートセミナー*, 2008 年 2 月.
- [特記 3] 黒田久泰, 片桐孝洋: 疎行列ソルバにおける MS-MPI の実行時自動チューニングの効果, *Windows HPC アップデートセミナー*, 2008 年 2 月.

スーパーコンピュータのための高速通信機構

松葉 浩也

1 概要

2007年度はクラスタ間 MPI 通信を高速化するための新しい手法について研究を行った。本研究で提案した手法では複数のクラスタを同時に使用して並列計算を行うことができ、従来の手法に比べて最大 3.3 倍の性能を得た。また、スーパーコンピュータサービスの品質向上への貢献として、高速ファイル通信ソフトを作成して公開した。

2 クラスタ間 MPI 通信機構の研究

2.1 背景

近年、PC クラスタは並列計算機の形態として一般的なものとなっている。通信を伴う並列計算に PC クラスタを使用する場合、一般的には通常のサーバが備えるネットワークでは性能が不十分であるため、Infiniband, Myrinet に代表される高性能インターコネクが使用される。これらのインターコネクは高性能である分、導入コストも高い。特に一台のスイッチに収容できないほど多数のノードを結合する場合、導入コストのみならず物理的配線の量も膨大となり故障率の増加を招く。

本センターのように多数のユーザーがスーパーコンピュータを使用する環境では、スーパーコンピュータ全体を単一のアプリケーションが占有することはほぼ皆無であり、通常はスーパーコンピュータをいくつかの部分に分割し、独立したアプリケーションを同時に実行する。したがって、スーパーコンピュータの導入にあたっては、全体をインターコネクで高速に接続する必要はなく、通常の運用での最大ノード数を高速に接続しておけば十分である。実際、2008年に導入を開始するオープンスーパーコンピュータでは 952 台の計算ノードが 512 台、256 台、128 台、56 台の 4 つに分割されて接続されている。一方で、通常の運用とは別に特別に大規模な問題を解く場合のために、高速接続されていないクラスタ同士が連携して並列計算を行う仕組みも準備しておく必要がある。本報告ではこのような場合のために、複数のクラスタを使用して並列計算を行うための効率的な通信ライブラリについて研究した成果について報告する。

複数のクラスタを接続して通信を行う必要性は本センターに限った課題ではなく、複数の計算センターを接続して計算を行う「計算グリッド」の試みにおいても重要である。そのため、主にグリッドの研究としてクラスタ間通信を高速に行う手法が提案されてきており、産業技術総合研究所などが開発した GridMPI などが代表的である。これらグリッド向けの手法を本センターにおけるクラスタ間接続方式として使用することも可能ではあるが、次のような性能上の問題がある。

TCP に関わる問題

クラスタ間を接続するネットワークは低速であるため、多数のノードが一斉にクラスタ間ネットワークにデータを送信するとすべてのデータが通信できず、データの消失が発生する。データの

消失に対しては通常 TCP 通信プロトコルがソフトウェア的に再送処理を行うため通信結果には影響しないものの、性能は低下する。

ネットワークインタフェースに関わる問題

クラスタ内は 10Gbps のインターコネクトを複数本使用する高速な通信が可能である一方で、クラスタ間通信に使えるネットワークは通常のサーバが備える 1Gbps のネットワークのみである。クラスタ同士を接続するネットワークスイッチが 10Gbps 以上の容量を持っていたとしても、その帯域を生かすには 10 台以上が同時に通信を行う必要がある。多数のノードを持つクラスタ同士を接続する場合、10 台程度ならば常に通信を行っていると思われるが、実際、並列計算は計算と通信が短い周期で繰り返されるため、複数のノードが厳密に同じ時間に通信を行う可能性は低い。そして、通信を行っているノードが少ない時間帯は、各ノードのネットワークインタフェースがボトルネックとなり、クラスタ間接続の帯域を十分に生かすことができない。

本研究の具体的な目標はこの 2 点の問題を解決する通信方式を考案することである。また、将来的に本センターにおいて複数のクラスタを使った並列計算を行う必要が発生した場合にはすぐにでも使用できるよう、実用的なレベルにまで安定した実装を行うことも目標とする。

2.2 内容

提案手法の概略を図 1 に示す。クラスタ間には特殊なネットワーク接続が存在しないため、クラスタ間通信のために TCP プロトコルを使用すること自体は避けられないが、TCP 通信を行う計算機の数制限することにより、通信の混雑が起きないように使い方をする。低速なネットワークインタフェースしか使用できない問題に対しては単一のメッセージを分割して複数のネットワークインタフェースを使用して通信することで解決する。この手法を aggregate router と呼ぶ。

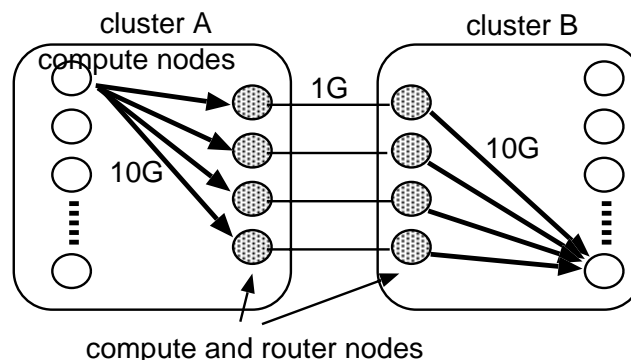


図 1: Aggregate Router Method

提案手法では通信は次のように行われる。

1. 計算プロセスがクラスタをまたぐ送信要求を出す
2. メッセージはクラスタ内の高速ネットワークを用いてルータープロセスと呼ばれる外部との接続を持ったプロセスに転送される。このときメッセージは複数に分割され、複数のルータープロセスに転送される
3. それぞれのルータープロセスは送信先クラスタのルータープロセスにデータを転送する
4. 送信先のルータープロセスがクラスタ内高速ネットワークを用いて実際の宛先プロセスにデータを送信する。宛先プロセスは異なるルータープロセスを経由したメッセージを組み立て、送信元のデータを復元する

図1に示された構成の場合、ルータープロセスは各クラスタに4プロセス準備されている。ルータープロセス同士は対一でTCPの接続を確立する。ルータープロセスの対外接続はそれぞれ1Gbpsであるため、クラスタ間接続が4Gbps以上あればデータ混雑によるパケットロスとは決して発生しない。複数の計算プロセスが一個のルータープロセスにデータを送信した場合はデータの混雑が発生するが、クラスタ専用高速ネットワークは混雑への対応にも優れているため、TCPプロトコルのような性能低下は発生しない。また、ルータープロセスの対外接続がクラスタ間ネットワークより低速な場合においても、本手法では単一のメッセージを複数のルータープロセスに分割することで常にクラスタ間ネットワークの有効活用を可能としている。

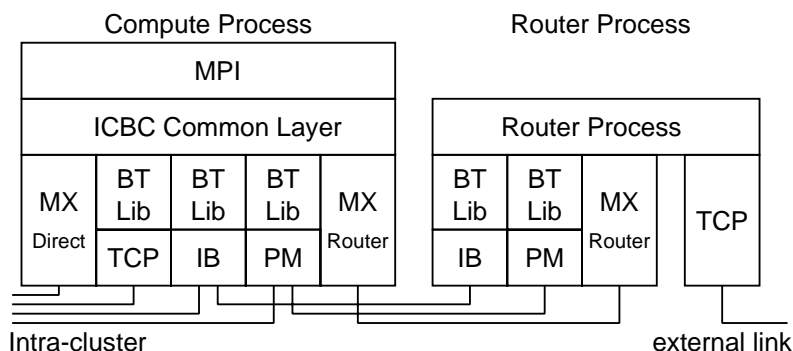


図 2: ICBC Protocol Stack

提案手法を MPI 通信ライブラリに組み込むため、図2のようなプロトコルスタックを設計した。上で述べたように、計算プロセスとルータープロセスは Myrinet あるいは Infiniband といった高速ネットワークで接続される。そのためには、図中に MX, IB, PM と示されている高速ネットワーク用のドライバが使用される。計算プロセスにおいてはデバイス依存部分の上位にデバイス共通のプロトコル処理層がある(図中の ICBC Common Layer)。ここでは MPI 通信に規定されている送受信順を守るための処理が実装されている。その上位にユーザープログラムへのインタフェースである MPI ライブラリが実装されている。一方、メッセージの送信順序はすべて受け取り側で管理されるため、ルータープロセスでは MPI の送受信順を守る必要はない。そのためルータープロセスにおいてはメッセージ順序を管理するプロトコル層は省略されており、デバイス依存部分から受け取ったデータは直ちに他方のルーターに転送される。

実装は本センターの石川教授が中心となって開発した YAMPI 通信ライブラリの低レベル通信層を置き換える方式をとった。この新しく開発した YAMPI は MPI 1.2 規格のすべてと MPI 2 規格の一部が実装されており、MPI 規格の開発元である米国アルゴンヌ国立研究所が提供する MPI のテストプログラムがすべて動作している。また、本研究で提案したクラスタ間通信方式のみでなく、通常構成のクラスタでも使用でき、通信方式が異なるクラスタに移動した際にも再コンパイルすることなくバイナリがそのまま動作するように実装されている。本研究で実装した部分はライブラリ自身が使用するメモリのキャッシュ効率まで考慮されており、通信時にキャッシュから追い出されるアプリケーションデータがなるべく少なくなるように実装されている。これらの工夫により本ライブラリはクラスタ間通信を伴わない通常の MPI 通信ライブラリとしても既存の実装を上回る性能を得ており、実用に耐える実装であると言える。ただし、本稿で報告する研究とは直接関係しないため、一般的な MPI 実装としての性能評価は割愛する。実装したライブラリは YAMPI3 として公開している [公開 1]。

2.3 具体的成果

提案手法を図3の環境で評価した。クラスタ A,B はそれぞれ 16 台のクラスタであり各ノードは Dul-Core Opteron プロセッサ (2GHz) を 2 個 (合計 4 コア) のサーバコンピュータである。それぞれのクラ

スタ内は Myrinet-10G インターコネクトで接続されている。クラスタ間接続には 10Gbps の Ethernet を使用する。ただし 10Gbps であるのはスイッチ間の接続のみであり、各ノードからノード間ネットワークのスイッチには 1Gbps のネットワークで接続される。提案手法の評価ではそれぞれのクラスタ内で 8 個のルータープロセスを動作させる。また、ルータープロセスが動作するノードでも他のノードと同じ数の計算プロセスを実行する。

この後示す性能のグラフでは提案手法である aggregate router を AR と表記する。比較対象の既存手法として、クラスタ間通信を行う各プロセスが独立した TCP 接続を持つ All-to-All と呼ばれる方法を AA、提案手法と同じく複数のルーターを用いるがそれぞれのノードが使用するルーターが固定されており単一のメッセージを分割して複数のルーターで転送することができない Load-Balanced router と呼ばれる方法を LB と表記する。また、クラスタ間通信が不要な 32 ノードのクラスタとして使用した時の性能と比較するため、クラスタ A,B のすべてのノードを高速インターコネクトで接続する方法 (MX)、クラスタ A,B のすべてのノードを TCP で接続する方法 (TCP) とも比較する。

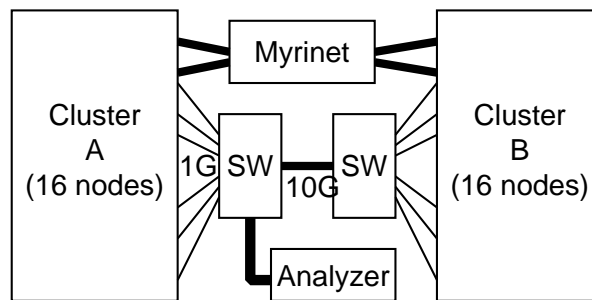


図 3: Cluster Configuration

ネットワーク性能

一対一のネットワークバンド幅測定の結果を図 4 に示す。クラスタ間通信の既存手法である AA と LB は共に 116MB/s にとどまっている。これはそれぞれのノードが 1Gbps のネットワークインタフェースしか持たないためである。一方、提案手法では単一のメッセージを複数ルーターで通信できるため、一対一の通信であっても単一のルーターのバンド幅以上の帯域が得られる。グラフからわかるように 710MB/s が得られており、既存手法の 6.1 倍の性能である。

MPIFFT

HPC Challenge Benchmarks 中の MPIFFT を用いて性能評価を行った結果を図 5 に示す。FFT はネットワーク性能が重要なプログラムであり、クラスタ間通信を行っている AR, LB, AA の中では提案手法である AR が最も高い性能を示している。これがネットワーク性能の差によるものであることを示すため、クラスタ間ネットワークに流れるデータのバンド幅を測定した。測定はベンチマーク中のある 25 秒間について行い、10ms 間隔で平均のバンド幅を計算した。結果を図 6 に示す。

(A) の提案手法では通信が始まるとバンド幅は直ちに 8Gbps 近い値となっている。ルータープロセスが 8 個なので、これは理論値に近い値である。一方、同じ 8 個のルーターを使う Load-Balanced router(B) では最高時は 8Gbps 近いバンド幅となっているが、通信するプロセスの数が少ない時間帯はそれに満たないバンド幅しか得られていない。さらに全プロセスが個別に TCP 接続を張る方法 (C) ではパケットロスなどのために通信がまったく安定しないことがわかる。この通信性能の差が図 5 に示される性能差である。

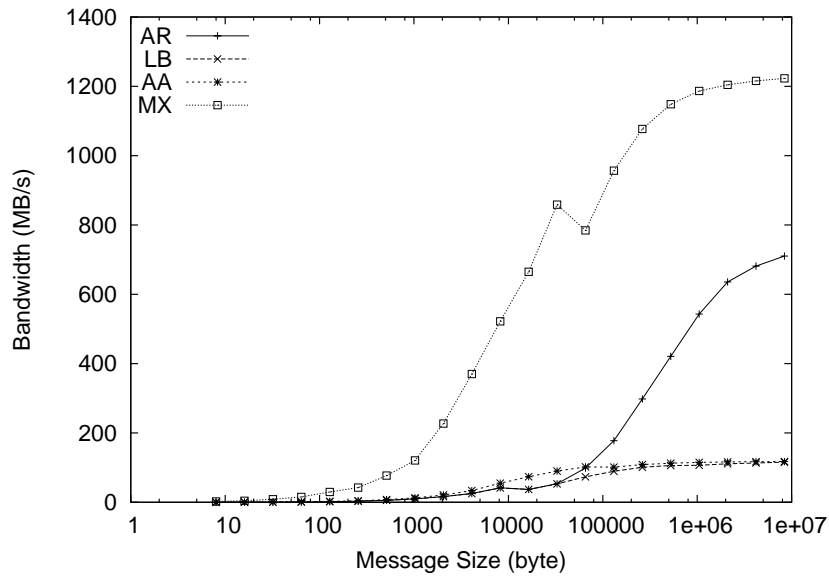


図 4: Bandwidth

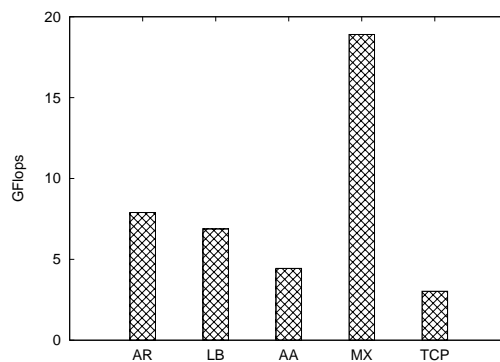


図 5: MPIFFT Results

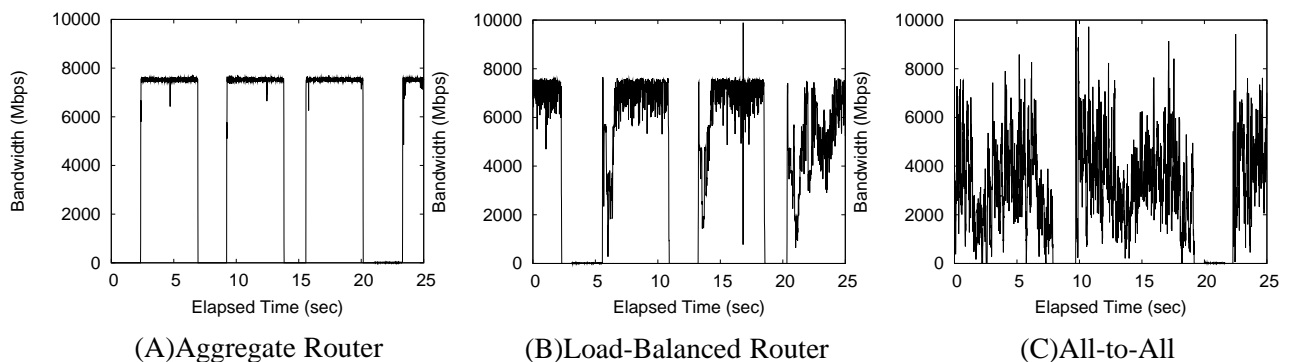


図 6: Traffic Sampling of MPIFFT

HPL

High Performance Linpack の結果を図 7 に示す。このベンチマークはほとんどが計算時間であり通信性能は全体の性能にあまり影響しない。それでもクラスタ間通信を行う手法の中では提案手法 (AR) が最も高性能である。MPIFFT と同じようにクラスタ間ネットワークにおけるバンド幅測定の結果を

図 8 に示す。MPIFFT と同じ傾向が出ており、提案手法が最も高いバンド幅を得ている。

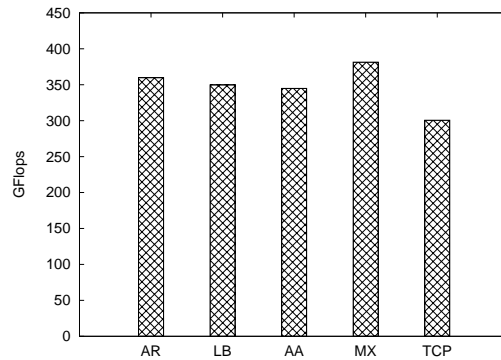


図 7: HPL Results

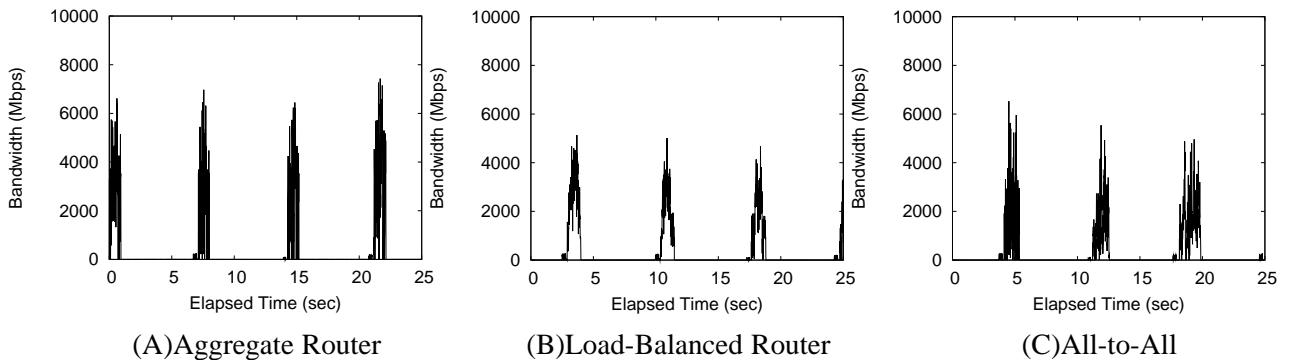


図 8: Traffic Sampling of HPL

CG

NAS Parallel Benchmarks から CG(Class C)の結果を図 9、11 に示す。CG は短い周期で通信と計算が入れ替わるベンチマークであり、図 11 のトラフィックサンプリングは MPIFFT や HPL とは違い、一秒間の通信状況が示されている。提案手法が最も有効なのはこのような通信パターンのアプリケーションであり、複数のプロセスが厳密に同時に通信を行わなくてもクラスタ間接続のバンド幅を有効活用できる提案手法が他に比べて高い性能を示している。このベンチマークでは提案手法は既存手法の 3.3 倍の性能を得ている。

MG

最後に提案手法がうまく働かない例があることも示す。図 10、12 は NAS Parallel Benchmarks の MG (Class C) の結果である。このベンチマークでは既存の手法でもパケットロスなどが発生せず、全体を高速ネットワークで接続した場合も含め、通信方式によらず似たような性能となっている。このようなアプリケーションに対しては提案手法は既存手法と同程度の性能となるのみであり、大きな性能向上は得られない。

3 高速ファイル転送ソフト SRFT の作成

3.1 背景

2007 年現在でサービス提供中の SR11000、また 2008 年より稼働開始予定のオープンスーパーコンピュータは共に HSFS と呼ばれるファイルシステムを使用している。このファイルシステムは特殊な

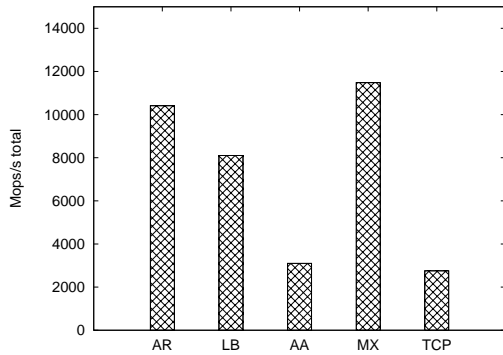


図 9: CG Results

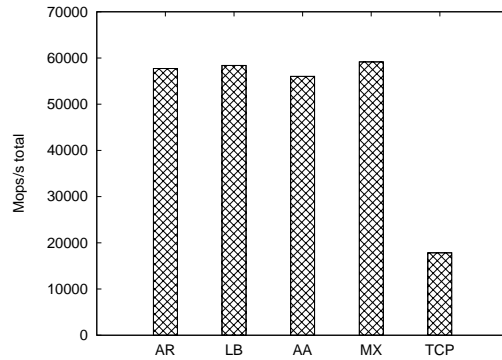


図 10: MG Results

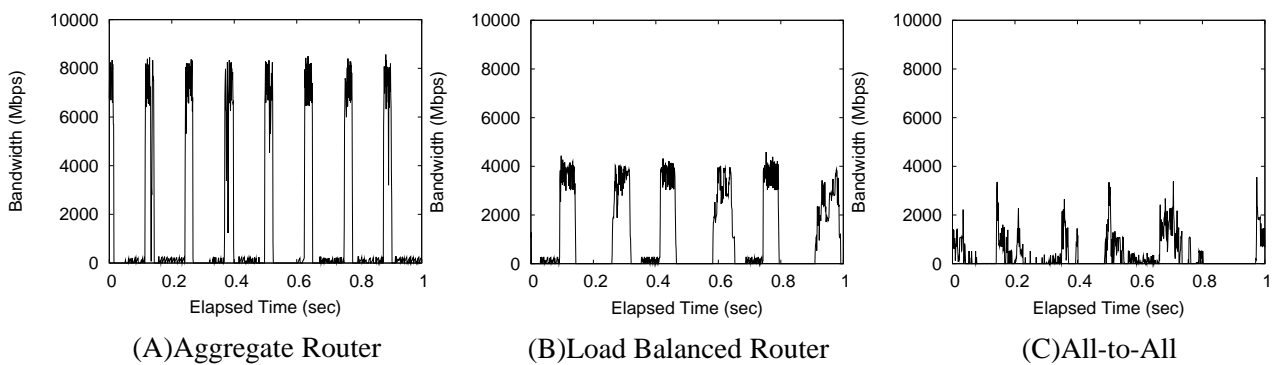


図 11: Traffic Sampling of CG

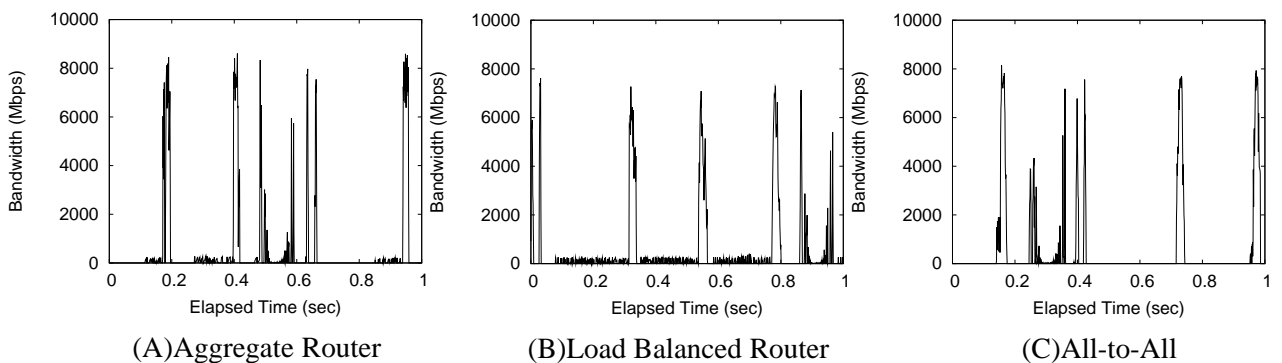


図 12: Traffic Sampling of MG

ファイルシステムであり、通常では考えられないほどの巨大なバッファを確保してファイルへのアクセス要求を出さないと高い性能が得られない。具体的には、通常のアプリケーションソフトは 8KB 程度のバッファでファイル入出力を行っているのに対し、HSFS で高い性能を得るには 16MB 程度のバッファが必要である。

SR11000 のユーザーが外部とファイル交換を行う場合、通常は scp と呼ばれるソフトウェアを使用する。これは SR11000 特有のソフトウェアではないため HSFS には最適化されておらず、上で述べたように 8KB 程度のバッファでファイル入出力を行っている。したがってファイルアクセスの性能が非常に悪く、3MB/s 程度の性能しか得られない。サービス品質向上のためにはこの転送速度を解決する必要があった。

3.2 内容

ファイル転送の速度に影響するのは以下の3点である。

1. ネットワーク性能
2. 暗号化処理の性能
3. ファイルアクセス性能

SR11000の現状を調べたところ、ネットワーク性能、暗号化処理性能にはほぼ問題がないことがわかった。したがって、ファイルアクセス性能のみを解決すればファイル転送速度は改善可能な状況であった。scpでファイル入出力性能がよくない要因は背景で述べたバッファサイズの問題であるが、さらに詳しく調べたところHSFSではファイルを開く操作に100ms前後の時間がかかることがわかった。一個のファイルを開くために0.1秒という人間でも体感できるほどの遅延があるファイルシステムは珍しいため、通常のソフトはこのような遅延があることは想定していない。これらのファイル入出力に関する問題を解決するため、以下のような特徴を備えるHSFS向けファイル転送ソフトSRFTを作成した。

- 16MBの大きなバッファを確保し、ファイル入出力は16MBずつオペレーティングシステムに要求を出すようにした。これによりHSFSでも高速にファイル入出力ができる。
- ファイルを開くための時間を隠蔽するため、複数ファイルの転送の際にはファイルオープン専用のスレッドがあらかじめファイルを開いて待つ仕組みにした。
- 16MBのファイル入出力の間もネットワークアクセスが可能になるよう、ネットワーク入出力とファイル入出力を別のスレッドに割り当て、I/O待ち時間を隠蔽した。ただしSR11000のログインノードは一ユーザーが同時に複数のプロセッサを使用しないことをルールとしているため、必ずI/O待ちと他の処理のみが重なるようにし、複数のスレッドが同時にCPUを使うことのないように実装した。
- OpenSSLによる暗号化機能も組み込み、安全にファイル転送が行えるようにした。
- scpとまったく同じ方法で使用できるよう、ユーザーインタフェースを作成した。

3.3 具体的成果

研究室内のファイルサーバからSR11000にファイル転送を行った際に実際に得られた性能を示す。上が作成したSRFT、下が既存の方法であるscpである。scpが2.7MB/sなのに対し、SRFTは70.4MB/sを達成しており、26倍もの高速化を実現した。

```
[matsuba@fs ~]$ srft x1 x2 xxxxxx@sr11000-s.cc.u-tokyo.ac.jp:
xxxxxx@sr11000-s.cc.u-tokyo.ac.jp's password:
x1                500.0MB/500.0MB (100.0%)  72.4MB/s    0h00m07s
x2                500.0MB/500.0MB (100.0%)  73.3MB/s    0h00m07s
Total             1.0GB/  1.0GB (100.0%)  70.4MB/s    0h00m14s
```

```
[matsuba@fs ~]$ scp x1 x2 xxxxxx@sr11000-s.cc.u-tokyo.ac.jp:
xxxxxx@sr11000-s.cc.u-tokyo.ac.jp's password:
x1                100%  477MB    2.7MB/s    02:59
x2                100%  477MB    2.7MB/s    02:59
```

本ソフトウェアはソースと SR11000 向けのバイナリを Web で公開している [公開 2]。また、スーパーコンピューティングニュース 2007 年 9 月号に解説記事を掲載し、SR11000 ユーザーが使用方法を解説している。本ソフトはセンターの公式サービスとしては位置付けていないが、個人的にユーザーサポートを行っており、ユーザーからの使用方法の問い合わせや性能報告を受けている。

4 成果要覧

査読付論文

[査読付 1] Akihiro Nomura, Hiroya Matsuba, Yutaka Ishikawa: Network Performance Model for TCP/IP Based Cluster Computing, IEEE International Conference on Cluster Computing (Cluster 2007), pp. 194-203, 2007

[査読付 2] Masato Sakai, Hiroya Matsuba and Yutaka Ishikawa: Fault Detection System Activated by Failure Information, Proceedings of the 13th Pacific Rim International Symposium on Dependable Computing (PRDC'07), pp. 19-26, December 2007

その他の発表論文

[発表 1] 藤田 肇, 松葉 浩也, 石川 裕: 柔軟な負荷分散を可能にする分散型シングル IP クラスタ, 情報処理学会 研究報告 2007-OS-106, pp. 1-8, 2007

[発表 2] 平野貴仁, 藤田 肇, 松葉 浩也, 石川 裕: PBus : 柔軟なカーネル機能拡張のためのインタフェース, 情報処理学会 研究報告 2007-OS-106, pp. 63-70, 2007

[発表 3] 北村 裕太, 松葉 浩也, 石川 裕: 大規模メモリ空間の利用を支援する遠隔スワップメモリシステム, 情報処理学会 研究報告 2007-HPC-111, pp.121-126, 2007

公開ソフトウェア

[公開 1] MPI 通信ライブラリ YAMPI3, <http://www.il.cc.u-tokyo.ac.jp/~matsuba/yampi3/>

[公開 2] 高速ファイル転送ソフト SRFT, <http://www.il.cc.u-tokyo.ac.jp/~matsuba/srft/srft.html>

ソフトウェア自動チューニングおよび 高性能数値計算ライブラリの研究と HPC 教育

片桐 孝洋

1 概要

本報告は、以下の研究・教育および業務報告に大別される。

- ソフトウェア自動チューニングに関する研究
 - PC からスーパーコンピュータに至る広範な計算機環境において、自動的に性能チューニングが達成できるソフトウェアの理論、方式、言語、およびライブラリの総合的研究。
- 高性能数値計算ライブラリに関する研究
 - スーパーコンピュータなどの先進的計算機上で高性能を達成するための数値計算アルゴリズムおよび数値計算の実装方式の研究。本年度は、近年米国カリフォルニア大学バークレー校で開発された、新しい固有値問題の解法である **MRRR** 法を取り扱う。
- 高性能計算 (HPC) における学部・大学院教育業務、およびスーパーコンピュータのユーザ教育に関連する業務
 - センターのスーパーコンピュータを用いて行う並列プログラミングの授業を、東京大学工学部および工学系研究科共通科目として行う講義、および、駒場キャンパスの教養学部学生を対象に行う全学ゼミの講義。
 - スーパーコンピュータユーザに並列プログラミングの教育を行う制度、および新規ユーザを獲得するための制度に関する業務。
- スーパーコンピュータの民間開放に関する業務
 - 文部科学省「先端研究施設共用イノベーション創出事業」【産業戦略利用】「先端的大規模計算シミュレーションプログラム利用サービス」に関する事業

2 ソフトウェア自動チューニングの研究

2.1 背景

携帯電話や自動車の制御系に代表される組み込みシステム用ソフトウェア(以降、組み込みソフトウェア)技術は、我が国が世界に誇る先端情報技術の一つである。しかしながら、組み込みソフトウェアの開発は、技術的要求水準の厳しさと開発サイクルの短さから、既存の工学的手法が適用できない分野である。すなわち、ソフトウェアの性能が、プログラマの職人芸にいまなお依存している。

組み込みソフトウェア開発の現状に類似して、ソフトウェア工学を基盤とするソフトウェア開発方法論やプログラム自動生成技術による開発工期削減が必要な分野がある。それは、スーパーコンピュータ上の数値計算ソフトウェアの開発分野である。

近年の数値計算ソフトウェアは、性能に影響を及ぼす計算機上およびアルゴリズム上のパラメタが多数存在する。このことは容易な利用の妨げになるばかりか、高性能を達成するため複雑で時間

を浪費する、職人芸的なパラメタチューニングを人手でしなくてはならないという現状がある。人手によるチューニングの要因、および問題点は、以下にまとめられる。

- **先進計算機アーキテクチャの複雑化:** 非対称メモリアクセス、マルチコア化という近年の計算機アーキテクチャのトレンドは、コンパイラによる自動最適化を複雑化させるばかりか、エンドユーザのアプリケーション性能が安定しないなどのチューニングを複雑化させる要因となっている。
- **京速コンピューティングの困難性:** 京速コンピューティング(ペタ・フロップス・コンピューティング)を達成するスーパーコンピュータの開発プロジェクトが進んでいる。しかし、このような計算機環境で高性能を達成するには、対象となる数値計算ソフトウェアにおける通信処理と計算処理のパラメタ自動チューニングが必須であることが指摘されている。

以上の背景から本研究では、以下を目的とする。

1. ソフトウェア工学的観点(図1)から、ソフトウェア自動チューニング(以降、単に自動チューニング、ATとよぶ)という新しいソフトウェア・パラダイムの研究をおこなう。このAT研究により、パラメタチューニング作業が自動化され、最適プログラムが自動選択されることで、ソフトウェア再利用性を格段に高めることができる。ソフトウェア開発効率向上(ソフトウェア開発工期短縮)を狙うことができる。
2. AT機能を加えた実用となる数値計算ソフトウェアの開発を行う。これにより、日本発の高性能で自由入手可能なソフトウェアの普及を目指す。
3. AT機構をソフトウェアに追加する際に便利な計算機言語、ミドルウェア、システムソフトウェア、およびAT支援ツールの開発を目的とする。

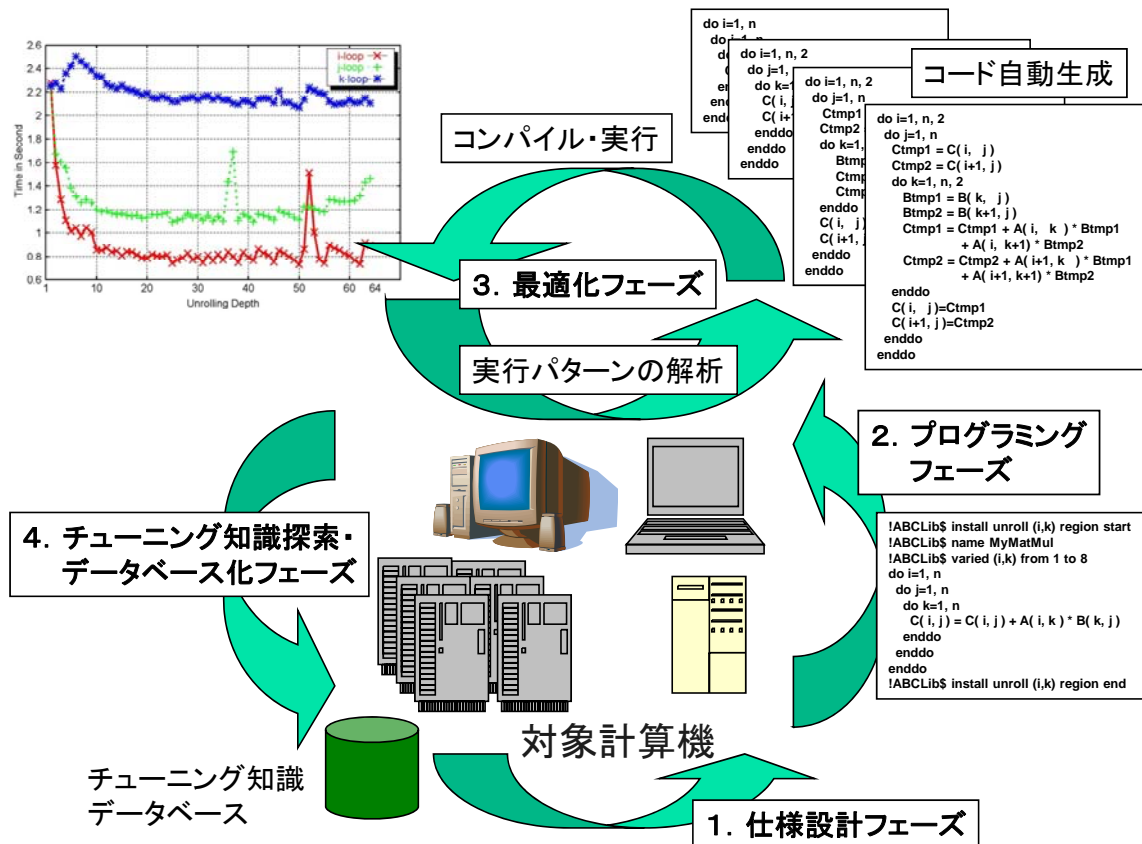


図1 自動チューニングソフトウェア工学の全体像

2.2 内容

本研究によるAT機能の提供から、数値計算および並列処理を専門としない利用者に対して、容易に利用できる高性能数値ソフトウェアが初めて開発可能となった。本研究では、世界に先駆けて数値計算ライブラリに限定しないATパラダイムの提案を行ってきた。

具体的には、3種の最適階層を有し、かつチューニングのタイミングも3種であることを特徴とする、汎用的な自動チューニングソフトウェア構成方式 *FIBER (Framework of Install-time, Before Execute-time, and Run-time optimization layers)* の提案と有効性評価を行ってきた。

2.3 具体的成果

本研究は、2002 年から継続して成果を出してきている。2007 年度における主な研究成果は以下の通りである。

1. **AT記述用言語ABCLibScript[公開1]における組み込みシステムへの適用と有効性検討[発表1]:** 科学技術研究費補助金・基盤研究C・18500018「数値計算と組み込みシステムのための自動チューニング方式」による成果である。AT方式を組み込みソフトウェア開発環境に適用した結果、最大で6.4倍のATによる速度向上を得た。
2. **Windows CCSにおけるMS-MPIの実行時自動チューニング方式研究[招待6、7]:** 近年基盤ソフトウェアが整備されてきているWindowsCCSクラスタによる、MPIの実行時自動チューニング方式の研究である。アプリケーションに依存するが、疎行列の連立一次方程式の解法において、最大で20倍も速度向上する例があった。なお本研究は、マイクロソフト産学連携研究機構(IJARC)戦略プロジェクトの支援による。本研究は情報基盤センター黒田久泰助教との共同研究である。
3. **AT記述用言語ABCLibScriptにおける超省電力記述法の研究[招待5]:** AT言語を用いて、CPUにおける周波数切り替え機能を用いた低電力最適化を実現するAT方式研究である。なお、本研究は、JST-CREST、「ULP-HPC:次世代テクノロジーのモデル化・最適化による超低消費電力ハイパフォーマンスコンピューティング」(研究代表者: 東京工業大学 松岡聡教授)の支援による。
4. **AT時間を削減する標本点抽出方式の研究[査読付2、3]:** AT時間を削減するための、適切な標本点を決定するアルゴリズムの研究である。数値計算における密行列直接解法、疎行列反復解法ともに有効性が確認された。なお本研究は、電気通信大学弓場敏嗣名誉教授、日立製作所田中輝雄博士との共同研究である。

継続した AT 研究の成果に対して、Microsoft 社が主催している Microsoft Innovation Award アカデミック部門において、最優秀賞を受賞した[受賞1]。

3 高性能数値計算ライブラリ開発(多固有値多分法の開発)

3.1 背景

高性能な数値計算ライブラリは、人類の叡智の結晶であり、極めて重要な知的財産であることはいままでもない。とりわけ固有値問題の求解においては、数学的な困難性に加えて処理の複雑さから、高い並列実行性能を先進的な計算機で達成できるライブラリ(固有値ソルバ)が、産学界から強く求められている。

二分法(Bisection Method)は、対称三重対角行列の固有値を計算するために広く用いられている方法である。しかしながら、計算すべき固有値が密集している場合、計算時間が増加するという問題がある。これは、幅広く用いられている数値計算ライブラリ LAPACK の二分法のルーチンにおいても例外ではない。この問題を解決するため、二分法の並列化が強く要請されている。特に、共有メ

メモリ型並列計算機での並列化方式が、多くのスーパーコンピュータの計算ノードが共有メモリ型並列計算機となっているため、特に重要な技術となっている。

本報告では、複数固有値を計算するための二分法における新並列化方式を提案することを目的とする。対象は共有メモリ型並列計算機である。この手法の主なアイデアは、複数固有値計算に必要な処理の自然な並列性と、固有値の存在区間における複数の探索点の並列性の2並列性を利用し、性能を極限まで抽出することにある。我々はこの新手法を、**多固有値多分法 (Multi-section with Multiple Eigenvalues method, MME)**と呼んでいる。

二分法を並列化するため、複数の探索点を利用することは新しいアイデアではない。この方法は、多分法 (Multisection Method) と呼ばれている。たとえば、Lo らや Simon の論文で、多分法の提案がなされている。しかしながらこれらの手法は、ベクトル型計算機を対象とした上で、複数の探索点を利用しベクトル化を行うものであった。ベクトル化をするため、演算中心となるループ(カーネル・ループ)において、IF 分がその中心に位置されている。

これら従来手法に対し我々が提案する手法は、多数のプロセッサ・エレメント(PE)を有する共有メモリ型並列計算機によるスレッド実行を意識し、カーネル・ループにおける最外ループの自然な並列性に着目するものである。加えて、複数固有値を計算する際の並列性も利用する。この2並列性を利用することで、最外ループの長さ(ループ長)を従来手法に対して長く保つことができる。それにとどまらず、探索点が少数しかとれない場合においても、最外ループ長を長く保てるという長所がある。これらの特徴はいずれも、共有メモリ型並列計算機において高並列効率を達成するため、極めて重要な要素となる。現在のスーパーコンピュータでは、ノード内 PE 数は 16PE 程度が普通であり、今後さらにノード内 PE 数が増加していくという計算機アーキテクチャのトレンドを考慮すると、ノード内並列性を少しでも高く保つことが、高性能計算で決定的要因となる。

なお、本研究は米国カリフォルニア大学バークレー校コンピュータサイエンス学科 James Demmel 教授との共同研究である。

3.2 内容

ここで示される二分法の実装は、LAPACK 3.1.1 の DSTEGR ルーチン中で呼び出されている二分法を用いた固有値計算ルーチンの dlarrb ルーチン(dlarrb.f)を基にしている。具体的に、図 2はdlarrbルーチンで使われる二分法の演算カーネルであるdlaneg関数(dlaneg.f)の概略を示している。

図 2のカーネルでは、 σ より小さな固有値数を返す。ここで

図 2のカーネルでは、ターゲットとなる固有値1つについて、探索点 σ 以下の固有値数が計算される仕様となっている。またMRRRアルゴリズムの特徴から、dlarrbルーチンで計算される固有値数は、入力された三重対角行列に対するRelatively Robust Representation (RRR) という形式を基に決まる。 σ の値は、対象となる固有値の存在範囲から決まる。この存在範囲は、MRRRアルゴリズムの場合、入力された三重対角行列に対するRepresentation Treeで決定される。

dlarrb ルーチンでは、MRRR アルゴリズムで必要となる $L D L^T$ 分解に「特化された」二分法を実行する。すなわち、以下の3つの部分がある：

- I. $L D L^T - \sigma I = L_+ D_+ L_+^T$ に関する「上三角行列部分」
- II. $L D L^T - \sigma I = U_- D_- U_-^T$ に関する「下三角行列部分」
- III. ねじれ部分

高速化のため、NaN を考慮した IEEE-754 浮動小数点形式に特化した処理が実装されている。

図 2の σ の値は、探索中の固有値に対する二分探索の基準点である。この σ は、現在の固有値存

在区間を[a,b]とした場合、 $a+(b-a)/2$ で計算される。

我々は、二分法のカーネルを、I部分の< 2 >~< 7 >行とする。なぜなら、II部分においても、計算に必要な変数のアクセスパターンはI部分と同様である。

```

< 0 > S = 0; P=0; NEG1=0; NEG2=0; NEGCNT=0;
< 1 > I) 上三角行列部分
< 2 > do J = 1, R-1
< 3 >   T = S -  $\sigma$ 
< 4 >   DPLUS = D(J) + T
< 5 >   S = T * LLD(J) / DPLUS
< 6 >   if ( DPLUS .lt. ZERO ) NEG1 = NEG1 + 1
< 7 > enddo
< 8 > if ( S .eq. NaN ) 上記ループより遅いバージョンを使う;
< 9 > NEGCNT = NEGCNT + NEG1
< 10 > II) 下三角行列部分
< 11 > do J = N-1, R, -1
< 12 >   DMINUS = LLD(J) + P
< 13 >   P = P*D(J) / DMINUS -  $\sigma$ 
< 14 >   if ( DMINUS .lt. ZERO ) NEG2 = NEG2 + 1
< 15 > enddo
< 16 > if ( P .eq. NaN ) 上記ループより遅いバージョンを使う;
< 17 > NEGCNT = NEGCNT + NEG2
< 18 > III) ねじれ部分
< 19 > GAMMA = S + P
< 20 > if ( GAMMA .lt. ZERO ) NEGCNT = NEGCNT+1
< 21 > return (NEGCNT)

```

図 2 LAPACK 3.1.1 dSTEGR における二分法の演算カーネル。変数 R は、三重対角行列 T のねじれ分解における「ねじれ点」を意味している。

3.2.1 二分法カーネルはなぜ並列化できないか？

図 3は、いま対象としている一つの固有値に対する二分法のカーネルとする。

```

< 0 > S=0; NEG1=0;
< 1 > do J = 1, R-1
< 2 >   T = S -  $\sigma$ 
< 3 >   DPLUS = D(J) + T
< 4 >   S = T*LLD(J) / DPLUS
< 5 >   if ( DPLUS .lt. ZERO ) NEG1 = NEG1 + 1
< 6 > enddo

```

図 3 二分法のカーネル

図 3のカーネルは、そのままでは並列化ができない。これは、変数 S について、<4>行で定義された値が J-ループをまたいで、<2>行で参照されるためである。このようなデータの依存関係を、ループ伝搬フロー依存 (loop-carried flow dependency) と呼ぶ。

図 3のループを並列化するためには、いろいろな方法がある。二分法については、2 種の方法が知られている。

まずはじめの方法は、二分法で用いる固有値の存在区間を分割し、それぞれについて並列に探索していく方法である。しかしながらこの方法では、図 3のカーネル自体が並列化されない。さらにベクトル型計算機においてはカーネルがベクトル化できないので、演算効率は悪くなる。

もう一つの方法は、並列接頭辞法(parallel prefix method)という方法を用いて、固有方程式 $p_k(x) = \det(T_k - x I)$ に関する<二分木状の>処理を並列化する方法である。この方法は並列性を抽出でき、さらにベクトル計算機においてもベクトル化ができるようになるが、残念なことに数値安定性が悪く、固有値数のカウントを誤る。したがって実用上使われない。

3.2.2 多分法のカーネル

図 3のカーネル自体を並列化する方法は、多分法を用いることである。図 4に、対象の一つの固有値を計算する際にML点の探索点を使う場合における、多分法のカーネルを示す。

```

< 0 >  S(1:ML)=0; NEG1(1:ML)=0;
< 1 >  do I = 1, ML
< 2 >    do J = 1, R-1
< 3 >      T(I) = S(I) -  $\sigma(I)$ 
< 4 >      DPLUS(I) = D(J) + T(I)
< 5 >      S(I) = T(I)*LLD(J) / DPLUS(I)
< 6 >      if ( DPLUS(I) .lt. ZERO ) NEG1(I) = NEG1(I) + 1
< 7 >    enddo
< 8 >  enddo
    
```

図 4 多分法のカーネル

図 4の配列 $\sigma(1:ML)$ には、現在の固有値存在区間を[a,b]とすると、 $\sigma(i) = a + h * i$ ($i=1,2,\dots,ML$)、 h は $(b-a)/(ML+1)$ と定義したものが設定される。図 4のカーネルは、最外ループであるIループについて、並列化ができる。なぜなら、ループ変数Iに関連する変数(配列)は、まったくデータ依存がない。

しかしながら、多分法のカーネルにも問題がある。もし並列度を上げるために、最外ループIの長さ(ループ長)を長くしたいとする。このとき、図 4の探索点数MLを、できるだけ大きく設定しなくてはならない。ところがMLに大きな値を設定すると、今度は二分法の探索効率が悪くなる場合がある。なぜかという、もし従来の二分法において探索1回で見つかる固有値分布の場合、多分法で探索点を増やすと、探索点を増やした分だけ余分な計算をすることになり、結果として二分法に対し実行時間が増える。

つまり多分法には、<並列化効率>と<探索効率>のトレードオフがある。

本稿で提案する方法は、この多分法のトレードオフを解消するものである。つまり、多分法の探索点数を低く抑えつつも、同時計算する固有値の計算処理の並列性を使うことで、高速実行と高探索効率を同時に実現する方式である。我々はこの新しい方式を、**多固有値多分法(Multi-section with Multiple Eigenvalues method, MME 法)**と呼ぶ。

3.3 具体的成果

性能評価のための入力行列について、以下の4種類の行列を用いた。

- 行列サイズ: 2100
- 行列 1: (-1,2,-1) 行列
- 行列 2: 0 ~ 1 の値で生成される一様乱数.
- 行列 3: Wilkinson 行列 W^+_{2100}
- 行列 4: 「副対角要素を循環シフトした」Glued Wilkinson Matrix W^+_{21} .
 - この行列は以下のように定義する: $\text{Diag}(T) = \text{Diag}(\text{Glued Wilkinson Matrix } W^+_{21})$. $\text{Sub}(T)$ は $(\delta, 1, \dots, 1)_{21}$ を 100 回繰り返したものの。glue value の δ は $1e-1$.

対象アプリケーションは、LAPACK 3.1.1 の DSTEGR ルーチンである。詳細は以下のとおり。

- **対象ルーチン**: 多固有値多分法を DSTEGR ルーチンの二分法ルーチンである dlarrb ルーチンに実装したもの。
- **対象計算**: 全固有値および全固有ベクトルの計算。
- **モード**:
 - DQDS モード (LAPACK の DLASQ1)
 - ◇ このモードでは、DQDS 法を固有値計算部分で利用する。また、固有値の精度を改良するため、固有ベクトル計算部分で二分法を利用する。
 - 積極的な二分法モード
 - ◇ このモードでは、固有値計算部分および固有ベクトル計算部分の両方で、二分法を利用する。

以上の2モードを、本性能評価で利用する。

- **測定対象**: 多固有値多分法を実装した dlarrb ルーチンの総実行時間を計測する。
- **パラメタに関する「静的固定方式」**: MME 法の性能パラメタ EL と ML について、DSTEGR ルーチンが終了するまで固定値で実行する方式である。なお、この固定値は、以下の値について変動させて時間を計測します。最高速となるパラメタのものを、性能評価対象の各方式(多分法、多固有値多分法)の実行時間とする。
 - EL \ni [1, 2, 3, 4, 8, 16, 32]: 7 種類
 - ML \ni [1, 2, 4, 8, 16, 24]: 6 種類
 すなわち、EL * ML = 42 種類のうち最適なパラメタによる実行時間とする。

表 2 は、多分法に対する静的固定方式による多固有値多分法の効果を示している。表 1 より、多固有値多分法を利用することで、二分法に対して最大で 7.7 倍、多分法に対して最大で 4.3 倍の速度向上を得た。

表 1 多分法に対する静的固定方式による多固有値多分法の効果 (HITACHI SR8000)

(a) dlarrb ルーチンの総実行時間 (DQDS モード)

方法/行列種類	行列1	行列2	行列3	行列4
二分法[秒]	0.347	1.83	15.2	8.20
多分法[秒]	0.323	1.45	5.90	3.30
最適 ML	8	8	16	16
多固有値多分法[秒]	0.075	1.16	5.34	1.75
最適(EL, ML)	(16,2)	(2,8)	(2,8)	(16,1)
二分法に対する速度向上	4.6x	1.5x	2.8x	4.6x
多分法に対する速度向上	<u>4.3x</u>	1.2x	1.1x	1.8x

(b) dlarrb ルーチンの総実行時間(積極的な二分法モード)

方法/行列種類	行列1	行列2	行列3	行列4
二分法[秒]	3.94	7.11	23.9	17.2
多分法[秒]	2.11	3.69	8.70	6.19
最適 ML	16	16	16	16
多固有値多分法[秒]	0.51	2.93	7.89	2.75
最適(EL, ML)	(16,1)	(2,8)	(2,8)	(16,1)
二分法に対する速度向上	<u>7.7x</u>	2.4x	3.0x	6.2x
多分法に対する速度向上	4.1x	1.2x	1.1x	2.2x

MME 法は、スーパーコンピューティング部門所有のスーパーコンピュータ HITACHI SR11000/J2 モデルにおいても有効性評価を行った[招待 4]。MME 法の性能パラメタ EL、ML を自動チューニングで決定する AT 方式の開発も行っている[査読付 1]。

4 HPC 教育関連業務

4.1 背景

東京大学情報基盤センター特任准教授として 2007 年 4 月 1 日に赴任後、教育業務として並列プログラミングの基礎の講義と本センターのスーパーコンピュータを用いた演習を、東京大学工学部・工学系研究科の共通科目「スパコンプログラミング(1)および(I)」として、通年(夏学期・冬学期)各1コマで同じ講義を行った。冬学期には、駒場キャンパスの教養学部の学生に、工学部と同様の授業を教えることで、天才プログラマを早期から育成することを目指す、全学ゼミ「スパコンプログラミング研究ゼミ」を開講した。

内容は、通信ライブラリ MPI を利用し、情報基盤センターに設置されているスーパーコンピュータ HITACHI SR11000/J2 モデル(8PE/ノードまで)を受講生に無料で利用させるものであり、スーパーコンピュータの利用方法の講習を含む。したがって、情報基盤センターのスーパーコンピュータの将来的な新規ユーザ獲得も狙った講義である。本講義の報告は、[発表 5]を参照されたい。

また、センターユーザの HPC 教育を行う業務を、本年度に複数企画して実行した。この企画の詳細は次節に記述する。

4.2 内容(HPC 教育業務)

- **スーパーコンピュータ若手利用者推薦(試行)**(2007 年 10 月から実施)
 - **概要:** 東京大学情報基盤センター(以下、センター)では、概ね 35 歳以下の若手研究者(学生を含む)を対象とした利用者向けの、スーパーコンピュータ若手利用者推薦(試行)(以降、若手推薦(試行))による課題の公募を開始する。スーパーコンピューティング部門の教員により審査の上、採択された課題の計算機利用負担金(半年分)をセンターが負担する。
 - **詳細:** 年 2 回公募し、年間で 4 件程度の優れた研究提案を採択する。継続申請と再審査の上で、最大で 1 年間の無料利用ができる。採択の方には、報告書の提出、研究成果の発表の際に若手推薦(試行)を利用したことの明記、およびセンターが発行する「スーパーコンピューティングニュース」誌の原稿執筆を採択の条件とする。

- **お試しアカウント付き並列プログラミング講習会(試行)**(2008年3月に実施)
 - **趣旨:**情報基盤センター(以降、センター)に設置されたスーパーコンピュータの利用者における利用促進、および利用を検討している新規ユーザを対象とした、センターのスーパーコンピュータを用いた並列プログラミング講習会(MPI および OpenMP を用いるもの。終日の講習会。)を実施する。
 - **対象者:**センターのスーパーコンピュータのアカウントを所有しているユーザのみならず、センターのスーパーコンピュータの利用資格を有し、今後の利用を検討している方(大学教員、および学生など)を対象にする。
- **先進スーパーコンピューティング環境研究会**(2008年3月に実施)
 - **概要:**スーパーコンピュータでの高速化の鍵となる通信ライブラリや数値計算ライブラリ、およびスーパーコンピュータシステムの技術革新が日進月歩である。さらに従来の高速化要求に加え、スーパーコンピュータの利便性を追求する新しい価値観や、利便性を達成する技術要求が、スーパーコンピュータの利用分野においても必要とされている。これらの新技術要求は、ソフトウェア研究を行っている計算機科学者のみならず、数値シミュレーションをおこなっている計算科学者においても、研究を推進するための重要なテーマとなっている。このような時代の変化と要求にともない、情報基盤センター(以降、センター)に設置されたスーパーコンピュータに関係する最新の研究成果をユーザが知ること、スーパーコンピュータの利用成果を著しく創出するための必要な技術知識となる。また、センターのユーザが個別に所有する情報に関して、ユーザ間での情報交流の活性化も、研究の活性化に必要と考える。以上の趣旨から、スーパーコンピューティング部門が主催する、スーパーコンピュータ環境に関する研究会を発足した。
- **スーパーコンピューティングニュース特集号発刊**(2008年2月発刊)
 - 隔月で発刊している「スーパーコンピューティングニュース」の特集号を号担当教員として編集した。
 - 具体的には、SR11000/J2 の高速化、T2K オープンスパコンの高速化、共通高速化に関する技術について、国内の著名な若手研究者を中心とする 8 名に原稿を依頼して編集業務をおこなった。
- **情報基盤センターJSPS 招へい研究員の招へい**(2008年2月～)
 - 研究業務の活性化のため、日本学術振興会短期研究者招へい制度に応募し、採択された。東京大学情報基盤センター研究員として、米国ローレンスバークレー国立研究所の Osni A. Marques 博士を、2008年2月1日～4月6日までの約2ヶ月間、情報基盤センターJSPS 招へい研究員として招へいした。数値計算ライブラリ分野での本センターにおける研究業務活性化に貢献した。加えて若手研究者との交流により、HPC 分野の教育研究の活性化に貢献した。

4.3 具体的成果

- 東京大学における HPC 教育
 - 工学部共通科目において、夏学期は登録人数 80 名超、単位取得者数 30 名程度を達成した。冬学期においては、登録者 29 名、単に取得者 13 名であり、夏学期に比べて単位取得率の向上を達成した。詳しくは、[発表 5]を参考のこと。
 - 教養学部の全学ゼミにおいては、登録 13 名、合格 7 名であった。上位の 3 名は、工学部学生・大学院生より勝るレポートを提出した。
- スーパーコンピュータ若手利用者推薦(試行)(平成 19 年度後期)
 - 筑波大学、埼玉大学、大阪大学からの 3 件の課題を採択した。
- お試しアカウント付き並列プログラミング講習会(試行)

- 他大学ユーザを含む 10 名の参加があった。講習会に関するアンケートを実施し、1 から 5 までの 5 段階評価(5 が高い)による満足度において、平均 4.0 という高い顧客満足度を得た。
- Marques 博士は、東京大学理学部、理化学研究所情報基盤センター、名古屋大学工学部、京都大学学術情報メディアセンターで招待講演を行い、日本の若手研究者らと、当該分野の研究交流を大きく活性化させた。これらは、HPC 分野の若手研究者の教育および分野啓蒙にも貢献したと考えられる。
- その他、先進的計算機アーキテクチャ「Cell」を用いたプログラミングコンテスト「Cell スピードチャレンジ 2008」(情報処理学会計算機アーキテクチャ研究会主催)において、問題作成委員長として規定課題の問題(連立一次方程式の解法)を作成した[特記 7]。国内の大学の学生・大学院生 69 名の参加者が集まった。予選、本選の実施を通して、高度な並列プログラムの開発がなされた。2008 年 6 月 11 日～13 日に、つくば国際会議場で開催される「SACIS2008—先進的計算基盤システムシンポジウム」の特別セッションで入賞者の実装内容が講演される予定である。この成果は、並列プログラミング教育および HPC 分野の啓蒙に貢献するものである。

5 スーパーコンピュータの民間開放に関する業務

5.1 背景

本業務は、文部科学省「先端研究施設共用イノベーション創出事業」【産業戦略利用】「先端的大規模計算シミュレーションプログラム利用サービス」(以降、共用イノベーション事業)の支援に関する業務である。

全国共同利用施設を有するスーパーコンピュータを所有する 7 国立大学法人(北海道大学、東北大学、東京大学、名古屋大学、京都大学、大阪大学、九州大学)は、社会貢献の一環として、大学で開発された応用ソフトウェアとスーパーコンピュータの民間企業への提供を、共用イノベーション事業の支援のもとに開始した。7 大学が有するスーパーコンピュータの総計算能力(理論性能値)は、平成 19 年度 99TFlops、平成 20 年度には約 300TFlops(予定)になる。

本事業の特徴は、各大学が開発した応用ソフトウェア(シミュレーションソフトウェアなど)と各大学が持つ国内有数のスーパーコンピュータを、それらの利用支援サービスと一体化して民間企業に提供する。このようにサービス、ソフトウェア、ハードウェアの 3 つを同時に支援する試みは、国内外を通じて初めてとなる。この事業の目的は、大学の持つ知と施設によって我が国の経済発展に貢献することである。

5.2 内容

企業におけるもの作りでは創薬のための分子設計やナノ技術分野において、東大を中心とする 7 大学は、ASP (Application Service Provider) 事業を行う NPO 等の団体・企業と協力して、「戦略分野利用推進」と「新規利用拡大」の2つのカテゴリに分けて支援する。

「戦略分野利用推進」では、ライフサイエンス、もの作り技術、情報通信、環境エネルギー、社会基盤分野に関する戦略分野を対象に、大学で開発された先端的シミュレーションソフトウェアとスーパーコンピュータの利用課題を民間企業から募り、企業によるイノベーション創出を支援する。民間企業が有するアプリケーションを利用した利用課題、あるいは新規アプリケーションを開発する利用課題も募る。

「新規利用拡大」では、これまで利用したことのない利用者および利用分野を対象に、新たに有望な戦略分野につながる利用課題を民間企業から募る。必要に応じ、ASP 事業を行う団体・企業の協力を仰ぎ、今まで、スーパーコンピュータを利用したことのない民間企業を支援する。これにより、

計算シミュレーションによるものづくりの企業への浸透、スーパーコンピュータの新しい民間利用を促進する。

今年度は利用課題の公募を 2 回行い、7 大学で組織する採択委員会において利用課題を決める。戦略分野利用推進では最大 2 年間、新規利用拡大では最大 1 年間、無償でソフトウェアとスーパーコンピュータを利用できる。

5.3 具体的成果

- 平成 20 年度課題申込書、募集要項、成果報告書の作成業務
- 本事業に関する事務局業務(質問受け付け、課題受け付け、広報、その他の事務業務補助)
- 共用イノベーションのホームページの開設と保守管理業務
- 平成 19 年度、第 1 期(19 件)、第 2 期(2 件)、および平成 20 年度、第 1 期(9 件)の新規課題の選定業務(7 大学センターを含めた採択件数)

6 成果要覧

受賞関連

[受賞 1] 片桐孝洋:Microsoft INNOVATION AWARD 2007, アカデミック部門, 最優秀賞受賞, 2007 年.

招待講演／招待論文

[招待 1]片桐孝洋:自動チューニング記述用言語 ABCLibScript, 第 12 回日本計算工学会講演会, 「数値計算ライブラリにおける自動チューニング研究最前線」, 国立青少年オリンピック記念センター, 計算工学会講演論文集, 第 12 巻, 第 2 号, pp.563-566, 2007 年.

[招待 2] 片桐孝洋:WindowsCCS 上の数値計算ライブラリのための MS-MPI の実装方式の自動チューニング, イノベーションジャパン 2007—大学見本市, 2007 年 9 月.

[招待 3] Takahiro Katagiri:Towards Auto-tuning Framework for Numerical Libraries, First French-Japanese Workshop -- Petascale Applications, Algorithms, and Programming (PAAP) --, Nov. 2007.

[招待 4] Takahiro Katagiri:Performance Of Multisection With Multiple Eigenvalue Method For Symmetric Tridiagonal Eigensolver On Next Generation Multicore Processors, APCOM'07, Organized Session, Future directions of large-scale scientific computing and parallel linear solvers (Organizers: Kengo Nakajima, Jonathan Carter, Guy Lonsdale, Hiroshi Okuda), Proceedings of APCOM'07, 10 pages, (CD-ROM), Dec. 2007.

[招待 5] 片桐孝洋:超省電力 HPC ソフトウェアのための自動チューニング記述法, 東京工業大学グローバルCOE「計算世界観の深化と展開」, 計算世界観ワークショップ #4:高性能計算における超省電力化, 2007 年 12 月.

[招待 6] 片桐孝洋, 黒田久泰:HPCサーバにおけるソフトウェア自動チューニングのインパクト: MS-MPIの実行時自動チューニング機構の開発を例にして, Windows HPC アップデートセミナー, 2008 年 2 月.

[招待 7] 黒田久泰, 片桐孝洋:疎行列ソルバにおける MS-MPI の実行時自動チューニングの効果, Windows HPC アップデートセミナー, 2008 年 2 月.

査読付論文リスト

- [査読付 1] Takahiro Katagiri, Christof Voemel, and James Demmel: Automatic Performance Tuning for the Multi-section with Multiple Eigenvalues Method for the Symmetric Eigenproblem, Selected Paper of Workshop On State-of-the-art In Scientific And Parallel Computing (PARA'06), Springer LNCS 4699, pp.938-948, 2007.
- [査読付 2] Teruo Tanaka, Takahiro Katagiri, and Toshitsugu Yuba: d-Spline Based Incremental Parameter Estimation in Automatic Performance Tuning, Selected Paper of Workshop On State-of-the-art In Scientific And Parallel Computing (PARA'06), Springer LNCS 4699, pp. 986-995, 2007.
- [査読付 3] 田中輝雄, 片桐孝洋, 弓場敏嗣: ソフトウェア自動チューニングにおける標本点逐次追加型性能パラメータ推定法の疎行列計算への適用, 情報処理学会論文誌: コンピューティングシステム, Vol.48, No. SIG13 (ACS 19), pp. 223-234, 2007 年.
- [査読付 4] 塚本智博, 吉瀬謙二, 片桐孝洋, 弓場敏嗣: Alpha アーキテクチャ用 COINS マシン記述の実装と GCC との比較, コンピュータソフトウェア, Vol. 24, No. 3, pp. 4-14, 2007 年.

公開ソフトウェア

- [公開 1] 自動チューニング機構付加言語 ABCLibScript の Fortran90 と MPI のプリプロセッサ ABCLibCodeGen: <http://www.abc-lib.org/>
- [公開 2] 自動チューニング機能付き固有値ソルバ ABCLib_DRSSSED: <http://www.abc-lib.org/>
- [公開 3] 自動チューニング機能付き BLAS3 ライブラリ AutoTuned-RB: <http://www.abc-lib.org/>

その他の発表論文リスト

- [発表 1] 片桐孝洋: 自動チューニング処理記述言語 ABCLibScript の組み込みシステムへの適用の一検討, 2007 年並列/分散/協調処理に関する『旭川』サマー・ワークショップ (SWOPP2007), 旭川クリスタルホール, 情報処理学会研究報告 2007-HPC-111, pp.207-212, 2007 年 8 月.
- [発表 2] 片桐孝洋, 田中輝雄, 弓場敏嗣: ソフトウェア自動チューニングのための疎行列ライブラリ用標本点追加型性能パラメータ推定法, 京都大学学術情報メディアセンター広報, Vol.6, No.2, pp.29-32, 2007 年.
- [発表 3] 片桐孝洋: 対称三重対角ソルバにおける多固有値多分法の開発, 東京大学情報基盤センタースーパーコンピューティングニュース, Vol.9, No.5, pp. 25-35, 2007 年 9 月.
- [発表 4] 片桐孝洋: C 言語による MPI プログラミング入門, 東京大学情報基盤センタースーパーコンピューティングニュース 特集号, Vol.9, No.Special Issue 1, pp. 169-188, 2008 年 2 月.
- [発表 5] 片桐孝洋: 東京大学のスーパーコンピュータを用いた並列プログラミング教育—工学部・工学系研究科共通科目「スパコンプログラミング1および I」(2007 年度夏学期)を通じて—, 東京大学情報基盤センタースーパーコンピューティングニュース, Vol.10, No.2, pp. 12-19, 2008 年 3 月.
- [発表 6] 黒田久泰, 片桐孝洋: Windows CCS 上における MS-MPI の実行時自動チューニング, 2008 年ハイパフォーマンスコンピューティングと計算科学論文集 HPCS2008, HPCS2008 論文集 pp.54, 2008 年 1 月.

- [発表 7] Takahiro Katagiri: Towards General Auto-tuning Description Language on Advanced Computing Systems, 13th SIAM Conference on Parallel Processing for Scientific Computing (PP08), Mini Symposium, MS1: Auto-tuning on Numerical Libraries and Advanced Computer Systems: Part I of II, Atlanta, Georgia, USA, Mar. 2008.
- [発表 8] Hisayasu Kuroda, Takahiro Katagiri: Auto-tuning Effect of Iterative Method Library on Windows CCS, 13th SIAM Conference on Parallel Processing for Scientific Computing (PP08), Mini Symposium, MS25: Auto-tuning on Numerical Libraries and Advanced Computer Systems: Part II of II, Atlanta, Georgia, USA, Mar. 2008.
- [発表 9] Tetsuro Ogi and Hanxiong Chen, University of Tsukuba, Japan; Takashi Furumura, University of Tokyo, Japan; Shoji Itoh, RIKEN Institute of Physical and Chemical Research, Japan; Takahiro Katagiri, University of Tokyo, Japan; Atsuyuki Morishima, University of Tsukuba, Japan; Kengo Nakajima, University of Tokyo, Japan; Osamu Tatebe: VINDAM: Data Visualization and Data Mining in Tele-immersion Environment, University of Tsukuba, Japan, 13th SIAM Conference on Parallel Processing for Scientific Computing (PP08), Mini Symposium, MS55: Towards Visual Data-mining of Ultra Scale Data-set under Distributed Environments, Atlanta, Georgia, USA, Mar. 2008.

特記事項

- [特記 1] スーパーコンピュータ若手利用者推薦(試行)制度の設立, 2007年10月.
- [特記 2] 平成19年度日本学術振興会外国人招へい研究者(短期)採択, 米国ローレンスバークレー国立研究所 Osni A. Marques 博士を情報基盤センターJSPS 研究員として招へい, 2008年2月-2008年4月.
- [特記 3] スーパーコンピューティングニュース特集号, 特集: コーディングしてみよう! ~スパコンプログラミングを極める I~, 編集担当教員, スーパーコンピューティングニュース, Vol.9, No.Special Issue 1, 2008年2月.
- [特記 4] 東京大学情報基盤センタースーパーコンピューティング部門主催, Advanced Supercomputing Environment (ASE)研究会の設立および ASE 研究会幹事, 第1回研究会の開催, 2008年3月.
- [特記 5] お試しアカウント付き並列プログラミング講習会(試行)の設立, 2008年3月.
- [特記 6] ミニシンポジウム・オーガナイザー, 13th SIAM Conference on Parallel Processing for Scientific Computing (PP08), Mini Symposium, MS25: Auto-tuning on Numerical Libraries and Advanced Computer Systems: Part I and II, Atlanta, Georgia, USA, Mar. 2008.
- [特記 7] マルチコアプログラミングコンテスト Cell スピードチャレンジ 2008, 実行委員, 問題作成委員長, 2007年-2008年.

報道関連

- [報道 1] Microsoft INNOVATION AWARD 2007, アカデミック部門, 最優秀賞受賞報道, マイクロソフト, NIKKEI NET, IT-PLUS(NIKKEI NET), ITpro, livedoor ニュース, ITmedia エンタープライズ, 2007年9月.

タンパク質量子化学計算支援統合環境の研究開発

吉廣 保

1 概要

本研究の内容は平成 19 年 9 月まで東京大学生産技術研究所の産学官連携研究員として研究を行っていたものである。平成 19 年 10 月から情報基盤センターの特任准教授として、文部科学省の先端研究施設共用イノベーション創出事業に携わってきた内容については後半で報告する。

「革新的シミュレーションソフトウェアの研究開発 (RSS21)」プロジェクトの「創薬・バイオ新基盤技術開発へ向けたタンパク質反応全電子シミュレーション・システム」グループでは、タンパク質のための密度汎関数法プログラム ProteinDF を基にタンパク質の統合的な量子化学計算シミュレーションシステムの開発を行ってきた。本システムは「計算エンジン ProteinDF」を中心に「自動計算法」、「構造最適化・ab initio 分子動力学(MD)」などのプログラムで構成されており、これらを統括するグラフィカル・ユーザ・インターフェース(GUI)が「ProteinEditor」である。現在では ProteinDF によるタンパク質の基底状態全電子計算達成をサポートする機能をほぼ全て装備している。

2 タンパク質量子化学計算支援統合環境 ProteinEditor の開発

2.1 背景

一般に、タンパク質のシミュレーションには初期構造に Protein Data Bank (PDB) で配布されている 3 次元構造データを用いることが多い。PDB に登録されている立体構造は X 線構造解析や多次元 NMR などで求められたものであるが、その 8 割以上が X 線構造解析によるデータで、水素原子の座標データを持っていない。さらに、化学的な見地からは異常な構造歪みや原子間衝突などを残したまま構造精密化が行われていることも多く、これらは量子化学計算には不向きな構造といえる。適切なタンパク質の計算構造を構築することは、今やタンパク質量子化学計算にとって最も重要なプロセスの 1 つであると言っても過言ではない。そこで ProteinEditor に、PDB データへの水素付加やアミノ酸置換、モデリング支援機能、MD と連動させて構造歪みを緩和する機能、および構造妥当性のチェックを支援する機能などが必要となる。

また、密度汎関数法などの分子軌道法計算では SCF と呼ばれる繰り返し計算で解を収束させるが、タンパク質のような巨大で複雑な系では、非常に良い初期値から計算を始めないと発散してしまうか正しい解に収束しない。そこで本研究グループでは、良い初期値を作成する方法として擬カノニカル局在化軌道 (Quasi-Canonical Localized Orbital: QCLO) 法を提案した。QCLO とは、ある領域 (フラグメント) に局在化しているがその領域内ではカノニカルとなる軌道である。QCLO 法は小さなペプチド鎖の QCLO をつなぎ合わせてより大きなペプチド鎖の高品位な初期値を作成し、計算分子 (フレーム分子) を徐々に伸長してタンパク質の全電子計算を達成する方法である。

伸長の諸段階をステップと呼ぶ。QCLO 法では計算分子の伸長の仕方、すなわちフレーム分子の定義やステップ数に任意性があり、これをシナリオと呼んでいる。計算者は安全に全電子計算が達成できるシナリオを決める必要がある。シミュレーションが大規模かつ複雑になるにつれて、煩雑さやミスが増すとともに、より高度な専門知識が要求されるようになる。これらを軽減することも ProteinEditor 開発目的の一つである。

2.2 内容

ProteinEditor は量子化学計算に関わる研究者のみならず、生物物理学や生化学の研究者を想定して開発され、現在では ProteinDF による複雑なタンパク質の全電子計算達成をサポートする統合環境としての機能をほぼ全て装備している(図 1)。量子化学計算シミュレーションシステムを統括する機能に加えてグラフィカルな編集機能、各種の機能表現に必要なタンパク質のための大規模分子グラフィックス、およびその GUI で構成されている。開発中の GUI は Windows に対応し、分子グラフィックスは OpenGL を用いて描画している。その中でも、Protein Data Bank (PDB)が配信する標準タンパク質立体構造データを表示し、解析する機能は充実している。ProteinEditor には、Ramachandran プロット機能、結合長・結合角・二面角の分布表示機能、異常原子間距離表示機能、アミノ酸置換(ミューテーション)機能、簡易 MD 組み込みによる構造緩和機能、MD アニメーション表示機能などが組み込まれており、処々の機能は様々な 3 次元グラフィックスと連動しているため、異常な構造がチェックしやすい。また、局所的な分子構造の編集のためにタンパク質モデリング支援機能を追加した(図 2)。タンパク質を表示した後、異常原子間距離を発見した場所などを選択し、新たな編集ウィンドウを開くことで局所的な修正を加えることが可能である。また生産技術研究所選定研究の支援を受け、タンパク質モデリング支援機能に Virtual Reality(VR)技術を取り入れた。

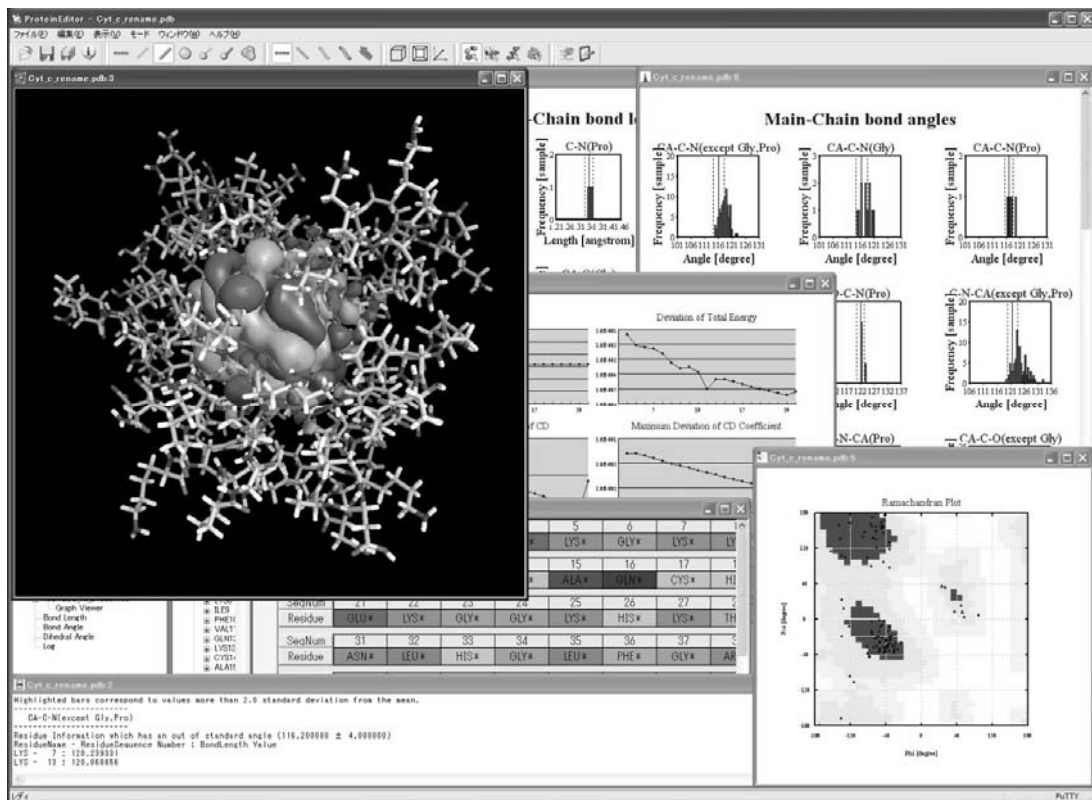


図 1 ProteinEditor のスナップショット

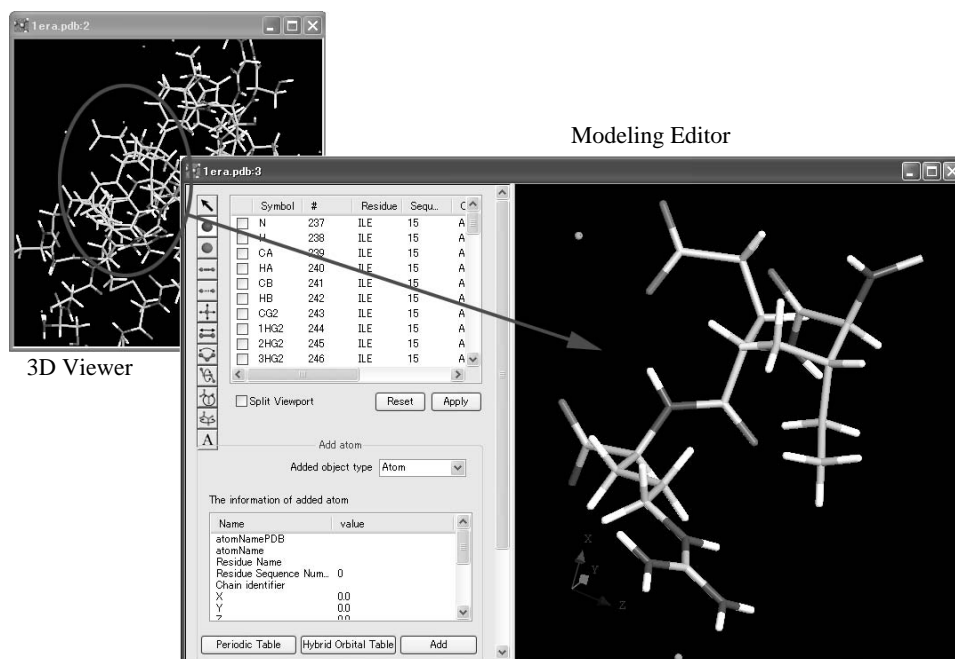


図 2 タンパク質モデリング支援機能
選択した構造を抜き出し、Modeling Editor で編集

2.3 具体的成果

タンパク質の密度汎関数法による全電子計算をサポートする新しい GUI、ProteinEditor の開発に成功した。本タンパク質反応全電子シミュレーションシステムでは解析する対象は動的となり、データ量も物理量の種類も格段に増える。独自の経験と技術をつぎ込んできた ProteinEditor をベースに、本システムの GUI へと発展させることが最適な開発方法である。薬剤設計のための構造モデリング機能を強化し、データベースとのリンクや新しい物性量を表現する手段を開発することによって、より高度な GUI へと成長させた。

a) モジュールの連携操作

これまでに ProteinEditor は 100 残基規模の一点全電子計算をサポートできるように開発してきた。今年度は計算規模を増すとともに、新しい自動計算法を導入し、ProteinEditor と連携して動作させた。これにより、ProteinEditor で作成した計算構造と自動計算シナリオを直接計算サーバーへジョブ投入することが可能となった。また、その計算結果も逐次計算サーバーにアクセスして取り出すことも可能となった。

さらに ProteinDF システムでは、タンパク質全電子計算を行うための計算構造として PDB(Protein Data Bank)データを用いているが、wwPDB のデータ修正プロジェクトにより IUPAC(国際純正・応用化学連合)命名法に従うフォーマットに修正されたため、ProteinEditor ではこの IUPAC 命名法に基づく PDB データ読み込みにも対応した。

b) タンパク質構造モデリング機能を強化

タンパク質のシミュレーションにおいてタンパク質立体構造の決定は重要で、計算結果に大きく影響を与える場合が多い。また、タンパク質はやわらかく、他の分子と相互作用する場合には変形することが多い。タンパク質の精細でフレキシブルな立体構造を作成するためには、細部にわたって構造の編集ができ、しかもその操作は分かりやすいものである必要がある。そこで今年度は、これまで ProteinEditor に実装されていた水素付加やアミノ酸置換といった機能の高品位化とともに、タンパク質構造のためのモデリング機能 Modeling Editor の機能を強化した。図 3 に Modeling Editor の操作画面を示す。プロテアーゼインヒビターの 35 残基目のアミノ酸残基の編集を行っている。

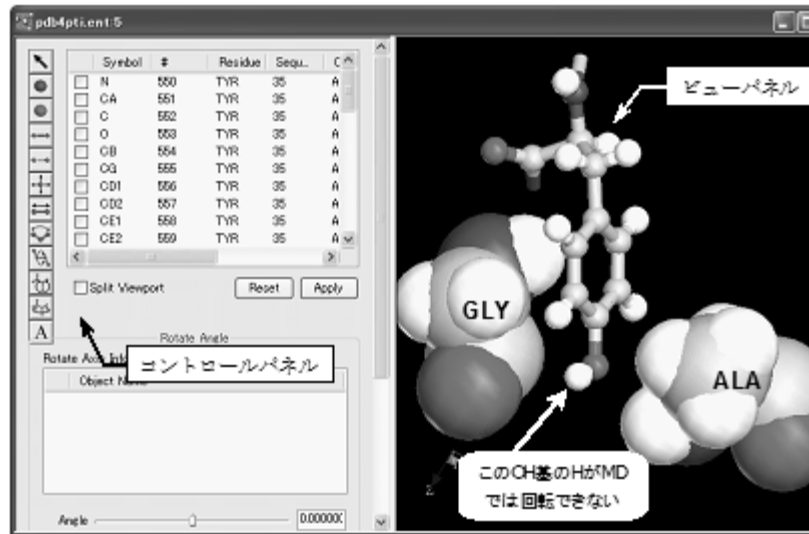


図3 Modeling Editor のスナップショット (TYR35 of Proteinase/Inhibitor: 4TPI)

左側(コントロールパネル): 構造編集機能のすべてを操作可能

右側(ビューパネル): 3次元構造の表示と編集操作が可能

この Modeling Editor では、分子グラフィックスで選択された構造に対してダミーチャージの追加、原子の追加と削除や原子位置、結合距離、結合角、二面角などの変更、また Modeling Editor 内で選択された部分構造に対して点中心の回転、軸中心の回転など分子グラフィックスからインタラクティブに、原子(単独および特定原子種集団)、基、側鎖、アミノ酸残基(群)、基質、配位子といった様々な単位で構造を編集することが可能となった。これに伴い、従来、分子動力学法や構造最適化法で局所解に落ち込み除去できなかった図3のような構造ひずみを対話的、かつ視覚的に編集することが可能になった。

また、タンパク質の構造編集後には、構造精密化を適切に実行することが必要である。加えて、ユーザが独自に力場を指定して局所部分構造最適化を行う手法も簡便かつ有効である。そこで構造妥当性評価機能を用いて評価された構造歪を除去するために、簡易部分構造最適化機能も実装した。この構造妥当性評価機能と簡易部分構造最適化は連携して動作させることができ、ユーザが直感的に構造編集を行えるようになった。

c) 反応シミュレーション解析のための新しい表現方法

反応シミュレーション解析では、全く新しいポスト処理が必要である。そこで今年度はタンパク質のような巨大で複雑な分子系で有効な表現方法を研究し実装した。タンパク質の反応シミュレーションでは解析の対象が動的になるとともに、データ量も物理量の種類も格段に増える。そこで、まず反応過程における分子軌道の変化をストレスなく観察できる表現手法を研究し、データ読み込みに多少時間がかかるものの、分子構造の時間変化はストレスなく表示が可能となった。

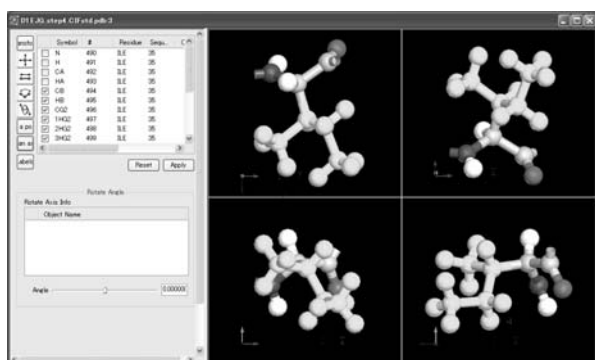
また、今年度はタンパク質表面のポケットの表示方法についても継続して研究を行った。タンパク質の重要な反応の一つに酵素反応がある。タンパク質(酵素)の表面や内部にはポケットと呼ばれるくぼみや空洞(鍵穴)があり、作用を受ける基質分子(鍵)はポケットに結合して、適切な化学変化を受ける。一般に、この活性部位の立体構造は、鍵穴と鍵の関係のように特定の基質とぴったり合うようになっており、酵素は基質の立体構造を認識することができる分子といえる。このドッキングのシミュレーション(Virtual Screening)を行う上で、ポケットの認識は非常に重要なステップである。ProteinEditor では、ポケットの検出方法としてよく使われている POCKET の派生である LIGSITE のアルゴリズムに基づき、ポケットの検出、及び表示を非常に高速に行うことができるようになった。

d) タンパク質モデリング VR システムの拡張(本研究の一部は東大生研・選定研究による)

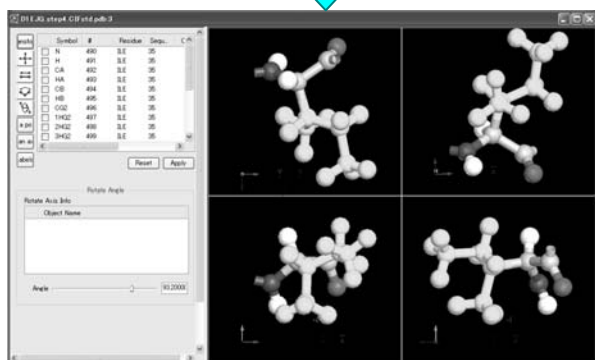
モデリング機能をベースに Virtual Reality (VR) 技術を取り入れ、複雑なタンパク質立体構造の 3 次元空間における原子位置関係と原子間に生じる力の場を正確に把握し、対話的な構造(原子、アミノ酸残基、基質、配位子)の編集を支援する機能として生産技術研究の選定研究により「タンパク質モデリング VR システム」の開発を行った。

従来の分子モデリングでは、奥行き感があまり得られない 2 次元的な表示と平面的な動きしか表すことの出来ないマウスを用いることが多かったため、編集する構造の位置を直感的に決定することは困難であった。また数十原子以下の一般分子ではその必要性も低かった。数千原子からなる巨大で複雑なタンパク質の密構造を編集する場合には、構造の 3 次元空間的位置関係や力を直感的に把握し、最適な状態へと構造を変形させる操作の機能開発が必要である。

本システムは反力デバイス PHANTOM による触感インターフェース、および裸眼立体視ディスプレイによる視覚インターフェースを取り入れた。モデリングは反力デバイスを使用してタンパク質の異常構造を変異・修正させる(図 4)が、その時に修正部位にかかる力を MD で計算し、デバイスにフィードバックさせることによって触感を通じてタンパク質が作る場を認識することを目指し開発を進めている。また、分子グラフィックスの立体視表示とこの触感によって 3 次元構造を正確に認識できると考えられるので、直感的かつ精密に構造を修正することができることが期待される。なお、このシステムにおいて対話的なモデリングを実現するには、タンパク質の力計算はリアルタイム性が重要となるため、MD シミュレーション専用計算機 MD-Grape3 の組み込みも行った。裸眼立体視の表示は、SHARP 製 3D 液晶を搭載したノート PC もしくは液晶ディスプレイで表示が行える。また、SGI 製の立体視ライブラリ ISL に対応したハードウェアでの立体視にも対応した。



構造編集前



構造編集後



- 3次元空間認識支援機能
- 3次元空間位置の直感的な認識が可能。
- 編集対象の構造とその周りの構造との衝突を判定。
- 編集支援機能
- 視覚とPHANTOMIによる触覚により原子間衝突の判別が可能。
- ペンの移動、回転(ロー/ヨー/ピッチ)に同期した構造の編集。

図 4 タンパク質モデリング VR システムの操作例

3 文部科学省 先端研究施設共用イノベーション創出事業【産業戦略利用】 先端的大規模計算シミュレーションプログラム利用サービス

3.1 背景

「先端研究施設共用イノベーション創出事業」は、大学、独立行政法人等の研究機関が有する先端的な研究施設・機器について、広範な分野における幅広い利用(共用)を促進し、イノベーションにつながる成果を創出するために、平成 19 年度から文部科学省が新たに開始した委託事業である。

本事業を通じて、産学官の研究者による戦略的かつ効率的な研究開発や、研究機関・研究分野を越えた横断的な研究開発活動を推進することにより、継続的に産学官の知の融合によるイノベーションを加速していくことを目指している。

全国共同利用施設であるスーパーコンピュータを有する7大学(北海道大学、東北大学、東京大学、名古屋大学、京都大学、大阪大学、九州大学)は、社会貢献の一環として、大学で開発された応用ソフトウェアとスーパーコンピュータの、民間企業への提供を実施している。7大学が有するスーパーコンピュータの総計算能力(理論性能値)は、平成 20 年度には約 330TFlops となった。

本事業の特徴は、各大学が開発した応用ソフトウェア(シミュレーションソフトウェア等)と各大学が持つ国内有数のスーパーコンピュータを、それらの利用支援サービスと一体化して民間企業に提供することにある。このようにサービス、ソフトウェア、ハードウェアの3つを同時に支援する試みは、国内外を通じて初めてのものである。この事業の目的は、大学の持つ知と施設によって我が国の経済発展に貢献することである。

3.2 内容

施設共用技術指導研究員として、「革新的シミュレーションソフトウェアの研究開発(RSS21)」で開発されたソフトウェアを7大学の計算機へインストール・整備・サポートを行う。また、RSS21 ソフトウェアのチューニングを行うための非常勤研究員と連携してチューニング作業を行う。ASP 事業を考えている企業に関しての全般的なサポートを非常勤研究員とともに行う。

施設利用の高度化に関する研究を行うことが許可されたため、グラフィカル・ユーザ・インタフェース(GUI)による操作性向上に関する研究に取り組む。

3.3 具体的成果

RSS21 ソフトウェアが東京大学情報基盤センターの SR11000/J2 と平成 20 年度から利用開始となる T2K オープンスパコン上で実行できるように移植作業を行った。現在、SR11000/J2 では ABINIT-MP、FrontFlow/blue、Phase、FrontSTR などが移植完了しており、チューニングも進んでいる。T2K オープンスパコンに向けての移植は、AMD Opteron のクラスタでの実行を確認しており、SR11000/J2 と同じプログラムの動作確認を完了している。また、本事業で第3者へアプリケーションの利用を提供する上で、RSS21 ソフトウェアのライセンスの確認などの作業も同時に行った。

GUI による操作性向上の取り組みを行うために、RSS21 で開発していた ProteinEditor を基に、他のプログラムの操作もこの GUI も統合できないか検討を行っている。また Windows クライアントから SSH を用いたスーパーコンピュータへのアクセスも、同じ GUI から操作できるような方法を開発できないかを調査中である。

事業としては、平成 19 年 8 月 22 日から 9 月 14 日までで平成 19 年度第 1 期公募を行い、19 件(うち東京大学は 5 件)の採択課題が 10 月 1 日より計算機利用を開始した。また、10 月 15 日より 11 月 30 日まで第 2 期公募を行い、2 件(うち東京大学は 0 件)の採択課題が 12 月 17 日より計算機利用を開始した。

平成 19 年度第 1 期採択企業のうち 3 社が ASP 事業を考えており、それらの企業と ASP を実現するために必要な技術的な問題の解決を行うためのミーティングを月に 1 度程度行った。

4 成果要覧

招待講演／招待論文

[招待 1] 佐藤文俊, 稲葉亨, 井原直樹, 恒川直樹, 西野典子, 西村康幸, 平野敏行, 吉廣保, 柏木浩: ProteinDF システムによるタンパク質電子構造解析, 第 10 回理論化学討論会, 2007 年 5 月.

著書／編集

[著書 1] 佐藤文俊, 恒川直樹, 吉廣保, 平野敏行, 井原直樹, 柏木浩: タンパク質密度汎関数法, 柏木浩, 佐藤文俊編, 森北出版, 東京, in press (平成 20 年 5 月中旬出版予定).

[著書 2] 佐藤文俊他: 生体分子量子化学計算 第 1 部, 森北出版, 東京, in press.

査読付論文

[査読付 1] 西村康幸, 吉廣保, 佐藤文俊: タンパク質全電子計算のための GUI・ProteinEditor, 生産研究, 59, pp93-100, 2007 年.

[査読付 2] Toru Inaba, Naoki Tsunekawa, Toshiyuki Hirano, Tamotsu Yoshihiro, Hiroshi Kashiwagi, Fumitoshi Sato: Density functional calculation of the electronic structure on insulin hexamer, Chem. Phys. Lett., 434, pp331-335, 2007.

公開ソフトウェア

[公開 1] 佐藤文俊他: 大規模タンパク質の量子化学計算システム (ProteinDF System Ver.1.4), ProteinDF システム用 GUI プログラム (ProteinEditor Ver.1.4), http://www.ciiss.iis.u-tokyo.ac.jp/rss21/result/download/index.php#download_1

その他の発表論文

[発表 1] 石川寛人, 西村康幸, 吉廣保, 佐藤文俊: タンパク質構造変動に伴うポケット形状変化の動的表示, 第 10 回理論化学討論会, 2007 年 5 月.

[発表 2] 西村康幸, 吉廣保, 井原直樹, 佐藤文俊: タンパク質量子化学計算のための分子モデリング環境の構築, 第 1 回分子科学討論会 2007, 2007 年 9 月.

[発表 3] 西村康幸, 吉廣保, 井原直樹, 佐藤文俊: Development of the molecular modeling environment for quantum chemical calculations of proteins, CBI 学会 2007 年大会, 2007 年 10 月.

[発表 4] 石川寛人, 西村康幸, 吉廣保, 佐藤文俊: Study on Dynamic Representation of Pocket Shape According to the Protein Structural Change, CBI 学会 2007 年大会, 2007 年 10 月.

[発表 5] 吉廣保, 西村康幸, 佐藤文俊: タンパク質量子化学計算支援統合環境 ProteinEditor の開発, 日本コンピュータ化学会 2007 秋季年会, 2007 年 10 月.

[発表 6] Fumitoshi Sato, Toru Inaba, Noriko Nishino-Uemura, Yasuyuki Nishimura, Tamotsu Yoshihiro, Naoki Tsunekawa, Naoki Ihara, Toshiyuki Hirano: Development of Quantum Chemical Simulation System for Proteins; ProteinDF, GCOE: International Symposium on Secure-Life Electronics - Advanced Electronics for Quality Life and Society -, 2008.

PKI プロジェクト

西村 健

大学におけるPKI の整備と既存システムとの連携

大学における PKI の整備と既存システムとの連携

西村 健

1 概要

PKI(Public Key Infrastructure, 公開鍵認証基盤)は技術的には成熟した分野であるが、現状では実用面において普及しているとは言い難い。これはある程度大規模な組織においては PKI の構築自体がコストがかかるものであることが大きな理由である [発表 1]。また、既存の非 PKI のシステムでは PKI 対応に必要なコストが膨大になり、応用できる分野が限られ PKI によるメリットを享受しにくいということも理由として考えられる。

本研究では、それぞれの問題に対してより具体的な対象を選択し、解決方法を提案している。

- 大学における登録局構築の課題
大学内での部局の独立性を考慮した低コストな PKI の構築、中でも登録局の構築に焦点をあてて論じる。
- PKI を応用した Personal Space Server の提案
PKI の応用範囲を広げるため、公開鍵の特性を利用した personal space server を提案する。さらに personal space server を利用した PKI 認証を従来の ID/パスワード認証と連携させるシステムを提案する。

2 大学における登録局構築の課題

2.1 背景

Web においてやりとりする情報には機密性の高いものがあり、それらは通常 SSL (Secure Sockets Layer) / TLS (Transport Layer Security) プロトコルを通して暗号化される。これはクライアント(ブラウザ)とサーバとの間の暗号化であり、サーバの認証のためにサーバ証明書というデータが使用される。サーバ証明書は PKI (Public Key Infrastructure) という枠組みの上で認証局によって発行される証明書の一種であり、認証局はサーバの実在性や発行申請者の本人性など適切な審査ののちに、サーバのドメイン名とサーバ自身の持つ公開鍵をペアにしたものに対して認証局私有鍵によるデジタル署名を施したサーバ証明書を発行する。ブラウザが認証局を信頼する、つまり認証局公開鍵を持っているという前提の下で、サーバ証明書をを用いることにより公開鍵暗号技術を用いた厳格な認証が行なえる。

PKI においては上記の審査の部分が大きなウェイトを持ち、この部分を効率化することは大きな問題となっている。本研究では特にサーバ証明書を取り上げ、大学の特殊事情も考慮した登録局の構築方法を提案する。東京大学のような総合大学においては、学部・研究科等の部局の独立性が高いこと、また部局内においても専攻等の下部組織の独立性が高い場合があることから、中央集権的な登録局を構築することは困難であるし、対応のための人員も不足する。

以下では認証局を、審査を行なう機関である登録局と発行を行なう機関である発行局に分けて論じる。

2.2 内容

大学におけるサーバ証明書の問題であるが、Web の歴史的経緯も含め、大学のサーバ管理は、特に研究室レベルのサーバの場合であれば、ルーズになされている場合が未だに存在する。例えば、教職員以上に知識を持つ学生が管理しているとか、研究目的のものを流用しているとか、助教・助手が仕事で手一杯なので学生が管理を手伝う、などである。特に管理者が学生の場合、何かあった場合の責任をとることができないので問題である。

このような責任の所在が不明確で管理体制があやふやなサーバにまで証明書を発行しては、SSL の信頼性崩壊と同様、サーバ証明書の信頼性が崩壊してしまう可能性がある。そのようなサーバがサーバ証明書を必要とし申請を行うかという問題もあるが、今回の場合証明書を無償で発行するということもあり可能性は高いと考えられた。

加えて、東京大学のような総合大学においては、部局の独立性が高く、また部局内においても専攻等の下部組織の独立性が高い場合があることから、サーバの所属は部局単位もしくはより細かい単位であることが必要である。部局毎に意思が異なる、つまり独立の法人のような立場であるので、仮にあるサーバから情報が発信されるとして、その情報源のサーバの所属は〇〇大学とするよりも〇〇学部、さらには〇〇研究室とするほうが誤解が少ない。部局まで明示するという事は、保証レベルを上げる、つまりより確かな保証を与えるサービスとなる。

具体的な登録局の構成は、階層的な構造となる。便宜的に中央の組織を TRA と呼び、下位組織を TLRA と呼ぶ。TRA は下記条件に合致しない証明書発行申請(直接申請)の審査も行なうが、条件が合えば一部の審査権限を部局単位の下位組織に委譲する。部局単位の発行審査組織が TLRA である。

審査は対面での確認を原則としており、証明書発行直接申請時および TLRA 設置申請時には、教職員による本人確認、および部局内でのドメイン管理体制を示した文書、TLRA 設置申請時には加えて申請者の本人確認方法を示した文書を提出してもらうことにより審査を行なう。

サーバ証明書の Organizational Unit Name (OU)に部局の英語名称を記載し、審査の上で発行証明書に記載する。これによって発行された証明書をインストールしたサーバの利用者は、部局名を確認しより確かな保証を受けることができる。

2.3 具体的成果

実際に東京大学におけるサーバ証明書発行のために階層的登録局を構築した。その詳細および発生した問題、課題等を [発表 2] において発表している。

これまで扱ってきたのはサーバ証明書に関する登録局だったが、ベリサインのマネージド PKI のように学内で登録局を持つ運用は今後増えていくと思われ、これまで得られた様々な知見とともに他の個人証明書への応用を予定している。

3 PKI を応用した Personal Space Server の提案

3.1 背景

近年様々な分野でウェブによるオンラインサービスが提供されているが、機密情報を扱う場面での利用は進んでいない。これはサービスを提供する側、特にそのサーバ管理者が不正にデータを利用できる環境にあることが大きいと思われる。加えてマーケティングやパーソナライゼーションのためにデータを利用することを明言されている場合もある。サービス利用者はそのサービスを利用するかしないかの二者択一を迫られることになる。サーバ上に保管されたデータのアクセス権を設定するのはそのサーバ管理者であると言える。本研究では、アクセス権を利用者の自由に設定できる personal space server を提案する。

3.2 内容

PKI には利用者が自由に自身が保持するデータを暗号化できるという特徴があり、この特徴を利用する。personal space server は、personal space と呼ばれる各利用者固有の領域を提供するサーバで、personal space には利用者以外、つまりサーバ管理者ですら自由にアクセスできない。この実装には利用者側での暗号化が有効であり、将来各利用者が持つであろうと想定される PKI の鍵ペアを用いることを想定している。

この研究は安全な後述するパスワードサーバの実現による PKI の利用促進のみならず、既存のオンラインサービスに personal space server の枠組みを適用することによって、従来オンラインサービスとして扱うことが難しかった機密情報を扱う分野にまで拡張することが可能となる。

3.3 具体的成果

personal space server の実現例として、personal space server をパスワードサーバに適用し、サーバ管理者ですら不正が不可能なパスワードサーバ LESSO を開発し、Internet Explorer および Firefox 上で動作するプロトタイプを実装した。

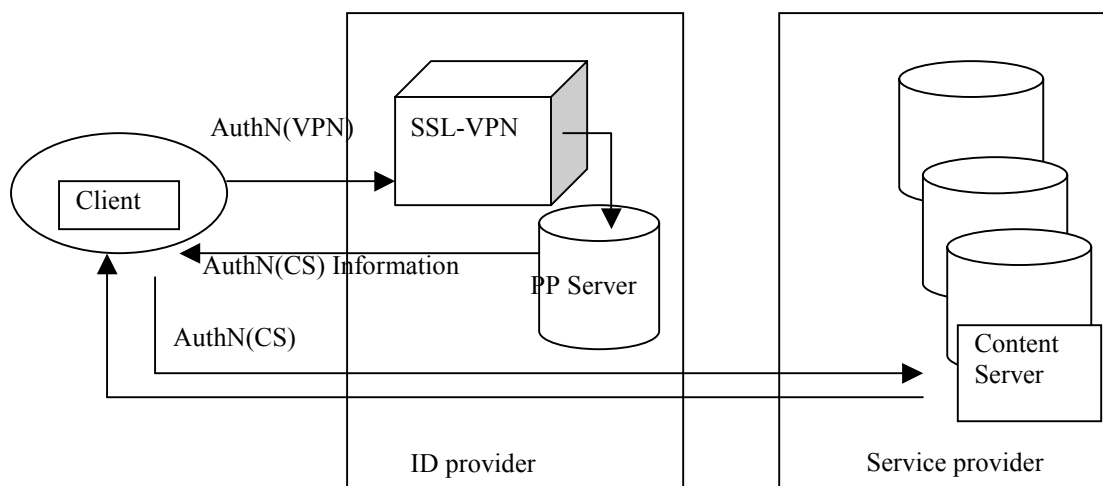


図 1: LESSO の構成

LESSO は既製の SSL-VPN 製品によって通信の安全性を担保しつつ認証において PKI 認証を用いたゲートウェイシステムである。このゲートウェイがクライアントとコンテンツサーバ群との仲介者となり、クライアントからの PKI 認証の後にゲートウェイが各サーバに対する ID およびパスワードを送出する。このシステムによってサーバを PKI 認証のために改変することなく、PKI による安全性を享受することができる。構成図を図 1 に示す。

Personal space server としてのポイントは、password pushdout server (PP server) である。PP server において利用者のパスワードを保持するが、これは利用者の鍵により暗号化され、サーバ管理者であっても中身を覗き見ることはできない。使用時は、PP server から暗号化されたパスワードを取得後、クライアント上で復号した上でサーバへ送付する。

システムが統一的でない大学のような組織で特に顕著であると思われるが、大小様々な形態で存在している既存システムと PKI との連携は容易ではない。連携を行ないそのメリットを享受するためには多大なコストが必要であるため、LESSO のような簡易的なシステムによりそれが実現できることが実証されれば PKI 普及を促進するものと考えられる。

現在は LESSO の通信路上およびコンピュータ上での攻撃に対する安全性の検証を行ない、従来の手法との比較を行なっている。

4 成果要覧

その他の発表論文

[発表 1] 西村健, 佐藤周行, “東京大学における認証局 UT-CA の全学展開に向けた取り組み,” ITRC 第 21 回インターネット技術第 163 委員会研究会, 2007 年 5 月.

[発表 2] 西村健, 佐藤周行, “東京大学におけるサーバ証明書発行体制の構築と課題,” 第 48 回分散システム/インターネット運用技術・第 26 回高品質インターネット合同研究発表会(2008-DSM-48), pp79-84, 2008 年 3 月.

PART 3

教育・サービス活動

情報メディア教育部門

図書館電子化部門

キャンパスネットワークング部門

(全国共同利用)

スーパーコンピューティング部門

PKI プロジェクト

情報メディア教育部門



新装の情報基盤センター1階大演習室(1)と教育用計算機システム標準端末

情報メディア教育部門 概要

柴山悦哉

平野光敏

情報メディア教育部門の主要なミッションは、教育研究目的で利用可能なさまざまな機器、ソフトウェア、サービスなどを、東京大学に所属する学生および教職員に対して提供することである。そして、これに付随する管理運用、ユーザサポート、システム設計、研究開発などの作業も行っている。現在提供中のシステムやサービスには以下のようなものがある。これらは、研究室、学科、部局等で個別に運用管理を行うことが困難であったり、大学全体で共有して管理運用を集約化の方が費用面等で有利であったりするものである。

教育用計算機システム (ECCS)：本郷、駒場、柏の3キャンパスに分散配置した1,176台のiMac端末と186台のWindows端末を約26,000名の登録ユーザに提供している。これらの端末には、Officeアプリケーションや数式処理システムなどの商用アプリケーションと多数のフリーソフトウェアが搭載されており、授業での利用をはじめとした教育目的を中心にさまざまな用途に利用されている。一般ユーザが直接触れることはないが、これらを一元管理するために、ブートサーバ、ファイルサーバなどを用いた集中管理体制を取っている。端末の他にも、プリントサービス、メールサービス、遠隔ログインサービスなどを各ユーザに対して提供している。また、新規利用者向け講習会などのソフトなサービスにも力を入れている。

メールホスティング：学内のドメインを単位としたメールホスティングサービスを提供しており、300以上のドメインに所属する約11,000名のユーザが利用している。ECCSユーザ向けのメールサービスと合わせると、延ユーザ数は約37,000名に達する。

Webホスティング：学内の500以上のドメインに対して、Webのホスティングサービスを提供している。

認証サービス：ECCSの登録ユーザ情報を活用した認証サービスも一部で行っており、無線LANの試行サービス、SSL-VPNなどがECCSのユーザ情報を利用している。さらに一部の部局に対しては、部局所有の端末にECCSの認証情報を提供するサービスも行っている。

ストリーミングサービス：映像配信用のストリーミングサーバを利用したサービスの提供を行っており、UTOCW, TODAI TV, 卒業式中継をはじめさまざまな用途に利用されている。また、映像作成のための機材の提供も行っている。

CFIVE：講義支援のためのオープンソースシステムとしてCFIVEの開発と運用を行っており、学生証を用いた出欠管理、課題の出題とレポートの提出、教材のオンライン配信などを支援する機能を提供している。CFIVEは、昨年度実績で100以上の講義で利用されている。

遠隔講義支援：テレビ会議システムを祖合えた遠隔講義室、多地点接続装置などを用意しており、遠隔講義や遠隔会議に利用されている。

教育用計算機システム運用報告

情報リテラシー教育支援係
情報教育支援係

1 ECCS2004 の運用

1.1 Mac(iMac, Xserve)

2008 年 2 月までの間は旧システム(ECCS2004)を前年度に引き続き運用した。ECCS2004 は標準端末として iMac 端末(1,149 台)を設置していた。iMac 端末はハードディスクを搭載せず、ブートサーバ(Apple Xserve)から OS(Mac OS X)のイメージを読み込んで起動する。1 台のブートサーバは iMac 端末 20 数台を制御し、49 台(駒場 34 台、本郷 15 台)のサーバを設置していた。また、RDC(Remote Desktop Connection)を使用して、Windows Server 2003 を利用可能であった。

システムイメージの更新は OS のバージョンアップ、セキュリティアップデート、アプリケーションのインストールやアップデートに伴って行っていた。

2007 年度の主な更新は以下の通りである。なお、不具合の修正や広報の表示の際にもイメージの更新を行うことがあった。

2007 年 4 月 16 日 iMac 端末のシステム更新

- Mac OS X 10.4.9
- Microsoft Office 2004 for Mac 11.3.4
- Adobe Reader 8
- Firefox 2.0.0.3
- iPod アップデータ 20060628
- Java for Mac OS X 10.4 リリース 5
- Opera 9.10
- Quicktime 7.1.5
- Thunderbird 1.5.0.10
- JMP 6.0.3
- MATLAB 7.3
- STATA/SE 9.2
- Firefox に live http headers プロトコルを追加
- 「Mozilla」を削除

2007 年 6 月 1 日 iMac 端末のシステム更新

- Mac OS X 10.4.9 Security Update 2007-005
- Microsoft Office 2004 for Mac 11.3.5
- Opera 9.21
- Quicktime 7.1.6
- Thunderbird 2.0
- Microsoft Office Open XML File Format Converter for Mac 0.1b
- nVIDIA Cg Toolkit 1.5
- 「Internet Explorer」を削除

2007 年 7 月 24 日 iMac 端末のシステム更新

- Mac OS X 10.4.10
- Security Update 2007-006
- Adobe Reader 8.1.0

- Firefox 2.0.0.4
- iTunes 7.3.1
- Microsoft Office 2004 for Mac 11.3.6
- Quicktime 7.2
- R 2.5.0
- Thunderbird 2.0.0.4
- VLC 0.8.6c
- Flip4Mac を追加
- Microsoft Office Open XML File Format Converter for Mac 0.1b を削除

2007年12月11日 iMac 端末のシステム更新

- Mac OS X 10.4.11
- Security Update 2007-007
- Adobe Reader 8.1.1
- Adobe Flash Player 9.0.47.0
- Firefox 2.0.0.11
- Flip4Mac 2.1.3.10
- iTunes 7.5
- Microsoft Office 2004 for Mac 11.3.9
- Opera 9.24
- Quicktime 7.3
- R 2.6.1
- Thunderbird 2.0.0.9
- Xcode 2.4.1
- SciFinder Scholar を追加

1.2 VID(Windows, Linux)

VID 端末は、Windows や Linux でしか動作しないアプリケーションを必要とする講義などで用いることを想定したディスクレス端末である。また、CAD などを使用するために大きなメモリ容量や高精細な画面が必要なアプリケーションを動かすための VID+端末もあり、本郷・駒場・柏に合計約 250 台設置していた。

Windows において、ディスクイメージファイルの更新時には OS のセキュリティアップデートとウイルス対策ソフトのパターンファイル更新を必須に行い、その他にソフトウェアの更新などを行った。

2007 年度の主な更新は以下の通りである。

2007年5月29日 VID 端末(Windows)のソフトウェア更新

- Windows Update / Office Update (2007/05)
- Windows Update / Office Update (2007/04)
- J2SE Runtime Environment 6 Update 1 (1.6.0_01)
- Flash Player 9.0.45.0
- VIRUS BUSTER 2006 update
- Ox 4.10
- Opera 9.21
- Office2007 ファイル形式用 Microsoft Office 互換機能パック
- 印刷時に「両面印刷」を指定できるようにした

2007年6月27日 VID 端末(Windows)のソフトウェア更新

- Windows Update / Office Update (2007/06)
- Firefox 1.5.0.12

- Thunderbird 1.5.0.12
- Windows Media Player 10
- VIRUS BUSTER 2006 定義ファイル等 update

2007年7月31日 VID 端末(Windows)のソフトウェア更新

- Windows Update / Office Update (2007/07)
- Firefox 2.0.0.5
- Opera9.22
- J2SE Runtime Environment 6 Update 2 (1.6.0_02)
- Flash Player 9.0.47.0
- VIRUS BUSTER 2006 定義ファイル等 update

2007年8月27日 VID 端末(Windows)のソフトウェア更新

- Windows Update / Office Update (2007/08)
- Firefox 2.0.0.6
- Opera9.23
- VIRUS BUSTER 2006 パターンファイル等 update

2007年10月24日 VID 端末(Windows)のソフトウェア更新

- Windows Update / Office Update (2007/10)
- Windows Update / Office Update (2007/09)
- Firefox 2.0.0.8
- Java 6 Update 3 (1.6.0_03)
- DocumentSkipper ブラウザ 3.00
- VIRUS BUSTER 2006 パターンファイル等 update

2008年1月31日 VID 端末(Windows)のソフトウェア更新

- Windows Update / Office Update (2008/01)
- Windows Update / Office Update (2007/12)
- Windows Update / Office Update (2007/11)
- Firefox 2.0.0.11
- Thunderbird 1.5.0.14
- Flash Player 9.0.115.0
- Opera 9.25
- VIRUS BUSTER 2006 パターンファイル等更新

2 新しい教育用計算機システム(ECCS2008)について

2.1 システム更新の経緯

2008年3月に教育用計算機システムの更新を行った。以下にその経緯と新計算機システム(ECCS2008)の概要について述べる。

(仕様策定)

情報教育専門委員会の下に「ECCS 導入作業部会」が組織され、2006年12月13日、2007年1月17日に同部会を開き、センターホームページで集計した「ECCS2004に関するアンケート」の結果、および2006年9月28日、29日に開催した公聴会での意見等を参考に、次期システム案の大枠について検討した。

(資料提供招請の公示)

2006年2月16日 資料提供招請の官報公示をした。

(導入説明書の作成)

ECCS 導入作業部会での検討結果を受け、導入説明書を作成した。

(導入説明会)

2月28日 導入説明会を開催した。

(仕様策定委員会)

教育用計算機システム仕様策定委員会を設置して、4月17日、4月27日、5月21日、6月18日、6月25日、7月2日の6回開催し、次期システムの仕様書を作成した。

(仕様書案の説明会)

6月18日 仕様書案について説明会を行った。

(入札の公示)

7月26日 入札の官報公示をした。

(入札説明会)

8月3日 次期システムの仕様書の説明、及び事務手続きの説明を行った。

(入札)

9月18日 入札締切

(技術審査委員会)

技術審査委員会が9月19日、9月25日に行われ、入札に参加した1社の提案を合格とした。

(開札)

10月1日 日本電気株式会社(NEC)が落札した。

(納入)

2月28日 納入期限

2008年3月1日から新教育用計算機システム(ECCS2008)によるサービスを開始した。

2.2 ECCS2008 システム説明会

教育用計算機システムを利用して授業を行なう教官、ティーチングアシスタントと分散配置端末管理・担当責任者を対象に、新システムの利用説明会を本郷・駒場の両会場で開催した。

駒場：3月19日(水) 15:10～17:00 情報教育棟3階大演習室2

本郷：3月21日(金) 13:30～15:20 情報基盤センター1階大演習室2

また、この説明会では教育用計算機システムの一部として運用している学習管理システム CFIVEの説明も行った。

2.3 新システム(ECCS2008)の概要

新しくなった教育用計算機システム(ECCS2008)の概要について紹介する。

【端末】

○標準端末(1176台)

- ・ 情報処理教育及びプログラミング教育用
- ・ 20インチ iMac
- ・ 1680 x 1050ドット(フルカラー)20インチ液晶ディスプレイ
- ・ 2.0GHz Intel Core 2 Duo
- ・ メモリ 2GB
- ・ Mac OS X 10.5, NetBoot によりブートサーバ(Xserve、端末30数台につき1サーバ)からOSイメージをロード
- ・ 仮想化ソフトウェア(Parallels Desktop for Mac)により、Windows Vista が利用可能
- ・ DVD+R DL、DVD±RW、CD-RW 対応光学ドライブ内蔵
- ・ Happy Hacking Keyboard Lite2 for Mac(JIS配列、カーソルキーあり)
- ・ 3ボタンレーザーマウス(うち1ボタンはホイール機能と兼用)

○CAD 端末(186 台)

- ・ CAD 講義用
- ・ Express 5800/51Le
- ・ 内蔵ディスクをキャッシュとして用いるシンククライアントシステム(Ardence)
- ・ 1680 x 1050ドット(フルカラー)20 インチ液晶ディスプレイ
- ・ 2.4GHz Intel Core2 Duo
- ・ メモリ 2GB
- ・ DVD-R、CD-RW 対応光学ドライブ内蔵
- ・ テンキー付小型キーボード(JIS 配列、カーソルキーあり)
- ・ 3 ボタンレーザーマウス(うち 1 ボタンはホイール機能と兼用)

ECCS2004 との相違

○標準端末

- ・ CPU 速度が単体で 2 倍以上, コア数が 2 に
- ・ メモリ容量が 4 倍に
- ・ 画面サイズが 15 インチから 20 インチワイドに
- ・ ネットワークインタフェースが 100Base-TX から 1000Base-T に
- ・ 光学ドライブが DVD+R DL、DVD±RW 対応に
- ・ マウスがレーザーマウスに
- ・ Windows 環境がローカル CPU で提供される

○CAD 端末

- ・ CPU が Pentium4 2.66 GHz から Core2 Duo 2.4GHz に(クロックは下がっているが単体性能は上、コア数が 2 に)
- ・ メモリ容量は 1.5GB→2GB
- ・ ネットワークインタフェースが 100Base-TX から 1000Base-T に
- ・ 全台に DVD-R、CD-RW 対応光学ドライブ内蔵(ECCS2004 は数台に 1 台の割合で DVD-ROM、CD-RW 対応光学ドライブ内蔵)

○端末全般

- ・ 台数減少
 - + ECCS2004 ではレンタル 1376 台 + 買い取り 58 台
 - + 新システムでは 1362 台
 - + 予算減少率より減少率を小さく押さえる

【ファイルサーバ】

- NEC iStorage NV7400

ECCS2004 との相違

- 総容量は, ECCS2004 の約 3 倍
- 利用の少なくライセンス料が高かった WebDAV サービスは停止

【Unix サーバ】

- Apple Xserve 3 台
- 外部から ssh で入って利用
- CUI アプリケーション及び X window アプリケーションのみ利用可能
- 一般の MacOS アプリケーションは利用不可

ECCS2004 との相違

- 台数が 4 台から 3 台に減少
- Solaris サーバは利用が少ないので廃止

【メールサーバ】

- Express 5800/120Rg-1
- DEEPSoft MailSuite 3.5(DEEPMail + SPAMBlock + Symantec AntiVirus/AntiSpam)
- ECCS2004 同様の IMAP(S), POP(S), Web-mail の運用
- 新規メールホスティングサービス申し込み分と共有
- 既存メールホスティングサービス用は別サーバ

ECCS2004 との相違

- ディスク容量は ECCS2004 の約 3 倍
- WWW インタフェースが Active Mail!から DEEPMail に変更

【ネットワークデバイス】

- コアスイッチは NEC BigIron RX-4/8(Foundry 社 OEM)
- エッジスイッチは D-Link DGS-3100
- 本郷・駒場間は WDM 1Gbps
- 情報基盤センター・本郷キャンパス間は 10Gbps 接続
- Juniper SSG 550 を用いた NAT によるインターネット接続

ECCS2004 との相違

- 末端までギガビット化
- WWW proxy を使用しないように変更(ログは従来同様に記録)

【ソフトウェア】

- ユーザー認証管理サーバー: NEC EDS(Enterprise Directory Server)
- プログラミング言語環境
 - ・ GCC, Java 言語環境 が MacOS 環境で利用可能
- 統計処理ソフトウェア
 - ・ SAS が Windows 環境で利用可能
 - ・ JMP が Windows、MacOS 環境で利用可能(別調達)
- 数式処理ソフトウェア
 - ・ Mathematica が MacOS, Windows 環境で利用可能
- CAD ソフトウェア
 - ・ AutoCAD Inventor、3ds max が Windows 環境で利用可能(別調達)
- オフィスアプリケーション
 - ・ MS-Office 2008 for Mac が標準端末 Mac OS 環境で利用可能
 - ・ MS-Office 2007 Enterprise 2007 が CAD 端末で利用可能
- ラスタ描画アプリケーション
 - ・ Photoshop Elements が Mac OS 環境で利用可能
- Unix フリーソフト
 - ・ MacOS 環境で、可能な限り現状と同様の Unix フリーソフトを提供

ECCS2004 との相違

- Mac 環境での MS-Office の互換性が向上したため、標準端末の Windows 環境には MS-Office はインストールしない
- PDF 作成やページ統合などが OS の標準機能で可能なので、Adobe Acrobat はインストールしない。

【プリンタ】

- リコー統合印刷管理システム「Ridoc IOGate」
- A4 カラープリンタ(IPSiO CX3500) 36 台、A3 カラープリンタ(IPSiO SP C810) 5 台
- ユーザの自己負担によるプリペイドカード課金に対応したシステム

ECCS2004 との相違

- 全台をカラープリンタに

3 稼働状況

3.1 部局別実利用者数

教育用計算機システムを一度でも利用したことのある学部学生、大学院生、研究生、教職員の月毎累積の部局別実利用者数である。比較のため、2007 年度(各月右側)と 2006 年度(同左側)を掲せてある。

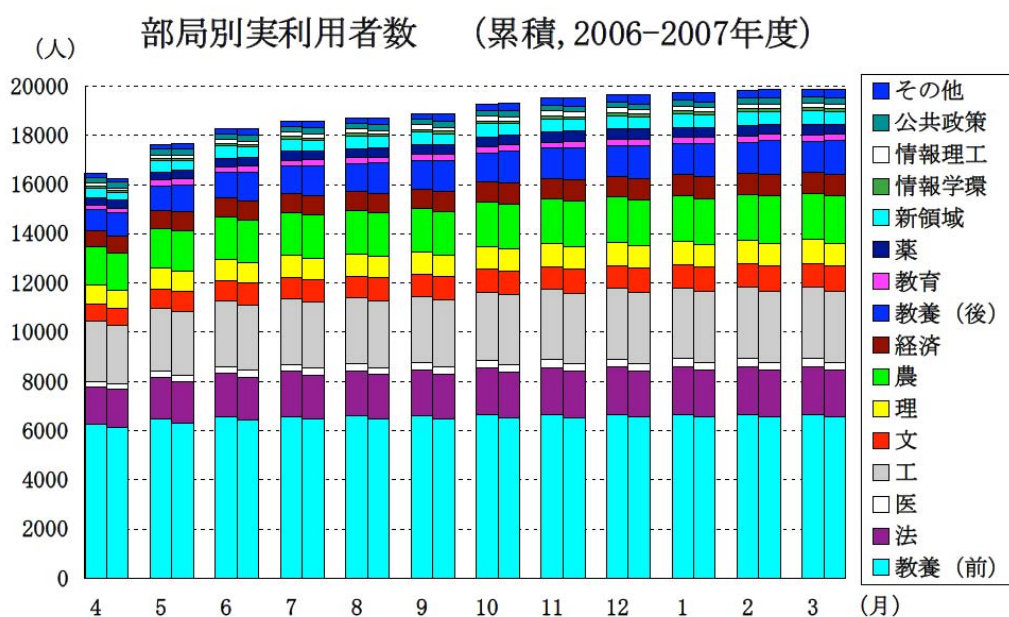


図 1. 部局別実利用者数

3.2 端末利用状況

情報メディア教育部門では、教育用計算機システムを管理しており、その端末等を本郷キャンパスと駒場キャンパスに配置している。本郷では、情報基盤センター(浅野地区)の演習室に端末等を集中配置すると共に、総合図書館、留学生センター、各学部／研究科等の30数箇所にも端末を分散配置している。駒場では、教養学部情報教育棟に端末等を集中配置すると共に教養学部図書館、数理科学研究科にも端末を分散配置している。

これらの配置場所のうち、利用の多い本郷の総合図書館と駒場の情報教育棟と駒場の教養学部図書館の iMac 端末(ECCS2004)利用状況は図2の通りである。(下から、総合図書館(2階メディアプラザ2と同3階メディアプラザ3の合計(51台))、駒場教養学部図書館2階メディアパークプラザ(35台)、情報教育棟3階大演習室2(195台)、同2階大演習室1(195台)、1階自習室(123台)の2008年1月28日におけるものである。)

2008年1月28日 iMac 端末(ECCS2004)利用状況

各部屋の時間帯ごとの端末利用数のグラフである。

縦軸は端末数で横軸は時刻(6時から24時まで)である。

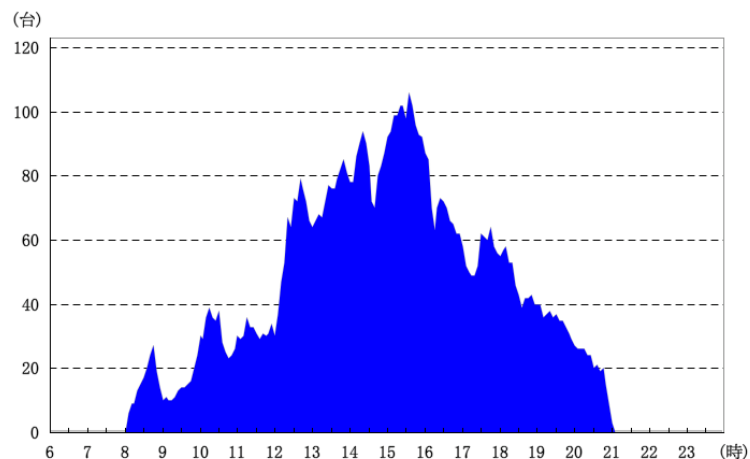


図2. 駒場情報教育棟 1階自習室 (123台)

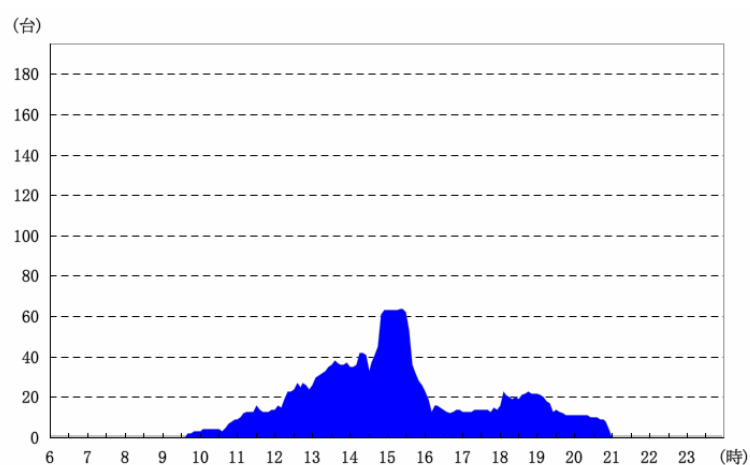


図3. 駒場情報教育棟 2階大演習室1 (195台)

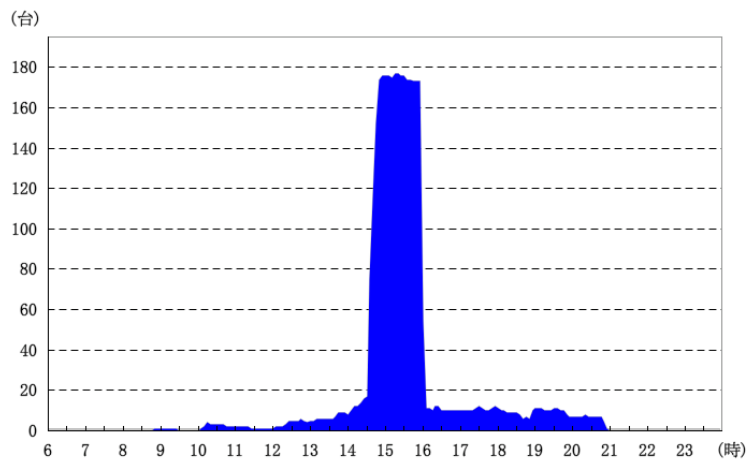


図 4. 騎場情報教育棟 3階大演習室 2 (195 台)

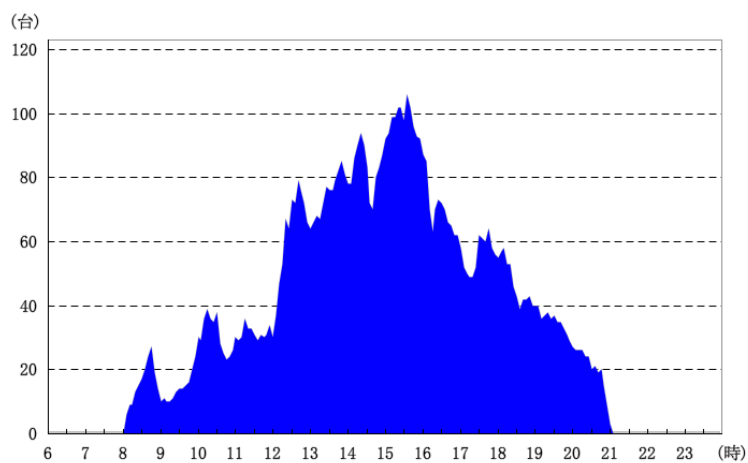


図 5. 騎場図書館 2F-メディアパークプラザ (35 台)

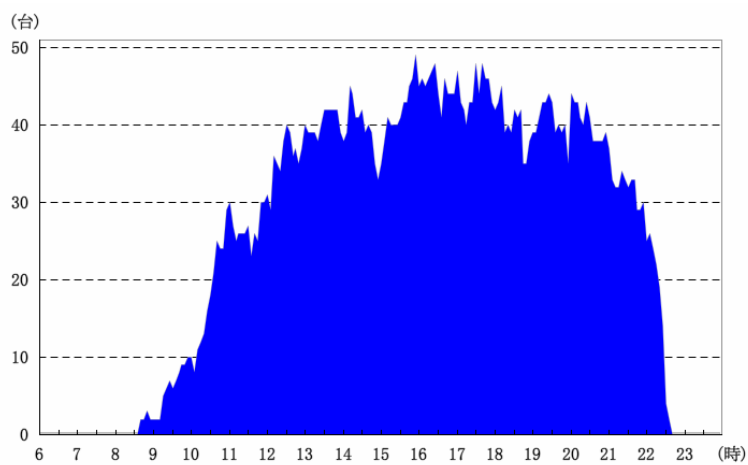


図 6. 総合図書館 2F-メディアプラザ 2, 3F-メディアプラザ 3 (51 台)

4 柏出張講習会

学部学生、大学院生および研究生が教育用計算機システムの利用を新規に申し込む際には、新規利用者向け講習会を受講する必要がある。しかし、柏キャンパスに所属する学部学生、大学院生については、本郷への移動は時間がかかるため、受講者の負担が大きく、また、遠隔キャンパスとしては所属人数が多いという事情を考慮して、年に1回程度の頻度で出張講習会を開催している。

2007年度の実施状況は以下の通りである。

- 4月19日(木) 講習会開催および利用申込書受付
 - 新規利用申込者 103名
 - 継続利用申込者 134名
- 4月26日(木) パスワード配付

学内組織向けメールサーバ(MAILHOSTING)運用報告

情報教育支援係

情報リテラシー教育支援係

1 経過

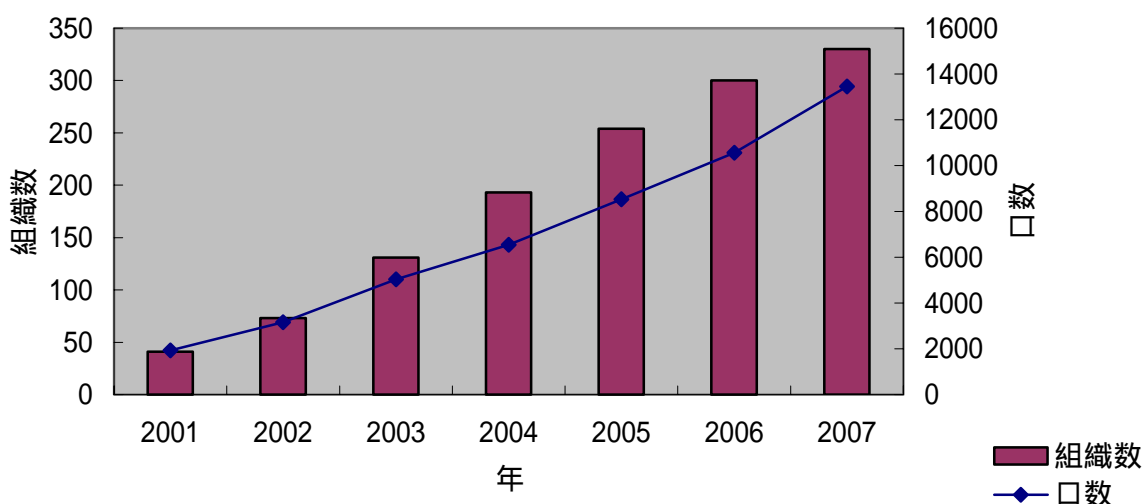
2007年6月29日 第27回情報メディア教育専門委員会において、利用負担金免除制度の適用継続を希望していた1組織に対し、制度の適用継続が認められた。

2008年2月新メールホスティングサーバへ移行した。

2 参加組織

平成19年度末現在、本サービスを利用している組織は330組織であり、それらの申込み口数の合計は、13450口である。平成18年度末に比べ、30組織、2855口増加している。

メールホスティング参加組織の推移



3 広報

本年度に行った主な広報は、以下のとおりである。

- ・ 4/3(火)メールサーバの障害について 更新:2007年04月03日
- ・ 緊急メンテナンスによるメールサーバの停止について(終了) 更新:2007年04月26日
- ・ メール等が接続しにくい障害について(終了) 更新:2007年06月29日
- ・ メール配送障害のお詫び 更新:2007年08月23日
- ・ Administration Suite の迷惑メールフィルタ設定方法 更新:2007年08月30日
- ・ 迷惑メールの対処方法について 更新:2007年10月26日
- ・ メールの受信に関する障害について 更新:2007年11月01日
- ・ ディスク資源の確保の為不要メール整理のお願い。 更新:2007年11月19日
- ・ 送信時の認証(SMTP AUTH)を使うためのメーラーの設定変更方法(Microsoft Outlook 2003) 更新:2007年12月05日

- ・メールシステムの障害について 更新:2007年12月09日
- ・メール転送設定でメッセージの保存を行う時の注意事項 更新:2007年12月14日
- ・xGateによるサービスの終了について 更新:2007年12月18日
- ・送信時の認証(SMTP AUTH)を使うためのメーラーの設定変更方法(Winbiff) 更新:2007年12月18日
- ・12/25(火)～12/27(木) メーリングリスト(Mailing List/ML)サーバの障害について 更新:2007年12月28日
- ・12/19～20 迷惑メールの誤判定によるメールの遅配のお詫び 更新:2007年12月28日
- ・パスワードに使用できる文字種・文字数の制限について 更新:2008年01月23日
- ・学内向けメール中継設定の変更について (追記) 更新:2008年02月01日
- ・login before smtpによるメール送信サービス停止について 更新:2008年02月01日
- ・Active!mailの使用について 更新:2008年02月07日
- ・メールシステムの移行作業における障害について 更新:2008年02月13日
- ・メールの送信について 更新:2008年02月14日
- ・NetscapeのSMTP AUTHに関するバグについて 更新:2008年02月14日
- ・サービス > Mailer 更新:2008年02月15日
- ・送信時の認証(SMTP AUTH)を使うためのメーラーの設定変更方法(Outlook Express) 更新:2008年02月15日
- ・送信時の認証(SMTP AUTH)を使うためのメーラーの設定変更方法(Microsoft Outlook 2007) 更新:2008年02月15日
- ・移行期間中のメール受信障害について 更新:2008年02月17日
- ・移行作業後に一部ユーザの認証が失敗する不具合 更新:2008年02月18日
- ・ECCS2004方式でのアカウント管理インタフェースの停止について 更新:2008年02月20日
- ・本来の運用形態のためのネットワーク機器切り替え作業 更新:2008年02月21日
- ・SMTP AUTHでメールが送信できない不具合 更新:2008年02月25日
- ・メールホスティングシステム新規受付の再開 更新:2008年03月05日
- ・Mac OS 9環境からのメールアクセスについて 更新:2008年03月06日
- ・パスワード付き zip ファイルが添付されたメールの取り扱いについて 更新:2008年03月10日
- ・Active!mailのログイン障害について(復旧) 更新:2008年03月12日
- ・SMTP AUTHの障害に関するお詫び 更新:2008年03月19日
- ・学内組織向けメールサーバ (MailHosting) のシステムの移行について 更新:2008年03月19日
- ・MailHosting MとMailHosting Dの違いは何ですか? 更新:2008年03月25日
- ・SPAMBlockによる迷惑メール判定 更新:2008年03月26日
- ・学内組織向けメールサーバの利用負担金について 更新:2008年03月30日
- ・3/26(Wed) ネットワーク障害によるシステム停止について 更新:2008年03月31日

WEB PARK サービス運用報告(2007年4月—2008年3月)

電子教材係

1 経過

2007/06/29(金) 第27回情報メディア教育専門委員会において、利用負担金免除制度の適用継続を希望していた5組織に対し、制度の適用継続が認められた。

2007/09/30(日)8:00～12:00 電源設備の定期点検のためサービスを休止した。

2007/12/01(土)8:30～18:00 電源設備工事のためサービスを休止した。

2008/03/27(木)8:30～18:00 電源設備工事のためサービスを休止した。

2 利用組織数

2007年度に利用を開始した組織数は93、利用を中止した組織数は13であった。

2008年3月末現在の総利用組織数は518である。

3 WEB PARK サービスを利用している学内組織一覧(2008年3月末現在)

(各カテゴリーにおいては申し込み順、各組織の名称は2008年3月末現在)

センター、委員会等：95組織

アジア生物資源環境研究センター(8組織)

- アジア生物資源環境研究センター (anesc)
- 地域資源評価研究分野 (horiken)
- 生物環境評価大部門 土地環境評価研究分野 福代研究室 (dinos)
- 樹木生理学・熱帯造林学研究室 (tree)
- 耐性機能開発研究分野 白子研究室 (anescmv)
- 土地環境評価研究分野 環境材料設計学研究室 (smd)
- 環境ストレス耐性機能研究室 (lestm)
- 共生研究室 (symbio)

遺伝子実験施設 (mgrl)

総合文化研究科／教養学部(14組織)

- AIKOM委員会 (aikom)
- 進学情報センター (agc)
- 学生相談所 (soudanjo-komaba)
- 情報教育棟 (i-edu)
- 駒場図書館 (lib-c)
- ドイツ・ヨーロッパ研究センター (desk)
- 文化施設運営委員会オルガン委員会 (organ)
- 「人間の安全保障プログラム」運営委員会 (human_security)
- 教養教育開発室 (komed)
- 広域科学専攻 広報委員会 (kouiki)
- 社会連携委員会 (friday-lecture)
- アメリカ太平洋地域研究センター (cpas)
- 美術博物館 (museum)
- 生命科学構造化センター (csls)

先端科学技術研究センター (14 組織)

- フォトニクス材料分野(宮野研究室) (myn)
- 知的財産権大部門 (ip)
- 知的財産マネジメント研究会 (smips)
- 藤井研究室 (findata)
- 御厨貴研究室 (mikuriya)
- バリアフリー分野 (bfp)
- 馬場研究室 (zzz)
- 研究戦略・社会システム大部門 資源・エネルギー・環境政策 (wlab)
- 化学認識機能材料分野 (crm)
- 菅研究室 (cbl)
- 渡辺研究室 (fennelweb)
- 藤井研究室 (finlab)
- NEDO 新環境エネルギー科学創成特別部門 (nedo)
- 香川研究室 (cm)

海洋研究所 (17 組織)

- 海洋研究所 (ori)
- 海洋生命科学部門 行動生態計測 (fishecol)
- 新プログラム「海洋生命系のダイナミクス」 (dobis)
- 海洋生態系動態部門 (ori-ecosystem)
- 海洋物理学部門 (ori-dpo)
- 先端海洋システム研究センター 海洋システム計測分野 (ori-cer)
- 海洋化学部門 (ori-co)
- 観測研究企画室 (ori-tso)
- 海洋生物資源部門 (ori-lmr)
- 海洋生命科学部門 (ori-bio)
- 国際沿岸海洋研究センター (icrc)
- 先端海洋システム研究センター (ori-camr)
- 海洋生命科学部門 生理学分野 (physiol)
- 海洋研究所図書室 (ori-library)
- 海洋科学国際共同研究センター 企画情報分野 (ori-cicplan)
- 先端海洋システム研究センター 海洋システム解析分野 (ori-msa)
- 海洋研究連携分野「生物圏環境学」 (ori-mbe)

低温センター (crc)

情報基盤センター (6 組織)

- 情報メディア教育部門 (itc-media)
- キャンパスネットワーク研究部門 若原研究室 (wakahara)
- 情報メディア教育部門 電子教材係 (elearn)
- 図書館電子化部門 (dl-itc)
- 片桐研究室 (katagiri)
- 小川研究室 (ogawa)

東京大学バリアフリー支援室 (spds)

生物生産工学研究センター (2 組織)

- 生物生産工学研究センター (biotec-res-ctr)
- 植物機能工学研究室 (ppk)

情報倫理審査専門委員会 (cie)

産学連携本部 (ducr)

埋蔵文化財調査室 (aru)
保健センター (health)
UT-CERT (ut-cert)
医療ナノテクノロジー人材養成ユニット (NBEP)
留学生センター (ic)
社会科学研究所 (1 組織)
 • 図書室 (iss-lib)
大学総合教育研究センター (1 組織)
 • 中原研究室 (jun)
大学総合教育研究センター／教養教育開発機構(学術俯瞰講義) (gfk)
科学技術インタープリター養成プログラム (STITP)
ナノバイオインテグレーション研究拠点 (CNBI)
生命科学教育支援ネットワーク (lse)
北京代表所 (beijing-office)
柏地区事務部 渉外・広報室 (utkk)
サステイナビリティ学連携研究機構 (ir3s)
情報の知識化・価値化技術協議会 (chishiki)
AGS 推進室 Intensive Program on Sustainability (ipos)
生物機能制御化合物ライブラリー機構 (cbri)
地球観測データ統融合連携研究機構 (editoria)
海洋アライアンス (OA)
武田先端知クリーンルーム (nanotechnet)
教養学部附属 教養教育開発機構／大学総合教育研究センター／情報学環
 駒場アクティブラーニングジオスタジオ (Kals)
知の構造化センター(ネットワーク) (cks)
環境安全研究センター (5 組織)
 • 環境安全研究センター (esc)
 • 戸野倉研究室 (tonokura)
 • 新井研究室内 大学等環境安全協議会 (daikankyo)
 • 刈間研究室 (karima-lab)
 • 尾張研究室 (owari)
総括プロジェクト機構 (2 組織)
 • ジェロントロジー寄付研究部門 (gerontology)
 • 知的資産経営研究寄付講座 (iam)

学部／研究科, 学科／専攻, 研究室等 : 511 組織

法学系研究科／法学部 (2 組織)
 • 久保研究室 (kubo)
 • 法学政治学研究科 道垣内弘人研究室 (hdogauchi)
医学系研究科／医学部 (26 組織)
 • 機能生物学専攻 統合生理学研究室 (iphysiol)
 • 病因病理学専攻 免疫学研究室 (immunol)
 • 健康科学・看護学専攻 成人看護学／緩和ケア看護学分野 (adng)
 • 国際保健学専攻 (sih)
 • 公共健康医学専攻 生物統計学 (epistat)
 • 健康科学・看護学専攻 地域看護学分野 (chn)

- 健康科学・看護学科 (hn)
 - 国際地域保健学教室 (ich)
 - 病因病理学専攻 微生物学講座 (microbiology)
 - 健康科学・看護学専攻 健康社会学分野 (hlthsoc)
 - 公共健康医学専攻 老年社会学分野 (sg)
 - 健康科学・看護学専攻 母性看護学・助産学分野 (midwifery)
 - COE 機能生物学・細胞分子生物学 (coem)
 - 公衆衛生学教室 (public health)
 - 健康科学・看護学専攻 看護体系・機能学分野／看護管理学分野(基礎看護学教室) (nurs-adm)
 - 国際保健学専攻 人類生態学分野 (humeco)
 - 健康科学・看護学専攻 老年看護学／創傷看護学分野 (rounenkango)
 - 機能生物学専攻 細胞分子生理学分野 (cmphysiology)
 - 健康科学・看護学専攻 家族看護学分野 (fn)
 - 分子予防医学教室 (macrophage)
 - 分子神経生物学 (mnp)
 - 国際保健学専攻 人類遺伝学教室 (humgenet)
 - 疾患生命工学センター 健康・環境医工学部門 (env-health)
 - 神経生化学 (neurochemistry)
 - 疾患生命工学センター 分子病態医科学 (tmlab)
 - システム疾患生命科学による先端医療技術開発拠点 (TSBMI)
- 工学系研究科／工学部 (146 組織)
- 航空宇宙工学専攻 (aero)
 - 航空宇宙工学専攻 藤本研究室 (crack)
 - 航空宇宙工学専攻 森下研究室 (tmorisi)
 - 航空宇宙工学専攻 ジェットエンジン研究室 (jetlab)
 - 物理工学専攻 鹿野田研究室 (kanoda_lab)
 - マテリアル工学系専攻 (material)
 - 材料学専攻 先端デバイス工学研究室 (adam)
 - 物理工学専攻 教務室 (ap_office)
 - 物理工学専攻 藤原研究室 (fujiwaralab)
 - 物理工学専攻 古澤研究室 (furusawa)
 - 物理工学専攻 五神研究室 (gono)
 - 物理工学専攻 十倉研究室 (cmr)
 - 物理工学専攻 初貝研究室 (ap_hatsugai)
 - 建築学専攻 (arch)
 - 機械系専攻 (microsystemlab)
 - 地球システム工学専攻 (geosys)
 - 地球システム工学専攻 物探・油層研究室 (gpre)
 - 計数工学科 (keisu)
 - 技術経営戦略学専攻 (knowledge)
 - 工学・情報理工学図書館 (cllib)
 - 附属総合研究機構 (sogo)
 - 電気工学科 山地・藤井研究室 (yamaji)
 - 電気工学専攻 横山研究室 (syl)
 - マテリアル工学系専攻 材料システム研究室 (mse)
 - 電子工学専攻 田中雅明研究室 (cryst)

- 都市工学科 都市デザイン研究室 (ud)
- マテリアル工学専攻 プラズマ工学研究室 (plasma)
- 社会基盤工学専攻 建設マネジメント・開発システム研究室 (cmis)
- マテリアル工学専攻 COE 推進委員会 (coe-material)
- 社会基盤学専攻 海岸・沿岸環境研究室 (coastal)
- 社会基盤学専攻 交通・都市・国土学研究室 (trip)
- 物理工学専攻 為ヶ井研究室 (tamegai_lab)
- 社会基盤学専攻 河川／流域環境研究室 (reel)
- 航空宇宙工学専攻 青木研究室 (aokilab)
- 社会基盤工学専攻 応用力学／岩盤力学研究室 (ohriki)
- 量子相エレクトロニクス研究センター 染谷研究室 (someya)
- 地球システム工学専攻 加藤研究室 (katolab)
- 量子相エレクトロニクス研究センター (qpec)
- 環境海洋工学専攻 (naoe)
- マテリアル工学科 山口研究室 (ssi-lab)
- 工学教育推進機構 国際化推進室(GWP) (gwp)
- 建築学専攻 環境系 (envarch)
- 環境海洋工学専攻 海洋機器研究室 (gisolab)
- 船舶海洋工学専攻 構造システム研究室 (nasl)
- 21 世紀 COE 強相関物理工学 (coe-ap)
- 物理工学専攻 (ap)
- マテリアル工学専攻 (metall)
- 航空宇宙工学専攻 河野－津江研究室 (kono-tsue_lab)
- 環境海洋工学専攻 安全評価工学研究室 (yuhalab)
- マテリアル工学専攻 マイクロフォトニクス研究室 (microphotonics)
- 物理工学科 香取研究室 (amo)
- 技術経営戦略学専攻 元橋研究室 (mo)
- システム創成学専攻 生産システム工学研究室 (msel)
- 物理工学科 樽茶研究室 (tarucha-lab)
- 物理工学専攻 永長研 CREST (nagaosa_crest)
- 環境海洋工学専攻 海洋空間計画研究室 (ospl)
- 化学システム工学科 山下研究室 (tcl)
- 精密機械工学専攻 (pe)
- 技術経営戦略学専攻 (biz-model)
- 精密機械工学専攻 小林研究室 (ein)
- 機械工学専攻 牛田・古川研究室 (tissue)
- 技術部 (ttc)
- 社会基盤学専攻 コンクリート研究室 (concrete)
- 化学生命工学専攻 (chembio)
- 総合研究機構 俯瞰工学 (metatechnica)
- 総合研究機構 強力 X 線実験室 (hpx)
- 精密機械工学専攻 新井・横井・太田研究室 (arai-lab)
- 社会基盤学専攻 上田研究室 (infra_economics)
- 環境海洋工学専攻 輸送・環境システム計画研究室 (suiso)
- 電子工学専攻 大津研究室 (uuu)
- 航空宇宙工学専攻 西成研究室 (tknishi)
- 機械工学専攻 ホロニックエネルギーシステム学講座 (hes)

- 総合研究機構 井上研究室 (ultracold)
- 総合研究機構 触媒反応工学領域 (nishiba)
- 地球システム工学専攻 藤田研究室 (tfujita-lab)
- 建築学専攻 岸田研究室 (design)
- システム創成学科 (si)
- 都市工学専攻 国際都市計画・地域計画研究室 (onishiken)
- 建築学専攻 難波研究室 (namba)
- マテリアル工学専攻 阿部研究室 (nanostructures)
- 社会基盤学専攻 橋梁研究室 (bridge)
- 応用化学専攻 平尾研究室 (qcl)
- 総合研究機構 石原研究室 (ishihara)
- 技術経営戦略学専攻 (TMI)
- バイオエンジニアリング専攻 (bioeng)
- 応用化学専攻 橋本研究室 (hashimoto_lab)
- マテリアル工学科 片岡研究室 (kklab)
- マテリアル工学専攻 吉田亮研究室 (yoshida-lab)
- 電子工学専攻 保立・山下・何研究室 (sagnac)
- 物理工学科 今田研究室 (imada)
- 都市工学専攻 関澤研究室 (fse)
- システム量子工学専攻 古田研究室 (cse)
- 物理工学専攻 求研究室 (motomelab)
- 航空宇宙工学専攻 岩崎研究室 (sal)
- 社会基盤学専攻 英語コミュニケーション教室 (english)
- 建築学専攻 野口研究室 (bmat)
- 都市工学科 都市交通研究室 (ut)
- システム創成学専攻 マリタイム・イノベーション(常石造船) 寄附講座 (mil_tsuneishi)
- 建築学専攻 伊藤研究室 (itolaboratory)
- 環境海洋工学専攻 設計研究室 (nakl)
- 1st Asian Biomaterials Congress (abc1st)
- 第 34 回 日本臨床バイオメカニクス学会 (biomech34)
- 地球システム工学専攻 登坂研究室 (tosaka-lab)
- 機械工学専攻 加藤研究室 (sstl)
- 化学システム工学専攻 土橋研究室 (dobashi)
- マテリアル工学専攻 複層鋼板プロジェクト (lism)
- 化学システム工学専攻 堂免・久保田研究室 (domen-lab)
- 化学システム工学専攻 大久保研究室 (zeolite)
- 化学システム工学専攻 (chemsys)
- システム量子工学専攻 大橋研究室 (ohashi)
- 化学システム工学専攻 越研究室 (koshi)
- 化学システム工学専攻 中尾研究室 (nakaolab)
- システム量子工学専攻 渡辺研究室 (fwlbg)
- 国際交流室 (oice)
- 建築学専攻 高田研究室 (tkdlab)
- 機械工学専攻 笠木・鈴木研究室 (thtlab)
- 機械工学専攻 鹿園研究室 (feslab)
- バイオエンジニアリング専攻 田畑研究室 (tabata-lab)
- 社会基盤学専攻 国際プロジェクト研究室 (intl-project)

- 建築学専攻 野口研究室 (noguchi)
- 総合研究機構 加藤研究室 (ykkato)
- 建築学専攻 松村・藤田研究室 (mf_lab)
- システム創成学科 B コース シミュレーション・数理社会デザイン (wisip)
- 精密機械工学専攻 医用精密工学研究室 (bmpe)
- 産業機械工学専攻 山田・ドロネー研究室 (lelab)
- 水環境制御研究センター (recwet)
- 都市工学専攻 都市環境工学コース (enveng)
- フロンティアエネルギー開発工学(JAPEX) 寄付講座 (fedl)
- 都市工学専攻 環境システム(花木)研究室 (esys)
- 工学系研究科 応用化学専攻藤田研究室 (fujitalab)
- 都市工学専攻 環境質リスク管理研究室 (envrisk)
- 都市工学専攻 都市環境工学講座 都市水システム研究室 (urbanwater)
- 物理工学専攻 押山研究室 (oshiyama)
- システム量子工学専攻 越塚研究室 (koshizukalab)
- 都市工学専攻 石川幹子研究室 (ishikawa)
- 機械工学専攻 リサイクルウェブ (recycle-w)
- 化学システム専攻 平尾研究室 (pse)
- 電子工学専攻 保立研究室 (hotalab)
- 電子工学専攻 山下研究室 (yamalab)
- 電子工学専攻 何研究室 (helab)
- バイオエンジニアリング専攻 古川研究室 文部科学省 科学技術振興調整費 重要課題解決型研究の推進プログラム (EMTE)
- 建築学専攻 前研究室 (labf)
- 物理工学専攻 就職幹事会 (ap-shushoku)
- 総合研究機構 ナノフォトニクス研究センター (npc)
- システム創成学専攻 (dsi)
- エネルギー・資源フロンティアセンター (frcer)

人文社会系研究科／文学部 (1 組織)

- 言語動態学研究室 (dootai)

理学系研究科／理学部 (5 組織)

- 技術部 (s-tech)
- 物理学専攻 藤森研究室 (atsushi)
- 物理学専攻 島野研究室 (shimanolab)
- 生物情報科学学部教育特別プログラム (upbsb)
- 物理学専攻 岡本研究室 (okamotolab)

農学生命科学研究科／農学部 (93 組織)

- 生産・環境生物学専攻 応用昆虫学研究室 (applant)
- 応用生命化学専攻 分析化学研究室 (bunseki)
- 応用生命化学専攻 生物機能開発化学研究室 (biofunc)
- 応用生命化学専攻 生物有機化学研究室 (seiyu)
- 応用生命化学専攻 土壌圏科学研究室 (soil-cosmology)
- 応用生命化学専攻 植物分子生理学研究室 (pmp)
- 応用生命化学専攻 植物栄養・肥料学研究室 (syokuei)
- 応用生命化学専攻 放射線植物生理学研究室 (radio-plantphys)
- 応用生命工学専攻 分子生命工学研究室 (molbiotech)
- 生物・環境工学専攻 生物機械工学研究室 (bme)

- 生物材料科学専攻 生物材料物理学研究室 (bmp)
- 森林科学専攻 (forest)
- 森林科学専攻 森林利用学研究室 (foresteng)
- 生圏システム学専攻 (es)
- 農業・資源経済学専攻 (arec)
- 水圏生物科学専攻 (fs)
- 附属牧場(高等動物教育研究センター) (bokujo)
- 放射性同位元素施設 (kanri)
- 農学生命科学図書館 (aglib)
- 獣医学専攻 (vetmed)
- 生物・環境工学専攻 農地環境工学研究室 (land)
- 水圏生物科学専攻 水産資源研究室 (shigen)
- 水圏生物科学専攻 水圏生物工学研究室 (suikou)
- 農業・資源経済学専攻 農業史研究室 (aghist)
- 森林科学専攻 森林風致計画学研究室 (fuuchi)
- 獣医学専攻 動物育種繁殖学研究室 (ikushu)
- 獣医学専攻 獣医解剖学研究室 (kaibo)
- 獣医学専攻 獣医生理学研究室 (seiri)
- 獣医学専攻 獣医薬理学研究室 (yakuri)
- 獣医学専攻 獣医微生物学研究室 (microbio)
- 獣医学専攻 獣医公衆衛生学研究室 (eisei)
- 獣医学専攻 比較病態生理学研究室 (hibyou)
- 獣医学専攻 獣医病理学研究室 (byouri)
- 獣医学専攻 獣医内科学研究室 (naika)
- 獣医学専攻 獣医外科学研究室 (geka)
- 獣医学専攻 実験動物学研究室 (jitsudo)
- 獣医学専攻 獣医臨床病理学研究室 (vcpb)
- 応用動物科学専攻 応用免疫学研究室 (immune)
- 応用動物科学専攻 応用遺伝学研究室 (iden)
- 応用動物科学専攻 動物細胞制御学研究室 (saibo)
- 応用動物科学専攻 細胞生化学研究室 (seika)
- 応用動物科学専攻 獣医動物行動学研究室 (koudou)
- 応用動物科学専攻 高度医療科学研究室 (koui)
- 生物・環境工学専攻 生物環境情報工学研究室 (beie)
- 森林科学専攻 森林植物学研究室 (fb)
- 応用生命化学専攻 有機化学研究室 (org-chem)
- 応用生命化学専攻 食品生化学研究室 (food-biochem)
- 生物・環境工学専攻 環境地水学研究室 (chisui)
- 応用生命工学専攻 細胞遺伝学研究室 (cell-gene)
- 水圏生物科学専攻 水圏生物環境学研究室 (fol)
- 応用生命工学専攻 醗酵学研究室 (hakko)
- バイオトロン研究室 (biotron)
- 農業・資源経済学専攻 経済学研究室 (agriecon)
- 森林科学専攻 造林学研究室 (silviculture)
- 生物・環境工学専攻 生物環境情報工学研究室 (joho)
- 応用生命工学専攻 微生物学研究室 (Lab_Microbiology)
- 食シグナル・生体統御系間相互作用(明治乳業) 寄付講座 (shoku-signal)

- 応用生命化学専攻 食糧化学研究室 (foodchem)
 - 応用生命化学専攻 生物化学 (inositol)
 - 生物・環境工学専攻 (bee)
 - 21世紀COEプログラム「生物多様性・生態系再生」研究拠点 (coe-ber)
 - 生圏システム学専攻 保全生態学研究室 (coneco)
 - 生産・環境生物学専攻 (aeb)
 - 生物・環境工学専攻 生物環境情報工学研究室 (agrinfo)
 - 生産・環境生物学専攻 植物分子遺伝学研究室 (pmg)
 - 付属家畜病院 (vmc)
 - 生物・環境工学専攻 生物プロセス工学研究室 (bipren)
 - 附属水産実験所 (suijitsu)
 - 生産・環境生物学専攻 園芸学研究室 (engei)
 - 生産・環境生物学専攻 植物分子遺伝学研究室 (ICPMB2007)
 - 電脳土壌センター (soilDB)
 - 農学国際専攻 (ga)
 - 農業・資源経済学専攻 食料・資源経済学研究室 (frec)
 - 産学官民連携型農学生命科学インキュベータ機構 (agri-cocoon)
 - 農学国際専攻 国際情報農学研究室 (iai)
 - 農学国際専攻 新機能植物開発学研究室 (pbt)
 - 森林科学専攻 森林経営学研究室 (forestmanag)
 - 農学国際専攻 国際森林環境学研究室 (gfes)
 - 森林科学専攻 森林動物学研究室 (zoology)
 - 応用生命化学専攻 栄養化学研究室 (ayo)
 - 農学国際専攻 国際水産開発学研究室 (lgfs)
 - バイオミネラルゼミ (biomineral)
 - 食の安全研究センター (frc)
 - 水圏生物科学専攻 魚病学研究室 (fishparasite)
 - 生物材料科学専攻 木材化学研究室 (woodchem)
 - 生圏システム学専攻 水域保全研究室 (suiiki)
 - 学術創成バイオミネラル (gakujutsubiom)
 - 水圏生物科学専攻 魚病学研究室 (jsfp2008)
 - 農業・資源経済学専攻 農政学研究室 (agripo)
 - 農学国際専攻 国際環境経済学研究室 (iee)
 - 生物・環境工学専攻 生物環境工学研究室 (kankyo)
 - 森林科学専攻 森林理水及び砂防工学研究室 (sabo)
 - 「共生社会基盤形成を通じた国土の保全管理に係る寄付講座」 (irrigationwater)
- 経済学研究科／経済学部 (2 組織)
- 都市経済学 金本良嗣研究室 (urban-eco)
 - 広域科学専攻 広域システム科学系 嶋田研究室 (shimada-lab)
- 総合文化研究科／教養学部 (51 組織)
- 教務課 (kyomu)
 - 広域科学専攻 生命環境科学系 丹野研究室 (tanno)
 - 国際社会科学専攻 (kiss)
 - 国際社会科学専攻 相関社会科学講座 (kiss-sr)
 - 国際社会科学専攻 国際関係論講座 (kiss-ir)
 - 国際社会科学専攻 山影進研究室 (yamakage-ken)
 - 国際社会科学専攻 山本泰研究室 (sociology-komaba)

- 化学部会 (chemistry)
 - 地域文化研究学科ドイツ分科 (doitsuka)
 - 複雑系生命システム研究センター (rcis)
 - 教養学部等事務部経理課 (keirika)
 - 総合文化研究科・教養学部 (c)
 - 地域文化研究専攻 (ask)
 - 地域文化研究専攻 イギリス科 (british section)
 - 超域文化科学専攻 表象文化論 (repre)
 - 地域文化研究専攻 若林研究室 (wakasemi)
 - 地域文化研究専攻「ジェノサイド研究の展開」(cgs)
 - 広域科学専攻 生命環境科学系 (bio-komaba)
 - 国際社会科学専攻 廣松研究室 (stat-komaba)
 - 英語教育支援室 英語I運営班 (eigoichi)
 - 超域文化科学専攻 比較文学比較文化コース (hikaku)
 - 広域科学専攻 関連基礎科学系 吉岡研究室 (yoshioka-lab)
 - 広域科学専攻 関連基礎科学系(共通) (dbs)
 - 国際社会科学専攻 内山研究室 (politics-komaba)
 - 文系三学科 地域文化研究科 ラテンアメリカ分科 (latinamerica)
 - 広域科学専攻 生命環境科学系 陶山研究室 (suyamalab)
 - 広域科学専攻 広域システム系 人文地理学 (humgeo)
 - 広域科学専攻 佐々研究室 (sasa)
 - 広域システム科学系 (system)
 - 基礎科学科 (kisokagakuka)
 - 広域科学専攻 広域システム科学系 増田研究室 (masuda_lab)
 - 広域科学専攻 染田研究室 (somedalab)
 - 地域文化研究学科 フランス分科 (sfr)
 - 文学・芸術の社会的統合機能の研究 (lac)
 - 広域科学専攻 関連基礎科学系 科学史・科学哲学研究室 (hps)
 - 地域文化研究専攻(古田元夫ゼミ) (furuta-semi)
 - 広域科学専攻 生命環境科学系 栗栖研究室 (xtalstructure)
 - 広域科学専攻 和田研究室 (lipid)
 - 広域科学専攻 関連基礎科学系 加藤雄介研究室 (kato-yusuke-lab)
 - 地域文化研究専攻 カンパニア地方の都市とヴィッラ集落をめぐる社会史的研究 (campania)
 - 超域文化科学専攻 文化人類学コース (bunjin)
 - 超域文化科学専攻／超域文化科学科 (choiki)
 - フランス語部会 (langue_fr)
 - 教養教育開発機構寄付研究部門「教養教育への囲碁の活用研究部門」(igo)
 - 情報システム研究グループ (graco)
 - 英語部会 (eigo)
 - 広域科学専攻 広域システム科学系 吉田研究室 (yoshidalab)
 - 国際社会科学専攻 木畑洋一研究室 (kibata)
 - 日独共同大学院プログラム (igk)
 - 広域科学専攻 関連基礎科学系 永田研究室 (nagatalab)
 - 前期宇宙地球部会 (ea)
- 教育学研究科／教育学部(2組織)
- 教育学研究科・教育学部図書室 (ikuto)
 - 附属中等教育学校 (hs)

新領域創成科学研究科 (72 組織)

- 先端生命科学専攻 分子認識化学分野 (molecular-recognition)
- 国際協力学専攻 佐藤研究室 (satoj)
- 国際協力学専攻 (cois)
- 国際協力学専攻 湊研究室 (minato)
- 国際協力学専攻 農業環境学研究室 (aee-labo)
- 環境学研究系 自然環境学専攻 (nenv)
- 環境学研究系 自然環境学専攻 生物圏情報学分野 (bis)
- 自然環境学専攻 自然環境形成学分野 (nef)
- 環境学研究系 社会文化環境学専攻 沿岸域環境研究室 (coastal-env)
- 物質系専攻 木村研究室 (phys)
- 環境学専攻 自然環境評価学研究室 (nee)
- 自然環境学専攻 自然環境構造学分野 (nes)
- 環境学専攻 久田研究室 (sml)
- 複雑理工学専攻 齊木研究室 (yukimuki)
- 人間環境学専攻 鳥居研究室 (torii-lab)
- 環境学専攻 環境健康システム学分野研究室 (envhlth)
- 環境学専攻 自然環境学コース 生物圏機能学分野 (lbf)
- 複雑理工学専攻 佐々木研究室 (sas)
- 国際協力学専攻 柳田研究室 (yanagita)
- 環境学専攻 社会文化環境学大講座 (sbk)
- 環境システム学専攻 徳永研究室 (tokunaga)
- 環境学専攻 環境システムコース 島田研究室 (shimada-labo)
- 人間環境学専攻 人間エネルギー環境学大講座 (hee)
- 物質系専攻 川合研究室 (surfchem)
- 先端生命科学専攻 動物生殖システム分野 (K-medaka)
- 国際協力学専攻 佐藤研究室 (rdm)
- 物質系専攻 和田研究室 (wadalab)
- 国際協力学専攻 中山幹康研究室 (nakayama)
- 国際協力学専攻 中山研究室「水のグローバルガバナンス」プロジェクト (ggwater)
- 複雑理工学専攻 鳥海研究室 (gaea)
- 環境システム学専攻 松橋・吉田研究室 (globalenv)
- 先端生命科学専攻 久恒辰博研究室 (hisatsune-lab)
- 環境システム学専攻 環境モデリング統合学研究室 (lemons)
- 先端生命科学専攻 資源生物制御学分野 (seigyō)
- 環境学専攻 保坂研究室 (ems)
- 環境学系 岡本研究室 (vis)
- 環境システム学専攻 (envsys)
- 先端生命科学専攻 医薬デザイン工学分野 (iyaku)
- 人間環境学大講座 メディア環境学研究室 (mesh)
- 人間環境学専攻 (h-k)
- 先端エネルギー工学専攻 藤井研究室 (fujiilabkashiwa)
- 環境学研究系 国際協力学専攻 國島研究室 (kunishimalab)
- 先端生命科学専攻 (ib)
- 先端生命科学専攻 人類進化システム分野 (jinrui)
- 先端エネルギー工学専攻 鈴木研究室 (aesuzuki)
- 社会文化環境学専攻 環境音響学研究室 (env-acoust)

- 社会文化環境学専攻 味埜研究室 (minolab)
 - 環境システム学専攻 地球環境工学研究室 (geelhome)
 - 環境システム学専攻 大島研 (oshima-lab)
 - 自然環境学専攻 自然環境構造学分野 春山研究室 (nesh)
 - 先端生命科学専攻 生命応答システム分野 (outou)
 - 人間環境学専攻 鳥居研究室 (droptech)
 - 自然環境学専攻 自然環境変動学分野 (changes)
 - 生涯スポーツ健康科学研究センター (hss)
 - メディカルゲノム専攻 生命分子解析学分野 (phys-biochem)
 - 基盤情報学専攻 高木研究室 (takagilab)
 - 環境システム学専攻 柳沢研究室 (yy-lab)
 - 複雑理工学専攻 溝川研究室 (mizokawa)
 - 環境学研究系 サステイナビリティ学教育プログラム (sustainability)
 - 社会文化環境学専攻 清水研究室 (rshimizu)
 - 人間環境学専攻 神保研究室 (jimbo_lab)
 - 複雑理工学専攻 高瀬・江尻研究室 (takase-ejiri-lab)
 - 複雑理工学専攻 武田常広研究室 (meg)
 - 国際協力学専攻 戸堂康之研究室 (yastodo)
 - 環境システム学専攻 海洋環境システム学研究室 (marenv)
 - 社会文化環境学専攻 北沢猛研究室 (kitazawalab)
 - 人間環境学専攻 室内環境保全学分野 (ieq)
 - 社会文化環境学専攻 鬼頭研究室 (kitosh)
 - 国際協力学専攻 田中幸夫研究室 (yukio)
 - メディカルゲノム専攻 田口研究室(科研費特定領域研究) (protein)
 - 基盤情報学専攻 竹内研究室 (lsi)
 - 海洋技術環境学専攻 (otpe)
- 情報学環・学際情報学府 (1 組織)
- 佐倉研究室 (sakuralab)
- 情報理工学系研究科 (14 組織)
- システム情報学専攻 第七研究室 (mels)
 - 知能機械情報学専攻 下山・松本・星野研究室 (leopard)
 - 21 世紀 COE「情報科学技術戦略コア」実世界情報システムプロジェクト (coe-rwisp)
 - 知能機械情報学専攻 (ist)
 - システム情報学専攻 第5研究室 (cyb)
 - 数理情報学専攻 武市研究室 (ipl)
 - 数理情報学専攻 (mist)
 - 国際交流室 (oir)
 - システム情報学専攻 (ipc)
 - 21 世紀 COE「情報科学技術戦略コア」 (coe-inside)
 - 情報理工学系研究科 (ist-inside)
 - 知能機械情報学専攻 CREST 移動体プロジェクト (crestidotai)
 - 創造情報学専攻 石塚研究室 (ishizuka-lab)
 - 知能機械情報学専攻 (irt)
- 公共政策大学院 (graspp)

留学生会、同窓会、クラブ等：7組織

東京大学中国留学生学友会 (acsut)

東京大学台湾留学生会 (todai-tsa)

ThaiTodai (thaitodai)

PAKISTAN STUDENTS CULTURAL SOCIETY (pakistan)

Tokyo University Islamic Cultural Society (tuics)

農学生命科学研究科 生物・環境工学専攻 同窓会 (shikou)

The University of Tokyo Nepalese Forum (TUNeF)

関連 URL : <http://park.itc.u-tokyo.ac.jp/>

遠隔講義支援サービス運用報告(2007.4 – 2008.3)

電子教材係

2007年度においては、遠隔講義支援に関わる設備等は、次の利用回数があった。
以下にその詳細を記す。

	会議等(回数)	講義等(講座数)
1) 本郷情報基盤センター 413 遠隔講義室	9	4
2) 本郷情報基盤センター 405 遠隔スタジオ	23	-
3) 駒場情報教育棟 E49 遠隔講義室	7	1
4) 貸出用テレビ会議システム	1	-
5) MCU:テレビ会議システム多地点接続制御装置	32	2
6) インターネットライブ中継	8	2

1 IP 遠隔会議システム

2007年03月26日、情報基盤センター 413 遠隔講義室のテレビ会議システムを Polycom ViewStation FX から Polycom HDX 9002(HD 対応テレビ会議システム) という機種へ変更した。

1.1 会議等利用状況(学外接続利用のみ抜粋)

1) 本郷 情報基盤センター 413 遠隔講義室

2007/05/18(金) 16:00-18:00

情報基盤センター 関谷助教 会議 慶応義塾大学と接続

2007/06/13(水) 10:00-12:00

情報理工学系研究科 山崎講師 学外との会議 大阪大学、京都大学と接続

2007/07/03(火) 13:00-17:30

情報学環 馬場教授 馬場ゼミ中継 香川県詫間電波高専 マルチメディアラボと接続

2007/07/24(火) 12:00-15:00

情報基盤センター 西澤主査 「先端研究施設共用イノベーション創出事業」記者会見 東京工業大学と接続

2007/08/21(火) 13:00-17:00

情報基盤センター 関谷助教 共同研究打ち合わせ 慶応義塾大学と接続

2007/09/07(金) 13:00-14:30

情報基盤センター 中山准教授 博士審査 奈良先端科学技術大学院大学と接続

2007/10/05(金) 13:00-17:00

情報基盤センター 関谷助教 研究打合せ 奈良先端科学技術大学院大学、北陸先端科学技術大学院大学と接続

2007/12/10(月) 10:00-12:00

情報基盤センター 中山准教授 Asia Future Internet SG meetin 事前テスト KOREN と接続

2007/12月18(火) 10:00-11:00

情報理工学系研究科 山崎講師 学外との会議 名古屋大学、京都大学と接続

2) 本郷 情報基盤センター 405 遠隔スタジオ

2007/04/17(火) 18:00-19:00

情報基盤センター 中山准教授 WIDE 研究会打合せ 北陸先端科学技術大学院大学、慶應義塾大学と接続

2007/04/18(水) 17:40-19:00

情報基盤センター 小川講師 会議 大阪大学と接続

2007/05/07(月) 14:00-16:30

情報基盤センター 加藤准教授 AI3 domestic meeting 慶應義塾大学と接続

2007/05/08(火) 08:00-09:00

情報基盤センター 加藤准教授 Studio Installation スタンフォード大学と接続

2007/05/17(木) 11:00-11:45

情報基盤センター 小川講師 会議 大阪大学と接続

2007/06/19(火) 11:00-14:00

情報基盤センター 中山准教授 研究打合せ 慶應義塾大学と接続

2007/07/24(火) 17:00-19:30

情報基盤センター 小川講師 会議 大阪大学と接続

2007/07/31(火) 16:30-18:00

情報基盤センター 小川講師 会議 大阪大学と接続

2007/08/03(金) 13:00-14:30

情報基盤センター 小川講師 会議 大阪大学と接続

2007/08/08(水) 09:30-12:30

情報基盤センター 中山准教授 博士審査 奈良先端科学技術大学院大学と接続

2007/08/30(木) 16:00-18:00

情報基盤センター 西村特任助教 mITF によるヒアリング 国立情報学研究所と接続

2007/10/15(月) 15:00-17:00

情報システム本部 安東特任講師 打合せ 九州大学と接続

2007/10/16(火) 10:30-12:00

情報基盤センター 加藤准教授 IETF 打合せ 広島市立大学、e-side と接続

2007/11/09(金) 15:00-17:00

情報基盤センター 加藤准教授 KMD 打合せ 慶應義塾大学と接続

2007/12/07(金) 14:30-16:30

情報基盤センター 小川講師 打合せ 大阪大学と接続

2007/12/18(火) 12:00-13:00

情報基盤センター 中山准教授 JGN2plus 打合せ NICT と接続

2007/12/26(水) 13:40-17:00

情報基盤センター 中山准教授 研究打合せ 慶應義塾大学と接続

2008/01/08(火) 13:00-15:00

情報基盤センター 中山准教授 NICT 打合せ KDDI と接続

2008/01/29(火) 10:30-12:00

情報基盤センター 中山准教授 NICT 研究打合せ NICT と接続

2008/01/29(火) 15:00-17:00

情報基盤センター 中山准教授 NICT 研究打合せ NICT と接続

2008/02/12(火) 15:00-16:00

情報基盤センター 中山准教授 NICT 打ち合わせ NICT と接続

2008/02/15(金) 13:30-15:00

情報基盤センター 西村特任助教 サーバ証明書発行プロジェクトに関する NII からのヒアリン

グ 国立情報学研究所と接続

2008/02/25(月) 17:30-18:00

情報基盤センター 西澤主査 東京工業大学打合せ 東京工業大学と接続

3) 駒場 情報教育棟 E49 遠隔講義室

2007/05/18(金) 16:00-21:00

人文社会系研究科 池田教授 修士の発表 バルセロナ自治大学と接続

2007/05/30(水) 16:00-21:00

人文社会系研究科 池田教授 修士の発表 バルセロナ自治大学と接続

2007/07/31(火) 13:00-15:30

総合文化研究科 刈間教授 モンゴル語中級 モンゴル国立大学と接続

2007/08/01(水) 13:00-15:30

総合文化研究科 刈間教授 モンゴル語中級 モンゴル国立大学と接続

2008/01/11(金) 13:00-14:30

総合文化研究科 金子教授 人事委員会 スイスと接続

2008/02/05(火) 16:00-18:00

総合文化研究科 阪本助教 講演 ワシントンと接続

2008/03/21(木) 09:00-12:00

総合文化研究科 山本リサーチフェロー インドシナ難民調査会議 世界銀行、立命館大学、名古屋大学と接続

4) 貸出用テレビ会議システム

2008/03/21(金) - 2008/03/31(月)

薬学系研究科 津谷客員教授 WHO ATC WHO Geneva と接続

1.2 講義等利用状況

1) 本郷遠隔講義室

2007/04/16-2008/01/21 毎週月曜日 13:00-14:45

新領域創成科学研究科 先端エネルギー工学輪講 柏キャンパスと接続

2007/04/10-2007/07/10 毎週火曜日 10:15-11:55

新領域創成科学研究科 平成 19 年度基盤情報学シミュレーション学 柏キャンパスと接続

2007/04/13-2008/01/25 毎週金曜日 10:15-11:55

新領域創成科学研究科 平成 19 年度基盤情報学専攻輪講 柏キャンパスと接続

2007/04/13-2007/07/20 毎週金曜日 13:00-14:45

新領域創成科学研究科 平成 19 年度基盤情報学並列数値処理論 柏キャンパスと接続

2) 駒場遠隔講義

2007/10/03-2008/01/23 毎週水曜日 16:45-19:00

情報学環 学際理数情報学基礎 A 本郷キャンパスと接続

2 MCU サービス

MCU:テレビ会議システム多地点接続制御装置 (2005 年度より導入)

2.1 会議等利用状況

2007/04/12(木) 09:30-12:00

情報基盤センター 米山係長 4 地点同時接続

2007/04/13(金) 12:30-16:00

新領域創成科学研究科 小野教授 7 地点同時接続

2007/04/12(木), 19(木), 26(木), 05/10(木), 17(木), 24(木), 06/07(木), 14(木), 21(木),

07/05(木), 12(木), 19(木), 08/23(木), 09/05(水), 06(木), 13(木), 20(木),

10/11(木), 18(木), 25(木), 11/08(木), 15(木), 22(木), 12/06(木), 13(木), 20(木),

2008/01/10(木), 17(木), 24(木), 02/14(木), 21(木), 28(木), 03/06(木), 13(木), 27(木) 12:00-15:00

人工物工学研究センター 浅間教授 4 地点同時接続

2007/04/24(火) 16:30-20:00

新領域創成科学研究科 小野教授 7 地点同時接続

2007/05/02(水) 12:30-16:00

新領域創成科学研究科 小野教授 7 地点同時接続

2007/05/07(月) 09:00-16:00, 05/08(火) 09:00-19:00

理学系研究科 藤代技術職員 4 地点同時接続

2007/06/12(火) 16:00-19:30

新領域創成科学研究科 小野教授 7 地点同時接続

2007/06/21(木) 12:30-15:00

情報基盤センター 米山係長 3 地点同時接続

2007/07/03(火) 12:00-18:30

情報基盤センター 中山准教授 5 地点同時接続

2007/07/04(水) 09:30-13:30

情報基盤センター 中山准教授 6 地点同時接続

2007/07/06(金) 12:30-16:00

新領域創成科学研究科 大井専門員 3 地点同時接続

2007/07/11(水) 12:00-16:00

情報基盤センター 中山准教授 12 地点同時接続

2007/07 月 13 日 (金) 15:30-18:00

情報基盤センター 米山係長 2 地点同時接続

2007/07/31(火) 09:30-12:30

情報基盤センター 中山准教授 6 地点同時接続

2007/08/01(水) 10:00-12:00

新領域創成科学研究科 大井専門員 3 地点同時接続

2007/08/08(水) 12:30-18:30

情報基盤センター 中山准教授 16 地点同時接続

2007/09/03(月) 09:00-13:00

新領域創成科学研究科 大井専門員 3 地点同時接続

2007/09/03(月) 13:30-17:00

新領域創成科学研究科 相田教授 5 地点同時接続

2007/09/03(月) 16:00-21:00

新領域創成科学研究科 小野教授 7 地点同時接続

- 2007/09/26(水) 13:00-17:00
 情報基盤センター 米山係長 3 地点同時接続
- 2007/10/08(月) 09:00-16:00
 理学系研究科 藤代技術職員 4 地点同時接続
- 2007/10/19(金) 09:30-12:00
 新領域創成科学研究科 油井係長 3 地点同時接続
- 2007/11/08(木) 13:00-17:00
 新領域創成科学研究科 油井係長 3 地点同時接続
- 2007/12/06(木) 17:00-18:00
 サステナビリティ学連携研究機構 小貫特任講師 5 地点同時接続
- 2007/12/21(金) 09:50-12:15
 新領域創成科学研究科 油井係長 3 地点同時接続
- 2008/01/18(金) 09:30-12:30
 新領域創成科学研究科 油井係長 3 地点同時接続
- 2008/02/07(木) 15:00-17:00
 新領域創成科学研究科 杉本准教授 3 地点同時接続
- 2008/02/08(金) 14:00-15:45
 新領域創成科学研究科 杉本准教授 3 地点同時接続
- 2008/02/12(火) 16:00-18:00
 情報基盤センター 中山准教授 4 地点同時接続
- 2008/02/13(水) 14:00-18:00
 情報基盤センター 中山准教授 8 地点同時接続
- 2008/02/28(木) 10:00-13:00
 サステナビリティ学連携研究機構 小貫特任講師 6 地点同時接続
- 2008/03/13(木) 10:45-13:00
 本部 白岩一般職員 3 地点同時接続

2.2 講義等利用状況

- 2007/04/16-2007/06/25 毎週月曜日 16:30-18:20
 新領域創成科学研究科 平成 19 年度夏学期 複雑理工学輪講 I,II 3 地点同時接続
- 2007/10/01-2008/01/28 毎週月曜日 16:30-18:20
 新領域創成科学研究科 平成 19 年度冬学期 複雑理工学輪講 I,II 3 地点同時接続

3 インターネットライブ中継

- 2007/04/10-2007/07/03 毎週火曜日 16:20-17:50 (配信範囲:学内のみ)
 大学総合教育研究センター 2007 年度夏学期 学術俯瞰講義「社会から見たサステナビリティ」
- 2007/04/19-2007/07/19 毎週木曜日 16:20-17:50 (配信範囲:学内のみ)
 大学総合教育研究センター 2007 年度夏学期 学術俯瞰講義「数理の世界」
- 2007/07/24(火) 12:00-17:00
 情報基盤センター 「先端研究施設共用イノベーション創出事業」記者会見
- 2007/08/22(水) 13:50-15:00
 情報基盤センター 「先端研究施設共用イノベーション創出事業」公募説明会

- 2007/10/15-2008/01/28 毎週月曜日 10:40-12:10 (配信範囲:学内のみ)
大学総合教育研究センター 2007年度冬学期 学術俯瞰講義「エネルギーと地球環境」
- 2007/10/10-2008/01/24 毎週木曜日 16:20-17:50 (配信範囲:学内のみ)
大学総合教育研究センター 2007年度冬学期 学術俯瞰講義「情報が世界を変える」
- 2007/11/26(月) 09:30-17:30
理学系研究科 東大-JAXA 学際理工学 20周年記念公開シンポジウム「宇宙科学と大学」
- 2007/12/03(月) 13:30-17:00
国立情報学研究所 SINET3 利用説明会
- 2008/03/11(火) 09:00-18:00, 03/12(水) 09:00-17:00
数物連携宇宙研究機構 IPMU Opening Symposium
- 2008/03/25(火) 09:00-12:00
本部 平成 19 年度東京大学卒業式

関連 URL <http://elearn.itc.u-tokyo.ac.jp/>

CFIVE 運用報告(2007 年 4 月—2008 年 3 月)

電子教材係

1 経過

CFIVE はプログラムソースを公開している学習管理システムであり、2004 年 4 月より東京大学でのサービスを開始した。サービス開始以来 CFIVE を利用する講義数は順調に増加しており、2005 年度は 35 講義、2006 年度は 63 講義、2007 年度は 104 講義で利用された。

以下に利用された講義名等のデータ、改良された機能、広報等を示す。

2 夏学期に CFIVE を利用した講義

夏学期(2007 年 4 月—2007 年 9 月) に CFIVE を利用した講義は以下の通りである。

項番	開講組織	講義名	担当教員(敬称略)
1	教養学部 (前期課程)	情報(火 2)	丹羽 清
2	教養学部 (前期課程)	情報(火 2)	玉井 哲雄
3	教養学部 (前期課程)	情報(火 4)	玉井 哲雄
4	教養学部 (前期課程)	情報(金 3)	中村 政隆
5	教養学部 (前期課程)	情報(金 4)	中村 政隆
6	教養学部 (前期課程)	情報(水 2)	開 一夫
7	教養学部 (前期課程)	情報(水 2)	藤垣 裕子
8	教養学部 (前期課程)	情報(水 4)	藤垣 裕子
9	教養学部 (前期課程)	情報(火 3)	植田 一博
10	教養学部 (前期課程)	情報(火 4)	植田 一博
11	教養学部 (前期課程)	情報(月 1)	増原 英彦
12	教養学部 (前期課程)	情報(水 3)	山口 和紀
13	教養学部 (前期課程)	情報(金 3)	山口 泰
14	教養学部 (前期課程)	情報(金 4)	尾上 能之
15	教養学部 (前期課程)	情報(木 2)	田中 哲朗
16	教養学部 (前期課程)	情報(月 3)	石崎 雅人
17	教養学部 (前期課程)	情報(月 4)	石崎 雅人
18	教養学部 (前期課程)	情報(木 2)	原田 至郎
19	教養学部 (前期課程)	情報(木 3)	原田 至郎
20	教養学部 (前期課程)	情報(木 1)	武市 正人
21	教養学部 (前期課程)	情報(月 5)	竹内 郁雄
22	教養学部 (前期課程)	情報(金 2)	廣田 光一,小穴 英廣
23	教養学部 (前期課程)	情報(水 4)	美馬 秀樹
24	教養学部 (前期課程)	情報(月 4)	田村 肇
25	教養学部 (前期課程)	情報(月 3)	河内谷 幸子
26	教養学部 (前期課程)	情報(火 3)	中谷 多哉子
27	教養学部 (前期課程)	情報(木 3)	辰己 丈夫
28	教養学部 (前期課程)	情報(金 1)	蔡 東生
29	教養学部 (前期課程)	情報(金 2)	蔡 東生
30	教養学部 (前期課程)	生命科学 I(1)(月 3)	佐藤直樹
31	教養学部 (前期課程)	生命科学(1)(木 1)	佐藤直樹
32	教養学部 (前期課程)	情報(水 3)	日暮 栄治
33	教養学部 (前期課程)	プログラム構成論(水 5)	山口和紀

項番	開講組織	講義名	担当教員(敬称略)
34	教養学部(前期課程)	全学自由研究ゼミナール(水5)	古田 幹雄
35	教養学部(前期課程)	図形科学演習II(火5)	山口 泰
36	教養学部(前期課程)	図形科学II(共通)	鈴木賢次郎ほか
37	教養学部(前期課程)	英語二列C(金4)	アルヴィなほ子
38	教養学部(前期課程)	情報科学概論I(月5)	増原英彦
39	教養学部(後期課程)	英語31	広瀬友紀
40	教養学部(前期課程)	図形科学II(水曜日3限)	鈴木賢次郎, 柏原賢二
41	教養学部(前期課程)	図形科学II(水曜日4限)	鈴木賢次郎, 柏原賢二
42	教養学部(前期課程)	英語二列C(木3)	河合祥一郎
43	教養学部(前期課程)	英語二列C(木2)	河合祥一郎
44	教養学部(前期課程)	英語二列R(金2)	河合祥一郎
45	教養学部(前期課程)	英語二列R(金1)	河合祥一郎
46	教養学部(前期課程)	青の革命と緑のダム	蔵治 光一郎
47	教養学部(前期課程)	図形科学II(火3)	田中一郎, 横山ゆりか
48	教養学部(前期課程)	図形科学II(火4)	田中一郎, 横山ゆりか
49	教養学部(前期課程)	情報システム利用入門(金5)	玉井哲雄, 金子知適
50	教養学部(前期課程)	英語二列R(金1)	荒木純子
51	教養学部(前期課程)	英語二列R(金2)	荒木純子
52	教養学部(前期課程)	図形科学II(月3)	加藤道夫, 金井崇
53	教養学部(前期課程)	図形科学II(月4)	加藤道夫, 金井崇
54	教養学部(前期課程)	英語二列R(R)(木2)	加納孝代
55	教養学部(前期課程)	英語二列R(R)(木3)	加納孝代
56	教養学部(前期課程)	宇宙科学実習II(月4,5)	鈴木 建
57	教養学部(前期課程)	宇宙科学実習II(火4,5)	鈴木 建
58	教養学部(前期課程)	図形科学II(金3)	深野 暁雄, 堤 江美子
59	教養学部(前期課程)	図形科学II(金4)	堤 江美子, 深野 暁雄
60	教養学部(後期課程)	国際協力	旭英昭
61	教養学部(前期課程)	図形科学II(木2)	長島 忍
62	教養学部(前期課程)	図形科学II(木3)	長島 忍
63	情報基盤センター	ICカードによる出欠管理システム	ICカードによる出欠管理システム開発グループ
64	情報基盤センター	講義自動収録システム	講義自動収録システム開発グループ

3 冬学期に CFIVE を利用した講義

冬学期(2007年10月-2008年3月)にCFIVEを利用した講義は以下の通りである。

項番	開講組織	講義名	担当教員(敬称略)
1	教養学部(前期課程)	情報科学(共通)	増原英彦
2	教養学部(前期課程)	英語2列(PW)(金3)	アルヴィなほ子
3	教養学部(前期課程)	英語二列P(PO)(金1)	小林薫
4	教養学部(前期課程)	英語二列P(PO)(金2)	小林薫
5	教養学部(前期課程)	英語II列P(PO)(火5)	中尾まさみ
6	教養学部(前期課程)	英語II列P(PO)(火2)	中尾まさみ
7	教養学部(前期課程)	宇宙素粒子物理学のフロンティア(木4)	竹内康雄
8	教養学部(前期課程)	情報科学(水4)	田中哲朗
9	教養学部(前期課程)	情報科学(月4)	河内谷幸子
10	教養学部(前期課程)	英語二列P(PO)(月4)	広瀬友紀
11	教養学部(前期課程)	英語二列P(PO)(月5)	広瀬友紀
12	教養学部(前期課程)	英語特修(火2)	広瀬友紀
13	教養学部(前期課程)	英語特修(火3)	広瀬友紀
14	工学部	コンピュータ及び演習(水1)	阿久津好明
15	教養学部(前期課程)	英語二列P O(火2)	河合祥一郎
16	教養学部(前期課程)	英語演習II(木2)	河合祥一郎
17	農学部	情報工学(木3)	中村典裕
18	教養学部(前期課程)	図形科学I(水4、金3)	鈴木賢次郎
19	教養学部(前期課程)	情報科学(水3)	玉井哲雄
20	教養学部(前期課程)	情報科学概論II(火5)	玉井哲雄
21	教養学部(前期課程)	情報科学(火4)	山口和紀
22	教養学部(前期課程)	情報科学(木2)	増原英彦
23	工学部	マテリアルシミュレーションII(火4)	渡邊聡
24	広域科学科広域システム	システム基礎数学3(金4)	山口和紀
25	工学部	プログラミング基礎演習(金5)	伊庭斉志, 峯松信明, 川原圭博
26	教養学部(前期課程)	情報科学(金4)	立堀道昭
27	システム創成学科	WWWアプリケーション入門(水3,4,5)	大和裕幸, 浅間一
28	教養学部(前期課程)	英語二列P(PW)(金1)	荒木純子
29	教養学部(前期課程)	英語二列P(PW)(金2)	荒木純子
30	広域科学科広域システム	情報システム科学実習I(月4,5)	山口泰
31	広域科学科広域システム	情報システム科学V(水2)	山口泰
32	経済学部	経営科学(月2,木3)	安田雪
33	教養学部(前期課程)	情報科学(金3)	上田哲郎
34	教養学部(前期課程)	情報科学(木3)	中尾彰宏
35	教養学部(後期課程)	情報システム科学実習II(水5)	山口和紀, 金子知適
36	教養学部(前期課程)	科学史(水1)	板垣良一
37	教養学部(前期課程)	方法基礎、データ分析(水1)	上田博人
38	情報基盤センター	技術職員研修	尾上能之
39	情報基盤センター	講義自動収録システム	講義自動収録システム 開発グループ
40	教養学部(前期課程)	初修外国語スペイン語(公開講座, 全員登録)	スペイン語部会全教員

4 組織毎の利用状況

組織毎にまとめた 2007 年度の利用状況は、以下の通りである。

開講組織	講義数
教養学部（前期課程）	88
教養学部（後期課程）	3
情報基盤センター	4
広域科学科広域システム	3
経済学部	1
農学部	1
工学部	3
システム創成学科	1
計	104

5 新たに追加された機能

2007 年度に新たに CFIVE に追加された機能は、以下の通りである。

- ・（複数プラグイン）各コース別の通し番号を割り当てられるようにする
- ・（ユーザ）ユーザ登録時 CSV ファイルのスペース処理の改良
- ・（ユーザ）LDAP に登録の無いユーザに関わるデータエラーの修正
- ・（授業）授業／出席一覧表のダウンロード（CSV/Excel/HTML の形式）
- ・（FAQ）「FAQ 一覧(カテゴリ表示)」の右に「FAQ 一覧(全件表示)」を追加
- ・（掲示板）ダイレクト URL でのアクセス時も公開終了日を有効にする
- ・（掲示板）既存メッセージ返信の際にもプレビュー表示する
- ・（掲示板）掲示板への投稿を RSS で通知する
- ・（テスト）学生がテスト内容を確認するときの参照可能時期の設定変更
- ・（テスト）期間内は何回でもテストを受けられる機能
- ・（教材）教材もカテゴリ(フォルダ)に分けて保存できるようにする
- ・（課題）課題評価の編集権限で、Author 以外の Lecturer, Assistant も編集可能にする
- ・（課題）誤って作成した課題の削除を可能にする
- ・（課題）グループ内で公開の制御を可能にし、学生がレポートを見せ合う機能
- ・（課題）レポート名にコロン(:)を使えない件の修正
- ・（課題）課題内容編集の設定結果が反映されていないように見える件の修正
- ・（WebDAV）提出期限を過ぎてもレポートの提出ができる件の修正
- ・（新機能）「WWW 経由での学生自身による CFIVE への利用登録」機能の実装

6 広報

2007 年度に行った主な広報は、以下の通りである。

お知らせ

- | | | |
|---|--------------------------------------|------------|
| 1 | CFIVE の緊急メンテナンスのお知らせ(終了) | 2008/04/10 |
| 2 | 2008/04/10 10:00 頃から発生している障害について(終了) | 2008/04/10 |
| 3 | 2008/04/09 13:57 頃から発生していた障害について(終了) | 2008/04/09 |

- 4 CFIVE 学生用マニュアルについて(教職員、学生向け) 2008/04/03
- 5 平成 20 年度夏学期の CFIVE のご利用について(教職員向け) 2008/04/03
- 6 UTask-Web のアカウント での CFIVE の利用について(教職員向け) 2008/04/03
- 7 IC カード による 出欠管理システムについて(教職員向け) 2008/04/03
- 8 CFIVE Support ではレポート 提出の受付は行っていません。(学生向け) 2008/04/03
- 9 課題作成の際の注意について(教職員向け) 2008/04/03
- 10 c5-support への問い合わせについて 2008/04/03
- 11 受講していない講義でも表示されることがあります 2008/04/03
- 12 CFIVE に関する 情報について 2008/04/03
- 13 CFIVE サーバのメンテナンスのお知らせ(2008/03/27) (終了) 2008/03/28
- 14 CFIVE サーバのメンテナンスのお知らせ(2008/01/04)(終了) 2008/01/04
- 15 CFIVE 問い合わせ用メールアドレスへの送信の際の障害について(終了) 2007/12/28
- 16 2007/12/18 の 12:00~12:40 頃に発生した CFIVE にアクセスできなかった現象について 2007/12/18
- 17 駒場キャンパス構内における計画停電による CFIVE への影響について(終了) 2007/12/09
- 18 課題機能における不具合について(2007/11/14)(終了) 2007/11/14
- 19 2007/11/13 の 12:00~12:35 頃の CFIVE へのアクセスが不安定になった現象について 2007/11/13
- 20 情報基盤センターの計画停電による CFIVE への影響について(終了) 2007/10/30
- 21 情報基盤センターの計画停電による CFIVE への影響について(終了) 2007/10/02
- 22 課題機能におけるレポート ファイル登録の際の注意について(学生向け) 2007/09/28
- 23 CFIVE のメンテナンスのお知らせ(2007/08/14) (終了) 2007/08/14
- 24 2007/06/29 に発生した障害について(終了) 2007/06/29
- 25 電気系統の障害によるサービス停止のお詫び(終了) 2007/05/29
- 26 2007/05/16 12:30 頃から発生している障害について(終了) 2007/05/16
- 27 CFIVE サーバのメンテナンスのお知らせ(2007/03/14)(終了) 2007/05/11
- 28 2007/04/16 頃から発生していた各機能におけるエラーについて 2007/04/19
- 29 CFIVE の緊急メンテナンスのお知らせ(2007/04/05) (終了) 2007/04/05

以上

図書館電子化部門



データベースリーフレット



ネットでアカデミック 学術情報へのアクセスガイド

GACoS Gateway to Academic Contents System



検索支援システムUT-Kiwi



書籍探索支援システム リttlナビゲーター



東京大学学術機関リポジトリ(UT Repository)



図書館電子化部門

概要

図書館情報係 本多 玄

図書館電子化業務部門は、図書館情報係、デジタル・ライブラリ係、学術情報リテラシー係の3係からなる。

図書館電子化業務部門は附属図書館と連携し、本学における学術情報リソースの提供とその利便性を高めることで本学の学術研究活動を支援する一方、本学の研究成果を社会に向けて発信する役割も担っている。

図書館情報係は主に全学の図書館室で使用される図書館システム(「附属図書館学術情報システム」)の運用・管理を行なっている。平成19年度は学務システム/予算執行管理システムといった外部システムとの連携のための改修を行なった。

デジタル・ライブラリ係は本学で生産される学術情報を電子化・発信することがその主な任務であるが、平成19年度は本学の学術機関リポジトリ(「UT Repository」)の本格運用により、収録コンテンツの増強及び著作権処理や広報活動、そして運用システムの開発等に取り組んだ。

学術情報リテラシー係は、本学に必要と思われる商業学術データベースの選定と契約を行なうとともに、学内で利用可能な各種データベースの利用に関する講習会やパンフレットの作成を通じて学内教職員・学生の学術情報リテラシーを高める活動を本年度も継続的に行なった。

デジタルコンテンツサービス

デジタル・ライブラリ係

1 運用報告

1.1 東京大学学術機関リポジトリの構築

東京大学学術機関リポジトリ～UT Repository～とは、東京大学で生産されたさまざまな研究成果を電子的な形態で集中的に蓄積・保存し、学内外に公開することを目的としたインターネット上の発信拠点である。平成 16 年度から附属図書館と連携し学術機関リポジトリの構築を行い、平成 18 年 4 月 1 日から「東京大学学術機関リポジトリ(UT Repository)」としてサービスを開始した。

(<http://repository.dl.itc.u-tokyo.ac.jp/>)

平成 19 年度は、昨年度に引き続き、国立情報学研究所の最先端学術情報基盤(CSI)構築事業による外部資金を調達・活用するなどして、附属図書館と連携協力して次の業務を行った。



1.1.1 広報活動

各部局教授会でのプレゼンテーション資料等の広報資料を改訂し、以下の各部局教授会にて附属図書館長が UT Repository のプレゼンテーションを行う際に、学内の研究者に向けた説明資料と著作権ガイドを1つにしたパンフレット「知っていますか？ UT Repository」と共に配付した。

また、紀要及び学位論文の UT Repository への登録を増強するため、紀要担当者及び学位論文の各部局各担当者への働きかけを行うと共に、各部局の特徴を考慮した著作権処理への対応を行った。

UT Repository の部局教授会でのプレゼンテーション

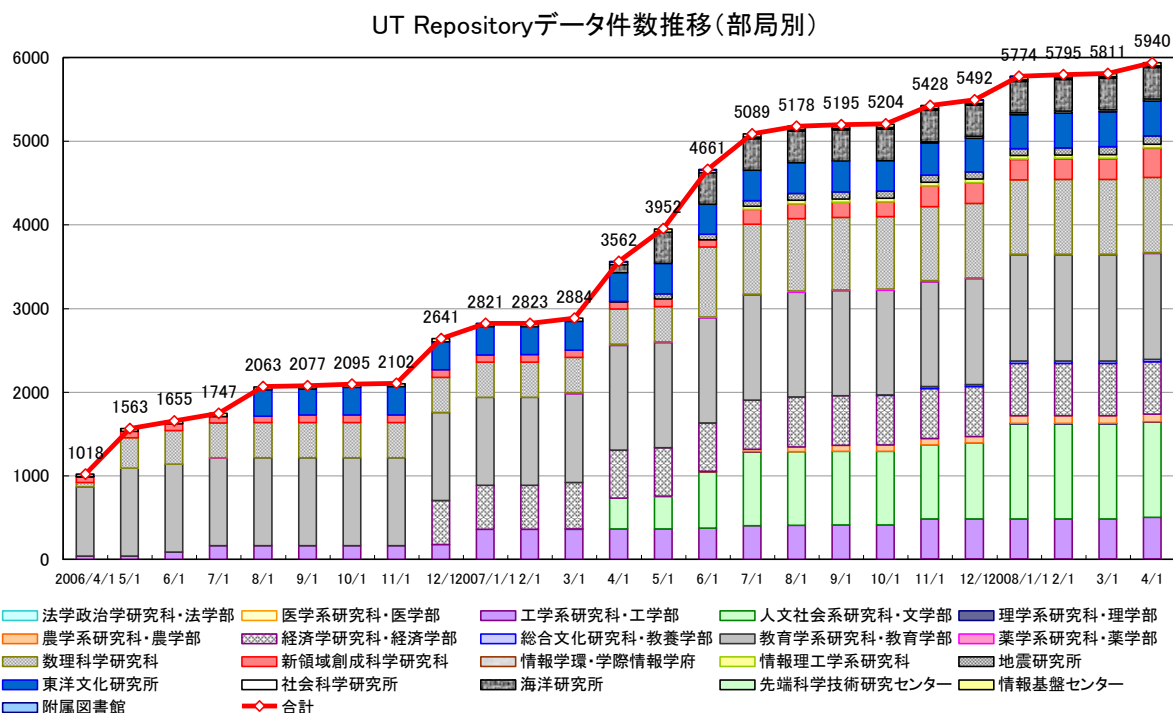
部局	月日	資料配付部数
薬学部・薬学系研究科	2007/4/11	77
東洋文化研究所	2007/5/31	70
地震研究所 + 分子細胞生物学研究所	2007/5/31	100
医学部・医学系研究科	2007/6/6	200
経済学部・経済学研究科	2007/6/13	60
教育学部・教育学研究科	2007/6/20	55
医科学研究所	2007/6/21	100
総合文化研究科	2008/1/24	250
数理科学研究科	2008/3/21	75



1.1.2 コンテンツ作成と著作権許諾処理

紀要、学位論文、学術雑誌論文を中心としてコンテンツの収集と作成を行い、3月末には5,940件のデータを公開した。

機関リポジトリでは本文そのものをPDF等で公開するため、単に電子化作業を行うだけでは無く、著作権の許諾処理を必須として作業を行っている。著者本人や共著者の許諾はもちろんであるが、学術雑誌掲載論文では出版者の許諾も必要であるため、機関リポジトリへの掲載許諾情報を調査し、必要に応じて学協会等へ個々に許諾の確認を取る作業を行った。また、紀要及びCOEリサーチアシスタント報告書等のバックナンバーの電子化にあたり、著者への許諾を電子メールと郵送により実施した。



1.1.3 システム開発

前述のとおり、著作権許諾処理は著者本人、共著者、出版者の許諾が必要であり、学生の場合は指導教員の許諾が必要となる場合もあるため、これらの許諾処理を同一のデータベースで一元管理し、全ての許諾処理が完了しないコンテンツは公開しないように制御するデータベースとシステムの開発を行い、著作権許諾管理とリポジトリへの登録管理業務の効率化を図る。平成19年度は昨年度開発した登録管理システムの根幹データベースに対して機能実装を行った。主な機能は次のとおり。

- ・アイテムのインポート機能
- ・GUIによるアイテム更新機能
- ・GUIによる検索及びエクスポート機能
- ・登録アイテムの許諾状況一覧機能
- ・DSpaceへの自動一括登録機能
- ・著作権許諾及び公開日制限をクリアしたアイテムの自動公開機能(DSpace連携)
- ・各紀要のホームページ作成による巻号単位での表示機能

今年度の開発により、当初予定していた著作権許諾管理とリポジトリへの登録管理業務の効率化のための運用システムが完成した。今後はシステムの機能改善と学内の関連データベースとの連携機能の開発を予定している。

1.1.4 学内システム連携

今後、標準実績データベース等の学内関連システムとの連携を図り、研究業績等を One Writing で登録できると共に、容易に各種研究成果リストの出力が可能となるようなシステムの構築を推進する予定である。

1.1.5 外部システム連携

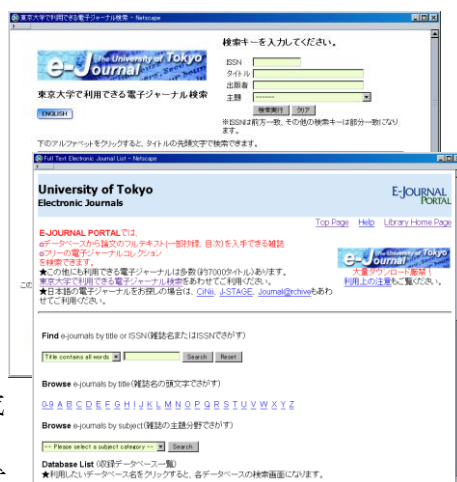
UT Repository は OAI-PMH (Open Archives Initiative Protocol for Metadata Harvesting) に対応しており、メタデータを外部へ提供して外部の検索エンジンにより検索することが可能となっている。既に昨年度には Google Scholar へ対応済みであるが、平成 19 年度は、国立情報学研究所の JuNii+、エルゼビアの Scirus、トムソンサイエンティフィックの Current Web Contents、フランス IFREMER の Avano(海洋分野)等とハーベストを実現し、外部からの検索への対応を拡大した。

1.2 電子ジャーナルのゲートウェイサービス

電子ジャーナルの提供を円滑に行うための電子ジャーナルリンク集データベース(約 8,500 件)のメンテナンス及び利用統計採取のためのプロキシサーバの維持・管理と、本学が契約するデータベースを経由して利用できる電子ジャーナルとフリーの電子ジャーナルについての管理ソフトを導入した E-Journal Portal(約 45,000 件)の学内へのサービスを行っている。

また、Elsevier 社の電子ジャーナルサービス ScienceDirect を学外から利用できるリモートアクセス用 ID を、教員・大学院生に対し発行する業務を行っており、平成 19 年度は 1,072 件の発行を行った。

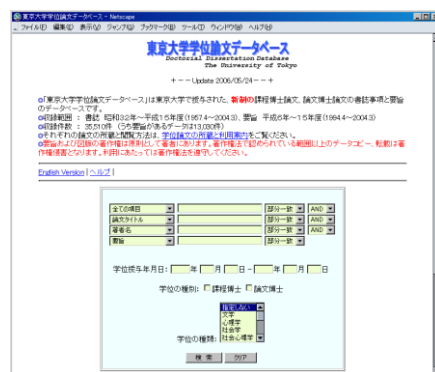
さらに平成 20 年度から開始予定の SSL-VPN Gateway 試行サービスによる電子ジャーナルの外部からのアクセス準備を行った。
(<http://ejournal.dl.itc.u-tokyo.ac.jp/>)



1.3 東京大学学位論文要旨データベースの構築

学位論文データベースは従来冊子体で刊行されていた「東京大学博士学位論文の内容の要旨と審査結果の要旨」を電子化して公開しているものである。

平成 19 年度は平成 17 年度論文要旨 1,404 件の電子化を行った。昨年度同様に、提出電子ファイルを全面的に利用し、データベース作成の効率化と経費の削減を図った。
(<http://gakui.dl.itc.u-tokyo.ac.jp/>)



1.4 貴重書のデータベース化支援

1.4.1 平賀譲 デジタルアーカイブ

東京大学工学系研究科環境海洋工学専攻、同新領域創成科学研究科人間環境学専攻、呉海事歴史科館の協力で作成された「平賀譲 デジタルアーカイブ」への協力を行い、本部門の公開用サーバからの公開を行った。

「平賀譲 デジタルアーカイブ」は、海軍造船官であった平賀譲(ひらが・ゆずる(1878-1943):海軍造船中将・第13代東京帝国大学総長(1938-1943))の遺した艦艇計画・建造関係の技術資料を中心とする資料(約40,000点)を電子化したものである。

(<http://rarebook.dl.itc.u-tokyo.ac.jp/hiraga/>)



1.4.2 写真帖『東京帝国大学』電子版

附属図書館と協力し、明治30年代の東京帝国大学を撮影した大判横長の写真帖を本部門の公開用サーバから公開した。

本写真帖は、フランス・パリの万国博覧会、米国・セントルイスの万国博覧会に出品するために製作されたもので、当時の総長、前総長や部局長、教授たちの肖像と校舎設備の状況を撮影した写真を収録し、序文や見出しなどには英文が併記されている。

(<http://rarebook.dl.itc.u-tokyo.ac.jp/shashincho/>)



1.5 貴重書展示会の支援

附属図書館展示委員会に協力し、通常展示および特別展示への協力を行っている。平成19年度の特別展示会「東京大学創立130周年・総合図書館再建80周年記念特別展示会—世界から贈られた図書を受け継いで—」では、パンフレット及びポスターのデザイン作成を行い、電子展示版(ホームページ)を作成・公開した。

(<http://www.lib.u-tokyo.ac.jp/tenjikai/tenjikai2007/index.html>)



学術情報リテラシー支援

学術情報リテラシー係

1 運用報告

1.1 情報探索ポータルサイト GACoS(Gateway to Academic Contents System)

デジタル学術情報にアクセスするためのポータルサイトとして GACoS(ガコス)を構築、運用している。日本語、英語、中国語、韓国語の 4ヶ国語版があり、Tips ページの項目を増やすなど、内容の充実に努めた。

(<http://www.dl.itc.u-tokyo.ac.jp/gacos/>)

1.2 リテラシー支援資料の作成・更新

1.2.1 「図書館利用ガイド」

附属図書館情報リテラシー教育部会編集の東京大学の図書館を初めて利用する学生等を対象とした「図書館利用ガイド 2008」及び、英語版「Guide to UT Libraries 2008」の作成に協力した。

1.2.2 「ネットでアカデミック」

東京大学における文献調査のガイドブックとして発行している小冊子「ネットでアカデミック」(日本語・英語・中国語・韓国語の 4ヶ国語版)を作成した。附属図書館で新規に導入した東京大学学術論文横断検索(UT Article Search)やリンクリゾルバ(UT Article Link)の解説を加えるなどの内容の改訂を行った。

冊子版を学内各図書館・室に配布するとともに、PDF 版を GACoS 上で公開し、パソコンからも参照できるようにした。

(<http://www.dl.itc.u-tokyo.ac.jp/gacos/net.html>)



1.2.3 データベースリーフレット

東京大学で利用できる以下のデータベースについて、それぞれの特徴や検索方法を紹介したリーフレット 6 種(日本語・英語の 2ヶ国語版)を作成し、学内各図書館・室に配布した。今年度は、新サービス「UT Article Search」や「UT Article Link」と「東京大学学術機関リポジトリ」のリーフレットを新規に作成した。また、「東京大学学位論文データベース」について、全文情報へのリンクの説明を加える改訂を行った。

PDF 版を GACoS 上で公開し、パソコンからも参照できるようにしている。

- ・ 日本語版: <http://www.dl.itc.u-tokyo.ac.jp/gacos/literacy.html>
- ・ 英語版: <http://www.dl.itc.u-tokyo.ac.jp/gacos/e/literacy-e.html>

リーフレット 6 種類は次のとおり。

- ・ GACoS / IRI : Index to Resources on Internet
- ・ 東京大学 OPAC
- ・ 東京大学学位論文データベース / 東京大学学術機関リポジトリ
- ・ 東京大学で利用できる電子ジャーナル検索 / E-JOURNAL PORTAL
- ・ 東京大学学術論文横断検索 (UT Article Search)
- ・ 東京大学学術論文リンク (UT Article Link)



1.3 Litetopi メールマガジン

平成 17 年 12 月に開始した東京大学所属者に各種データベースサービスや、情報探索ガイダンスの案内などを送る「Litetopi (リテトピ) メールマガジン」の発行を継続して行った。平成 19 年度は、月 1～2 回の頻度で、No.28 から No.46 までの 19 号発行した。登録者も約 750 人に増加した。

バックナンバーは、GACoS(<http://www.dl.itc.u-tokyo.ac.jp/gacos/literacy.html>)から参照できるようにしている。

1.4 学術データベースの拡充

1.4.1 人文社会科学系全文データベースの拡充

人文社会科学系の雑誌記事アーカイブデータベース Periodicals Archive Online (旧 PCI Full Text) に Collection 3 を追加契約し、Collection 1、2 と合わせ、約 300 誌が、全文までアクセス可能になった。

1.4.2 事典・辞書データベースの拡充

『日本歴史地名大系』(平凡社刊) 全 51 冊を完全データ化した「日本歴史地名大系オンライン版」と『日本国語大辞典』(小学館刊) 全 13 巻の Web 版である「日国オンライン」を導入した。SSL-VPN Gateway 試行サービスにより、学外からも利用可能にしている。

2 講習会・研究会開催報告

2.1 情報探索ガイダンスの実施概要

データベースをはじめとした電子的学術資料を学習・研究に効率的に利用してもらうことを目的とする情報探索ガイダンスを、年間を通じて各種開催した。平成 19 年度は 100 回開催し、合計で 1,124 人の参加があった。いくつかのコースでは、電子教材系の協力の下、テレビ会議システムを用いて、柏図書館に中継を行い、柏キャンパスの利用者が本郷キャンパスまで出向かなくても同じ内容を視聴できるようにした(以下の表で会場欄に※印のあるもの)。また、医学図書館や駒場図書館主催の講習会に配布資料や台本の提供、会場の貸出や広報等の協力を行ない、部局での講習会開催の支援をした。

2.2 平成 19 年度情報探索ガイダンス開催実績

2.2.1 テーマ別ガイダンス

1 つのテーマについて、検索実習を交えながら解説するテーマ別ガイダンスを 5 コース実施した。平成 19 年度は、「新聞記事を探すには?」コースを新設した。「電子ジャーナルを利用するには?」コースでは、新サービス UT Article Search や UT Article Link の解説を加え、内容の充実を図った。

コース名	月日	回数	人数	内容
図書館の蔵書を検索す	9/18,3/19	2	4	OPAC の使い方を中心に、図書館の蔵

るには?				書を検索できるデータベースを紹介。
日本の論文を探すには?	5/11,10/18,1/23	3	31	FELIX、CiNii の使い方を中心に、日本の雑誌論文を検索できるデータベースを紹介。
電子ジャーナルを利用するには?	5/17,10/12,12/4,12/12,2/15,3/13	6	48	電子ジャーナルにアクセスする方法や、代表的な出版社の電子ジャーナルサイトを例にとって実際の利用方法を解説。
自宅から検索するには?	5/23,7/9,10/24,1/29,3/7	5	45	学外からも利用できるデータベース、電子ジャーナルの利用手続きやアクセス方法を紹介。
新聞記事を探すには?	9/13,1/11	2	9	聞蔵 II、LexisNexis Academic の使い方を中心に、新聞記事の探し方を紹介。
合計		18回	137名	

2.2.2 データベースユーザトレーニング

特定のデータベースの使い方について、検索実習を交えながら解説した。平成 19 年度は、「JapanKnowledge」コースを新設した。また、2 月の Web of Science コースでは、講習時間を延長して文献管理ツール EndNote Web の解説を加え、好評であった。

コース名	月日	回数	人数	内容
Web of Science	9/26,10/30,2/4,2/28	4	26	Web of Science の特徴を解説し、キーワード検索や引用文献をキーにした検索方法を紹介。
JapanKnowledge	9/7,1/17	2	3	各種辞事典の一括検索や、東洋文庫、会社四季報等収録コンテンツを紹介。
合計		6回	29名	

2.2.3 分野別データベースユーザトレーニング

データベースの提供元から講師を招き、各分野の主要なデータベースごとの講習会を実習形式で実施した。新規導入やバージョンアップのあったデータベースを加えてコース数を増やすとともに、会場も駒場図書館の他、柏図書館でも実施した。

・総合分野

コース名	月日	時間	人数	会場
LexisNexis Academic	6/7	13:45-14:45	12	総合図書館
	11/14	15:00-16:30	13	
JapanKnowledge	6/12	15:00-16:30	12	総合図書館
	11/1	14:40-16:10	15	駒場図書館
EndNote Web	6/20	16:30-17:30	25(12)	総合図書館※
	6/29	16:30-17:30	22	駒場図書館
朝日新聞オンライン記事データベース「聞蔵 II ビジュアル」	7/3	16:00-17:00	9	総合図書館
MC-統計	7/5	15:00-16:00	14	総合図書館

・人文社会科学系分野

コース名	月日	時間	人数	会場
Literature Resource Center with MLA International Bibliography	5/30	13:45-14:45	6	総合図書館
InfoTrac Custom	5/30	15:00-16:00	14	

Business Source Elite	6/6	13:15-14:45	7	
PsycINFO・PsycARTICLES	6/6	15:00-16:30	5	
Lexis.com	6/7	15:00-16:30	11	
ASSIA・LLBA	6/8	10:30-12:00	10	
Eighteenth Century Collections Online (ECCO)	6/18	13:00-14:30	3	駒場図書館
	6/11	11:00-12:00	9	総合図書館
eol ESPer	6/21	15:00-16:00	16	駒場図書館
	6/13	13:30-14:30	4	総合図書館
JSTOR	6/15	15:30-16:30	7	総合図書館
Web of Science 人文社会科学系	6/20	10:30-12:00	10	
	6/29	14:45-16:15	11	

・自然科学系分野

コース名	月日	時間	人数	会場
AGRICOLA・ProQuest Agriculture Journals	6/4	15:00-16:30	7	総合図書館
SciFinder Scholar	6/19	15:00-16:30	4	駒場図書館
	6/22	15:00-16:30	10	総合図書館
	6/28	15:00-16:30	11	柏図書館
Web of Science 自然科学系	6/20	13:30-15:00	42(27)	総合図書館※
	6/29	13:00-14:30	2	駒場図書館
BIOSIS Previews	6/20	15:15-16:15	6(0)	総合図書館※
合計	28回		317名	

※柏図書館に中継したコース。人数欄に()で中継参加者を内数で示す。

2.2.4 総合図書館オリエンテーション

総合図書館オリエンテーションに協力し、春季は「OPAC 入門」、秋季は「データベース講習会」として、各 30 分のコースを担当した。

コース名	月日	回数	人数	内容
OPAC 入門	4/6～4/20	10	86	OPAC の使い方と GACoS の紹介
データベース講習会	10/10	2	24	OPAC、FELIX、電子ジャーナルの利用方法
合計		12回	110名	

2.2.5 留学生向け情報探索ガイダンス

2.2.4 記載の総合図書館オリエンテーションの英語コースに協力するとともに、データベースユーザトレーニング、テーマ別ガイダンスの英語編を開催した。11 月には、「留学生向け情報探索ガイダンス」を 4ヶ国語で開催した。日本語コースを除き、外国人講師によって講習を行った。

コース名	月日	回数	人数	
総合図書館オリエンテーション	OPAC 入門英語編	4/12,4/16,4/18	3	31
	データベース講習会英語編	10/12,10/16	2	18
Web of Science 英語編	6/26,9/28,2/12	3	18	
Electronic Journals 英語編	5/22,7/6,12/7	3	10	
留学生向け情報探索ガイダンス	中国語コース	11/7	1	18
	韓国語コース	11/13	1	4
	日本語コース	11/16	1	2
	英語コース	11/22	1	3
合計		15回	104名	

2.2.6 事務職員向け情報探索ガイダンス

事務職員が業務をする上で役立つデータベースの紹介を行った。

月日	時間	人数	内容
8/2	11:00-12:00	4	辞書・事典ツール、新聞・雑誌記事、官報、統計、企業情報の検索方法や図書館サービスを紹介。
8/8	16:00-17:00	9	
合計	2回	13名	

2.2.7 出張講習会

利用者(教員、学生等)からの依頼に合わせて授業や研究室単位などで行う出張講習会を実施した。教養学部、文学部、理学部の出張講習会では、関係部局の図書館・室と連携して行なった。また、昨年度に引き続き、教養学部の基礎科目「基礎演習」の授業の1コマでの講習も担当した。

研究科・研究室名等	月日	時間	人数	内容	会場
学際情報学府 「学際情報学概論I」	4/10	11:00-11:45	98	・JapanKnowledge、OPAC、Web of Science、CiNii、電子ジャーナル等の利用方法	工学部2号館
教養学部 「基礎演習」	4/20	9:00-10:30	24	・情報検索の基礎の解説と実習	駒場 情報教育棟
	4/20	10:40-12:10	27		
	4/23	14:40-16:10	22	・JapanKnowledge、OPAC、Webcat Plus、CiNii 等の利用方法	
	4/24	10:40-12:10	22		
	4/27	13:00-14:30	24		
	5/7	14:40-16:10	25		
新規採用職員研修	5/10	17:00-17:30	36	・JapanKnowledge、聞蔵II、日経BP記事検索サービス等の利用方法	情報基盤センター演習室
文学部 “Academic Writing”	6/14	17:00-18:00	13	・JapanKnowledge、OPAC、Webcat、CiNii 等の利用方法 ・文学部図書室案内	総合図書館
	11/29	17:00-18:00	9		
理学部 生物学科 動物学コース・植物学コース	6/27	14:45-15:45	11	・JapanKnowledge、OPAC、Web of Science、電子ジャーナル、PubMed 等の利用方法 ・生物学図書室案内	総合図書館
工学系研究科社会基盤学専攻 交通・都市・国土学研究室	7/3	18:00-19:00	16	・JapanKnowledge、OPAC、Web of Science、電子ジャーナル、Engineering Village 等の利用方法	総合図書館
工学系研究科社会基盤学専攻 国際プロジェクト研究室	7/5	13:00-14:00	11	・JapanKnowledge、OPAC、Web of Science、電子ジャーナル、Engineering Village 等の利用方法	総合図書館
工学部システム創成学科	10/4	13:30-15:00	31	・JapanKnowledge、OPAC、Web of Science、	工学部14号館

				電子ジャーナル、Engineering Village 等の利用方法	
文学部社会心理学研究室「社会心理学調査実習」	11/8	10:00-11:00	11	・OPAC、Webcat、電子ジャーナル、CiNii、PsycINFO、Web of Science 等の利用方法 ・図書館ツアー	総合図書館
	11/8	11:00-12:00	12		
教養学部 後期課程「英語」	12/4	9:30-10:15	5	・LRC with MLA International Bibliography の利用方法	駒場図書館
合計	17回		397名		

2.2.8 セミナー・説明会

トライアルサービスを実施した科学技術文献情報を収録した二次情報データベース「JDreamII」の説明会を実習形式で行った。また、2007年5月にサービスを開始した Littel Navigator の説明会を、図書館職員を対象に開催した。両説明会ともデータベースの提供元から講師を招いて行なった。

講習会名	月日	時間	人数	会場
JDreamII 説明会	7/24	13:30-15:00	8(2)	総合図書館※
Littel Navigator 説明会	8/22	10:30-11:30	9	総合図書館
合計	2回		17名	

※柏図書館に中継したコース。人数欄に()で中継参加者を内数で示す。

図書館関係システム運用・管理

図書館情報係

1 運用報告

図書館情報係では、附属図書館学術情報システム及び附属図書館 Web サーバも含めた図書館サービスに係るサーバ(リポジトリと電子ジャーナル関係を除く)を管理・運用している。

以下、それらの 2007 年度の運用状況について報告する。

1.1 附属図書館学術情報システム

「附属図書館学術情報システム」は全学の図書館業務システムであると同時に、OPAC や Web リクエスト等の利用者サービスシステムとしての面も有している。2007 年度は業務システム面では学務システム、予算執行管理システムといった外部システムとの連携強化、サービス面では本システムに組み込んだ形で新たなブックコンテンツ DB の本番運用を開始している。

1.1.1 外部システムとの連携強化

図書館は大学の中でも早期から業務の電算化に取り組んできた部署である。そのため外部システムとの連携を行なわずとも業務が可能な完結した世界を持ちつづけて来た。例えばシステム独自の予算コード、利用者 ID、部署コード等の体系をシステム内に持ち、学内の他システムの状況によらず一通りの業務を行なえるようになっていく。

しかしながら大学の中で他の基幹的業務においても業務の流れの見直し、新たなシステムの立ち上げが法人化を契機になされるようになり、これらの(後発ではあるが)新しいシステムとの連携、大学として標準で使用されるコード体系等は、同じく大学の業務の一部である図書館システムとしても考慮せざるを得ない状況となっている。

このうち利用者 ID の共通 ID 化は前回のリプレースを機に行なったところであるが、2007 年度は学務情報(学生・院生データ)の図書館システムへの自動取り込みを実現することで進学時期の図書館利用者データの大量な更新事務を無くすとともに、本学の学生証を有する者は図書館窓口での利用者登録手続きを経ずとも、自動的に「図書館利用者」として登録される、という姿を目指し情報課とも協力してシステムの改修を行なった(本番稼働は 2008 年度中に実施予定)。

また 2008 年度から稼働する「予算執行管理システム」へ図書購入データの引き渡しが必要になってくることから、本部、メーカーと打合せを行い当面必要な機能について改修が行なわれている。

「予算執行管理システム」との連携は図書館業務システムにおける発注・受入といった機能の見直しの契機になるとも考えられ、今後とも図書館あるいは本部と連携しつつ業務の効率化を視野に改修や次期システムでのあり方を検討していく必要がある。

1.1.2 ブックコンテンツ DB の本番運用

2006 年度に改修設計、テストを行っていたブックコンテンツ DB と附属図書館学術情報システムとの統合であるが、2007 年度から本番運用を開始した。利用者側から見た機能は従来と大きな差異はないが、OPAC で目次情報も含めて検索可能となり、またコンテンツデータの生成も従来から比較するとかなり容易になった。

しかしながら附属図書館学術情報システムの有する書誌データを利用することとなったため、仕様検討段階では想定していなかった内部データのな不都合も生じている。今後この点の解決が課題である。

1.2 その他の図書館関係サーバ

図書館情報係では dl.itcドメインのサーバだけでなく、総合図書館に設置されている libドメインのサーバ群の管理も行なっている。

2007年度は附属図書館の「原簿サーバ」(法人化以前の図書原簿の情報を管理する)が電源系統の故障で使用不能となったため、新たなサーバ環境の構築を行なった。

また、下記で述べる図書職員向け講習会で使用するサーバ環境構築も図書館情報係が担当した。

2 講習会・研究会開催報告

図書館情報係が事務局となり、社会科学研究所図書職員の前田 朗氏、本研究部門の二宮 崇講師を講習会講師として「図書系職員のためのアプリケーション開発講習会」を実施した。

本講習会の目的は、図書職員の中でプログラミングやサーバ運用に関心を持つ人材を発掘し、講習会を通して各自の技量の向上を図り、講習会の成果物を図書館界(あるいは一般の利用者)に還元しようという試みである。

募集に対して 14名の応募があった。講習会は統一の課題は設けずに受講者各自が設定したテーマの実現に向けて講師陣がアドバイスをを行なうという形式で、9月から翌2月にかけて全体会を14回開催した。2008年度も継続の予定である。

2007年度の本講習会の成果は

<https://mbc.dl.itc.u-tokyo.ac.jp/products.html>

を参照されたい。

「図書系のためのアプリケーション開発講習会」成果

「図書系のためのアプリケーション開発講習会」は、東京大学情報基盤センターが主催する東京大学の図書系職員向けの講習会です

このページでは、講習会の受講者が開発したアプリケーションを公開しています。

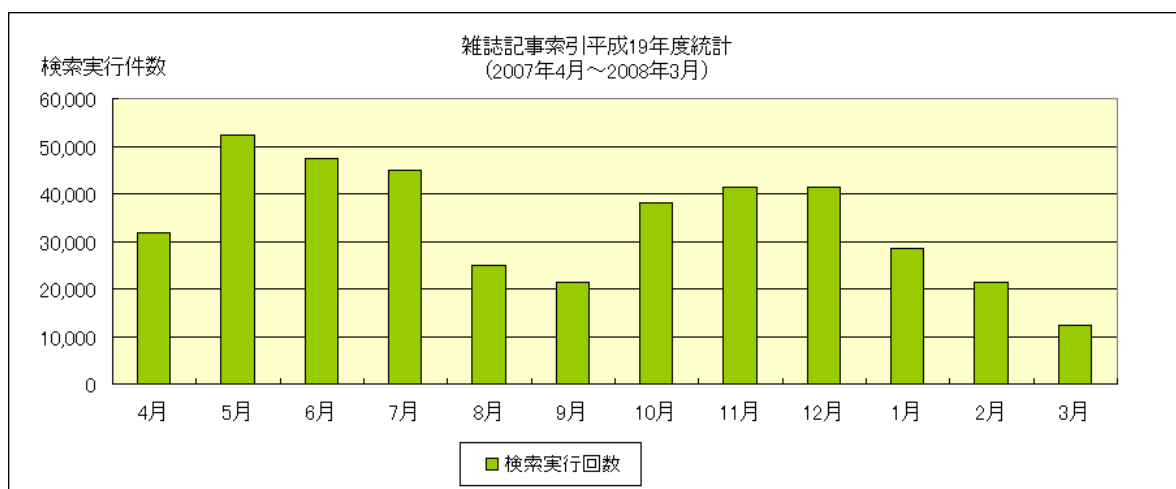
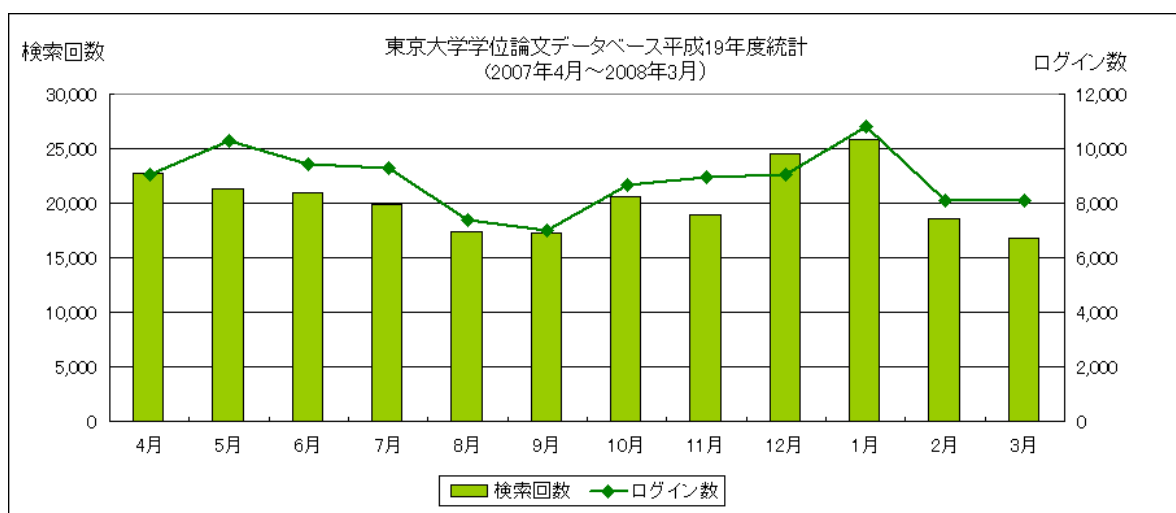
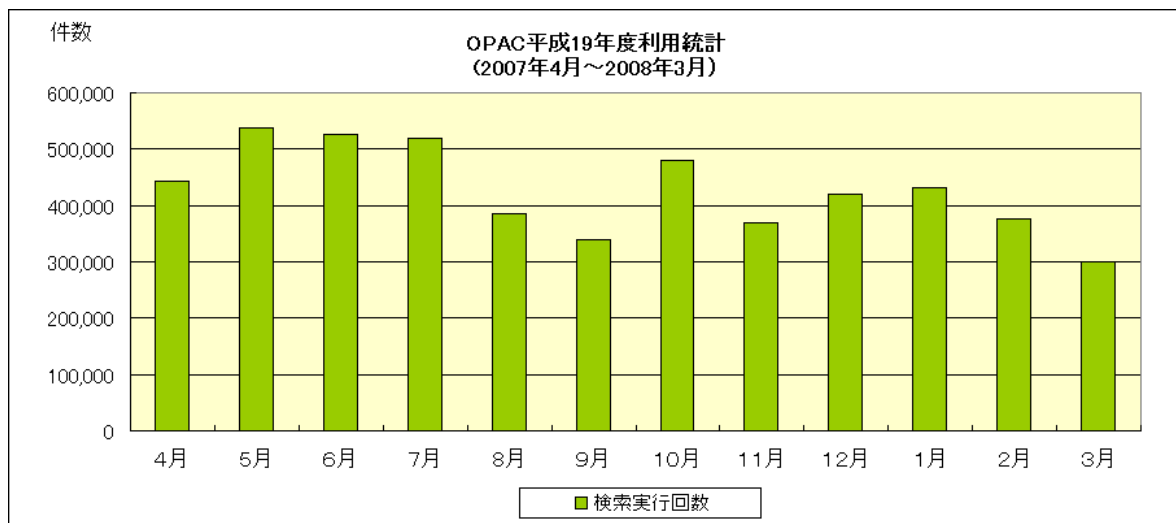
1. 利用者向けサービス

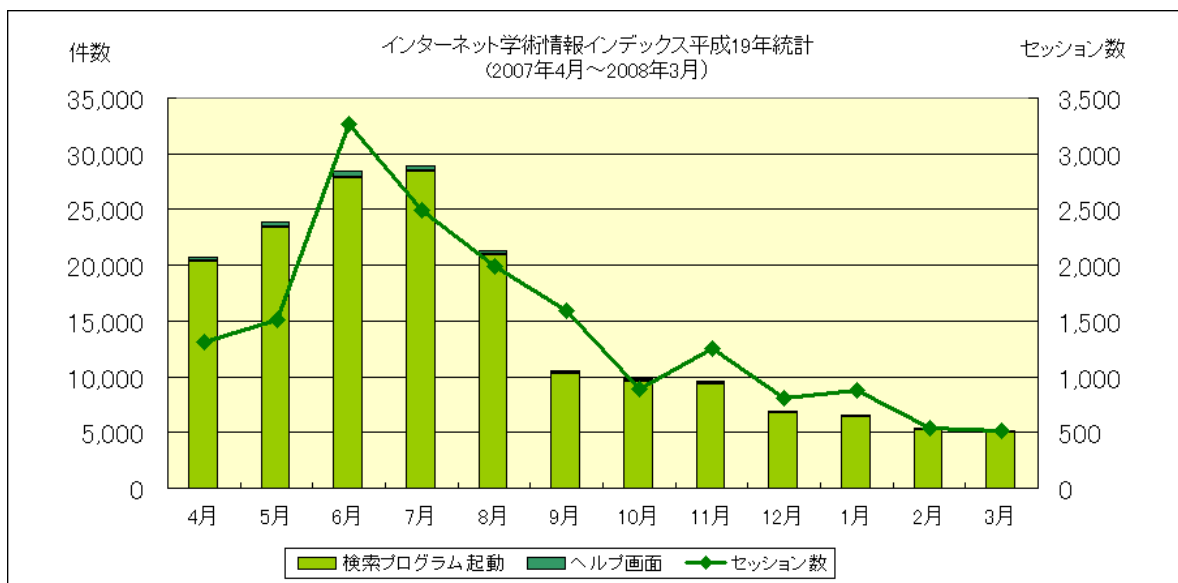
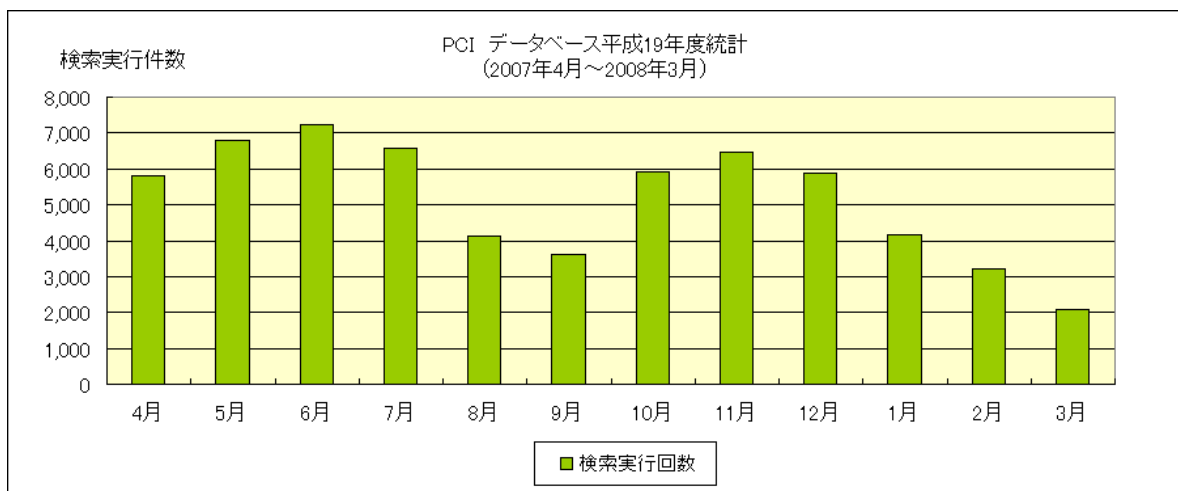
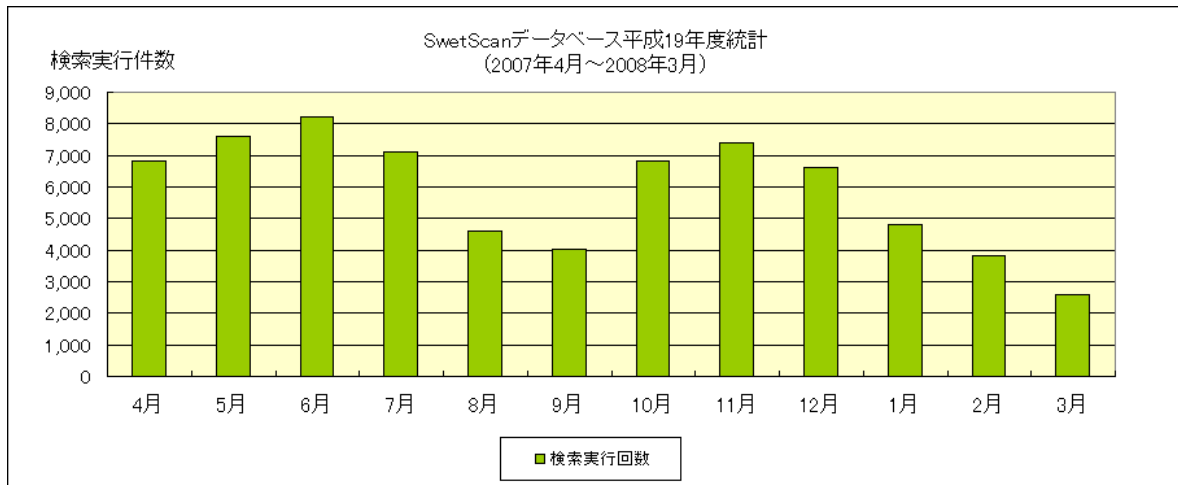
サービス及びリンク	解説	開発者
東京大学版 LibX	LibXは、Webブラウザ「Firefox」にも大学図書館向けの付加機能をつけるプラグインソフトです。「東京大学版LibX」は文字どおり、その東京大学版になります。Webブラウザと連携した東大OPACの特長機能、Amazonからの東大OPACリンク、UT Article Link 対応などさまざまな機能があります	手津 俊美
東京大学 OPAC の ジェット	ガジェットとは、デスクトップ上に常駐させる小範囲のアプリケーションのことです。東京大学OPACを特長できるガジェットの開発と配布を行っています。ガジェットには種類がいくつもありますが、まずはMacintoshガジェットボードから作成しました。	田口 忠祐
UT Article Search	東京大学のデータベースの検索検索=UT Article Search(学内限定)では、検索検索を行うデータベースを数多くのうちから選択する必要があります。このUT Article Searchは、使用するPCには検索対象データベースの選択結果を保持するものです。東京大学内限定のツールですが、プログラムのソースを変更することで他大学のCentral Searchにも対応できます。Adobe AIRにて実装を行いました。	前田 朗

2. 業務担当者向けツール(東京大学図書館職員限定)

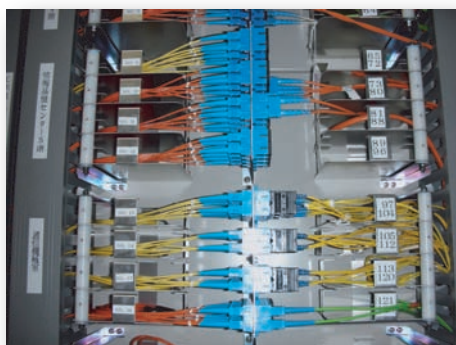
ツール名及びリンク	解説	開発者
全学図書館業務メモDB	図書館担当者の業務メモを管理、共有できます。オプション指定により検索対象を自図書館(室)限定と、他の学内図書館(室)を含めたものにも切り分けることができます。URLを変更しました。これにより、書籍ファイルワークオールからでもアクセス可能になっています(2008.03.25)	前田 朗
ISBN 3桁化変換	10桁のISBNを3桁に変換するWebアプリケーションです	小川 雅弘
雑誌図書館の重複チェック	雑誌図書館(ファイル形式)のデータと、所蔵データをもとに、これから雑誌する資料が既に受け入れられているかどうかを、チェックするためのWebアプリケーションです	小川 雅弘
貸出冊数集計	図書館システムによる貸出データから、自図書館(室)と学内の他の館室にわけて貸出冊数を集計します	戸田 宗洋
e-ODS 予算集計	e-ODS サービスの予算集計を出すWebアプリケーションです	前田 朗
資料IDチェックディジット計算	複数冊の資料IDのチェックディジット計算や検証を行います。また範囲指定により一括で資料IDを生成することも可能です	小林 幸彦
附属図書館FAX着信ニュースのメール自動配信(試行)	附属図書館FAXに到着したニュースが掲載された、くく(最大2時間以内)、到着ニュース内容を電子メール配信します。なお、利用には申請(電子メールアカウントの登録)が必要です。	本多 玄

3 サービス統計

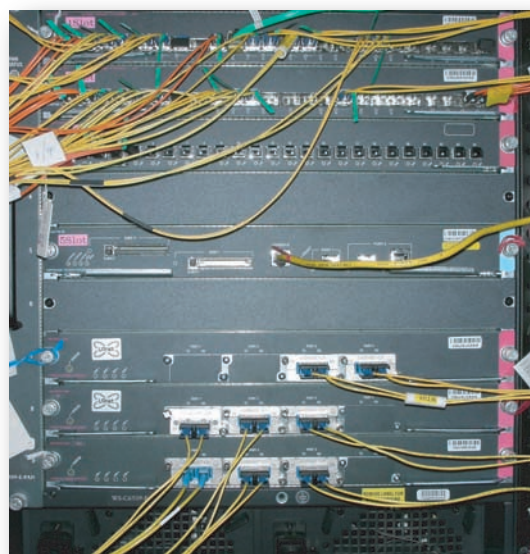




キャンパスネットワーク部門



光ケーブルが集線されているスプライシングボックス



UTnet3の代表的な機器(レイヤ3スイッチ)

キャンパスネットワークキング部門

概要

早野裕士

■東京大学キャンパスネットワークシステム(UTnet3)の運用管理

本学のキャンパスネットワークシステム UTnet3(University of Tokyo network system 3)は、各建物内の支線ネットワーク(支線)及び建物間接続や学外との接続のための基幹ネットワーク(基幹)とから構成されている。情報基盤センターの本キャンパスネットワークキング部門は、以下に示す通り、基幹の運用管理に加え、これに関連する業務も含めて取り組んだ。

1. 基幹ネットワークの運用管理

基幹の中核を占めるネットワーク機器は、本郷地区、駒場Ⅰ地区、駒場Ⅱ地区、柏地区、白金地区、中野地区の各ハブサイトに設置したギガビットスイッチ(L3)であり、基幹の基本トポロジーは情報基盤センターを中心にしたスター型になっている。基幹に対するトラフィック需要の増加に対応するため、ギガビットスイッチの転送処理ボードの増強、回線速度の高速化等を実施した。また遠隔研究施設については本郷地区との接続性を確保しているが、同様にトラフィック需要に応じた増強を実施した。

2. 基幹ネットワークと支線ネットワークとの接続

各建物にレイヤー2 スwitchを設置し、基幹のギガビットスイッチ(L3)と支線を接続しているが、ネットワーク利用の増大への対応と高機能化のため、2007 年度も 2006 年度に引き続き、レイヤー2 スwitchの順次更新を行った。

3. VLAN 対応

UTnet3 では仮想 LAN(VLAN: Virtual LAN)に対応している。VLAN によって、部局や研究室が複数の建物やキャンパスに分散して配置されている場合でも部局や研究室のまとまりごとに同一のサブネットに収容することを可能としている。本学では、建物の新設や組織変更等が少なくなく、それに伴った VLAN に関する要求は非常に多く、それら要求に応じてネットワーク機器の設定変更や増設等を実施した。

4. キャンパス間接続、及び対外ネットワーク接続

キャンパス間接続、及び学外への接続については、ギガビットスイッチを設置し、SINET3 をはじめとした広帯域回線による接続を行っている。2007 年度は、接続回線速度の高速化や高機能化を実施した。

4. 無線 LAN サービス

山上会館、武田先端知ビル内の武田ホール、弥生講堂・一条ホールに加えて、2008 年 3 月から柏地区の柏図書館内メディアホールで無線 LAN サービスを開始した。

5. その他関連業務

UTnet3 基幹ネットワークの運用管理に関連する業務として、基幹のギガビットスイッチ(L3)を設置している HUB サイト内に設置している空調機の老朽化に伴う更新、ならびに、全学法定点検による停電の回避策としての無停電電源装置(UPS: Uninterruptible Power Supply)の設置、ネットワーク機器監視業

務の改善、ドメイン名の割り当て、DNS(Domain Name System)のサービス、UTnet 光ファイバケーブル専用利用、及び無線 LAN サービスの各業務に取り組んだ。

■セキュリティ対応

最近のネットワークにおいてはセキュリティへの対応が必須となっており、本学においてもセキュリティインシデントは日常茶飯事と言っても過言ではないほどに頻繁に発生している。特にコンピュータへの不正侵入の試み、電子メールや Web ページを通してのウィルス感染、P2P(Peer to Peer)、SPAM メール等が問題となっている。そこで、本学のセキュリティの維持確保に向け、以下の通り、セキュリティ対応業務を実施した。

1. 不正侵入の監視

学外と学内との間の通信に関し、基幹部分において不正侵入の試みを検知し事前に予防する仕組みとして、UTnet3 では不正侵入検知装置 (IDS: Intrusion Detection System)と異常トラフィック監視システムを導入して監視し、その結果に応じて必要な対処を実施した。

2. ウィルス感染への対応

メールや Web ページ等を通してのウィルス感染に関する対策として、情報基盤センターでは、適切なウィルス対策ソフトウェア製品を学内利用者の希望に応じて配布し、予防に役立てている。2007 年度は、配布製品を拡大するとともに、利用者の増大を図って、ウィルス対策を積極的に進めた。

3. SPAM 対策

全学的に激増する SPAM メールに対して、本部事務組織では全学的な対策が検討されていて、情報基盤センターでは、製品の評価や運用テスト等を協力した。

■ 東京大学情報システム緊急対応チーム(UT-CERT)との連携

UTnet3 ではセキュリティに関する対応として、基幹部分のギガビットスイッチ(L3)で、全学的に問題となる事象や各部局の個別の事象についてのフィルタを設定している。東京大学情報システム緊急対応チーム(UT-CERT)から委託業務を受け、このフィルタを活用することによって、全学的なセキュリティ対策を行った。

■ 学内ソフトウェアライセンス

学内に利用者の多いソフトウェアについて、全学サイトライセンスを取得し、そのライセンス管理と配布サービスを行った。2007 年度には、JMP を有料化して運用開始と CHEMOFFICE を新規に運用開始した。

■ 関係委員会

キャンパスネットワークに関する事項について、全学的視点から企画、立案及び審議を行う情報基盤センターキャンパスネットワーク専門委員会が下記のとおり行われた。

2007 年 5 月 8 日 第 26 回情報基盤センターキャンパスネットワーク専門委員会

2007 年 7 月 17 日 第 27 回情報基盤センターキャンパスネットワーク専門委員会

2007 年 10 月 9 日 第 28 回情報基盤センターキャンパスネットワーク専門委員会

2008 年 1 月 16 日 第 29 回情報基盤センターキャンパスネットワーク専門委員会

主な検討事項および承認事項

- ・基幹ネットワークの構成変更について
- ・UTnet 部局負担金について
- ・ソフトウェアライセンスの負担金について

- 10G 接続の負担金について
- 光ファイバケーブル専用利用の業務利用について

東京大学キャンパスネットワークシステム(UTnet3)の運用管理

中村昇平
松本浩一

佐島浩之

佐々木馨

1 運用報告

本学のキャンパスネットワークシステムは UTnet(University of Tokyo network system)と称し、これまでの更新経緯に応じて現在は UTnet3 と呼ばれている。UTnet3 は支線ネットワーク(支線)と基幹ネットワーク(基幹)とから構成されている。支線は、各建物内に設置されたネットワークで、当該部局によって運用管理されている。基幹は、支線の相互接続及び学外との接続のために設置されたネットワークで、情報基盤センターの本キャンパスネットワークキング部門が運用管理している。本部門では、以下に示す通り、基幹ネットワークの運用管理に加え、これに関連する業務も含めて取り組んだ。

1.1 基幹ネットワークの運用管理

本郷地区の基幹には、5 箇所のハブサイト(情報基盤センター、電話庁舎、附属図書館、工学部 8 号館、農学部 3 号館)があり、各ハブサイトにはギガビットスイッチ(L3)を設置し、基幹は情報基盤センターを中心にしたスター型の構成としている。駒場 I 地区、駒場 II 地区、柏地区、白金地区、中野地区の各郊外地区についても、各ハブサイトにギガビットスイッチ(L3)を設置し、専用回線等を介して本郷地区と接続している。このような基幹の運用管理について、主に以下の取り組みを行った。

- －白金区、柏地区のギガビットスイッチ(L3)の CPU ボードの増強
- －駒場 II 地区の HUB サイトの停電対策としての無停電電源装置の導入
- －附属図書館、工学部 8 号館、農学部 3 号館の HUB サイトの空調機の更新

以上の増強、安定化運用の対策等を実施し、基幹に対する学内からの要求に応えることによって、本学の研究・教育の一層の円滑な推進を図った。基幹のネットワーク構成を図 1 に示す。また、本郷地区と、駒場 I 地区、駒場 II 地区、柏地区、白金地区、中野地区の各キャンパスの間のトラフィック量の推移を図 2 に示す。

1.2 基幹ネットワークと支線ネットワークとの接続

基幹と各支線の接続については、各建物にレイヤー 2 スイッチを設置し、このスイッチとハブサイト内のギガビットスイッチ(L3)との接続で実現している。UTnet3 導入当初にはレイヤー 2 スイッチを約 200 台設置したが、ネットワーク利用の増大への対応と高機能化のため 2005 年度より順次更新作業を開始した。これまでに約 160 台の更新を行い、2007 年度には、更に 21 台のレイヤー 2 スイッチの更新を行った。

1.3 VLAN 対応

本学では、部局や研究室が、複数の建物やキャンパスに分散して配置されている場合及び他部局の建物の一部に入室する場合が増えており、これに伴って、部局や研究室のまとまりごと同一のサブネットを利用できる建物間仮想 LAN(Virtual LAN:VLAN)の申請が増えてきている。VLAN に関して 2007 年度にあった申請に伴って実施した、基幹ネットワークの構成変更およびサブネットワークの割当は表 1 の通りである。

1.4 キャンパス間接続、及び対外ネットワーク接続

以下に示す通り、学内キャンパス間接続及び対外ネットワークとの接続に関して、変更作業を実施し、これによって、各接続に関して高機能化・高速化を達成した。

- 2007 年 9 月に、WIDE(Widely Integrated Distributed Environment) との接続を、1Gb/s から 10Gb/s に変更した。
- 駒場 I 地区、駒場 II 地区と本郷地区(情報基盤センター)間のキャンパス間回線を 1Gb/s から 10Gb/s に変更するためのテスト等を行った。

キャンパス間回線のトラフィック量の推移は図 2 の通りであった。

1.5 その他関連業務

(1) ドメイン名の割り当て

2007 年度に申請があり、キャンパスネットワーク専門委員会で承認されたドメイン名の割当を行った。内訳は、新規が 5 件で、詳細を表 2 に示す。

(2) UTnet 光ファイバケーブル専用利用

本学では建物間で光ファイバを専用利用することが可能で、2007 年度に行った UTnet 光ファイバケーブル専用利用の変更及び割り当てを表 3 に示す。

(3) 無線 LAN サービス

2007 年度の山上会館、武田先端知ビル内の武田ホール、弥生講堂・一条ホールにおける無線 LAN サービスの各部局別の申請利用状況を表 4 に示す。また、2008 年 3 月から柏地区の柏図書館内メディアホールで無線 LAN サービスを開始した。

2 講習会・研究会開催報告

[第 4 回 UTnet Meeting]

最近のネットワークの動向や管理の問題点に関して情報交換を行うための集まりとして開催している UTnet Meeting を、2007 年 10 月 25 日に理学部 1 号館の小柴ホールで行った。

プログラムは次の通りであり、参加者は、情報基盤センターの関係者を含め 67 名であった。

- | | |
|-----------------------------------|-------------------|
| 1. UTnet 報告 | 椿山惣一郎 |
| 2. SINET3 の提供サービスについて | 国立情報学研究所 下田哲郎 |
| 3. UT-CERT 報告 | 早野裕士 |
| 4. パブリックサーバ証明書発行プロジェクトのご案内 | 西村 健 |
| 5. 『Web からの脅威』の現状とトレンドマイクロの対応について | トレンドマイクロ株式会社 平原伸昭 |
| 6. フリーディスカッション | |

[平成 19 年度東京大学技術職員研修(コンピュータ関係)]

本学職員のうち、研究室やセンター等においてコンピュータやコンピュータ・ネットワークの初心者レベルの利用者、及びそれらの運用管理業務やセキュリティ対策業務に携わる初心者レベルの管理者を対象に、セキュリティ対応を含むコンピュータやコンピュータ・ネットワークの利用に関する研修を、以下の通り実施した。参加者は約 20 名であった。

- ・期間
平成 19 年 11 月 27 日(火)～平成 19 年 11 月 30 日(木) (3 日間)
- ・場所
情報基盤センター 4 階講義室

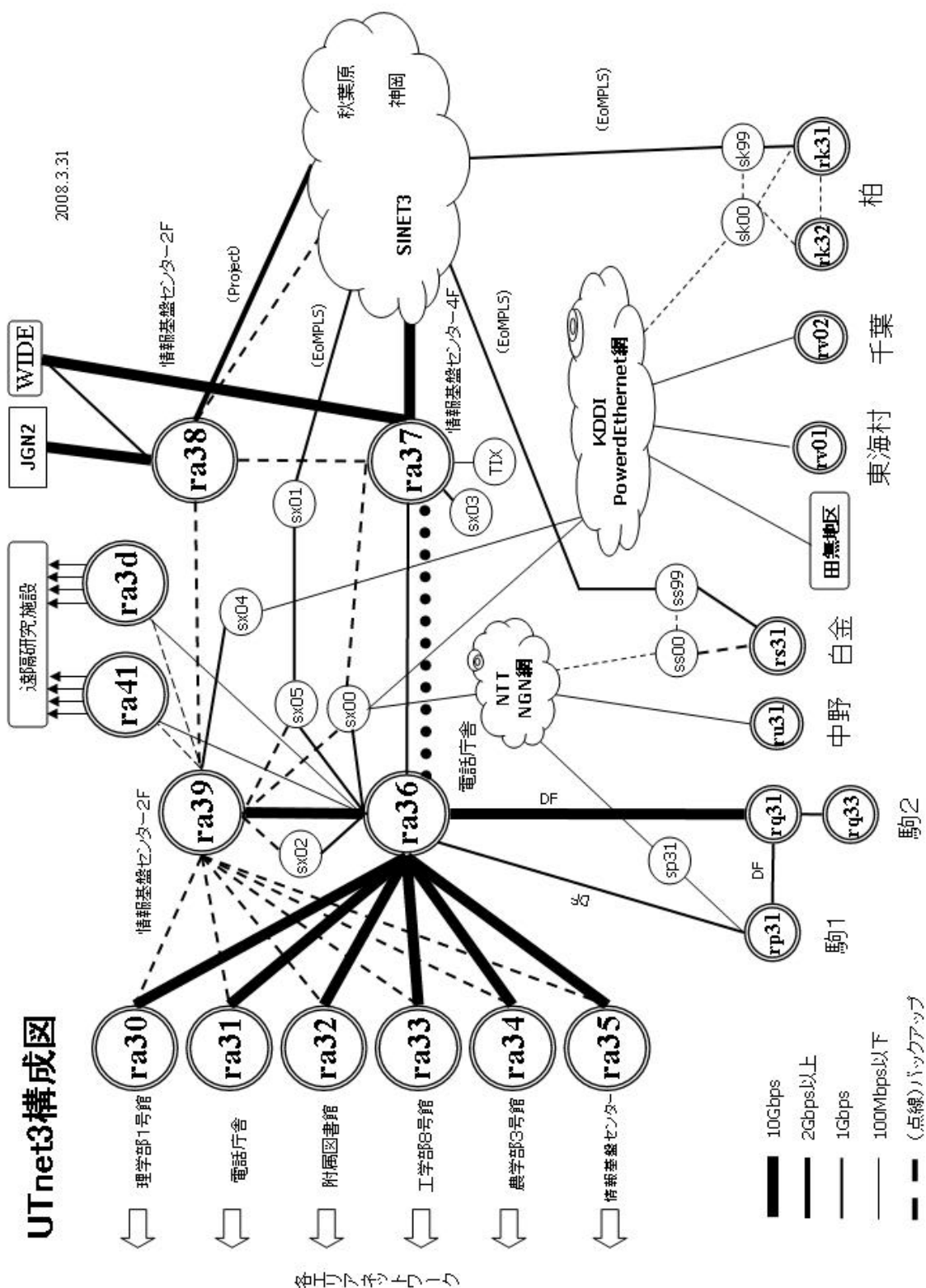
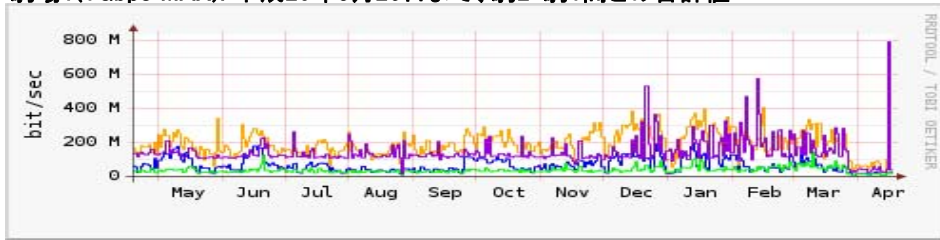


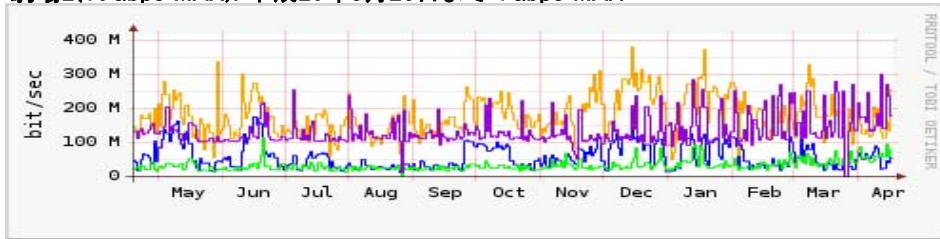
図1 UTnet 基幹構成図

キャンパス間回線トラフィック

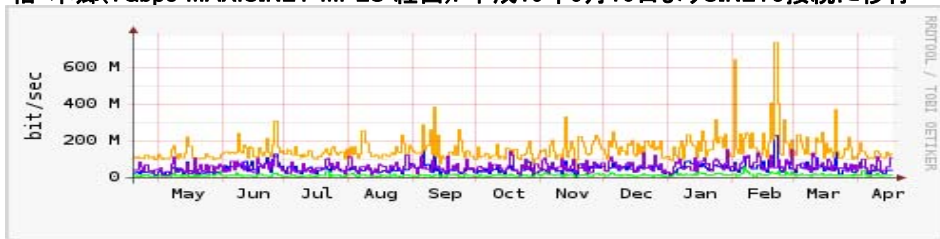
駒場1(1Gbps MAX): 平成20年3月26日まで、駒2-駒1間との合計値



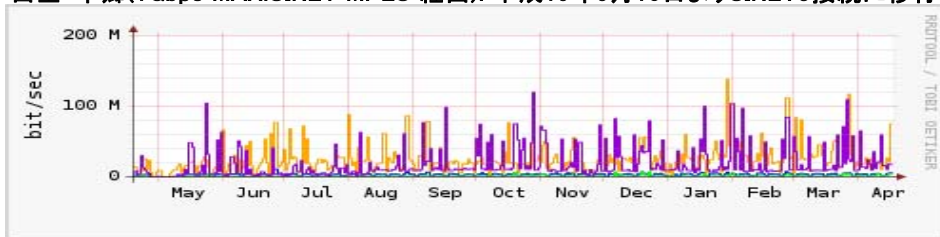
駒場2(10Gbps MAX): 平成20年3月26日まで 1Gbps MAX



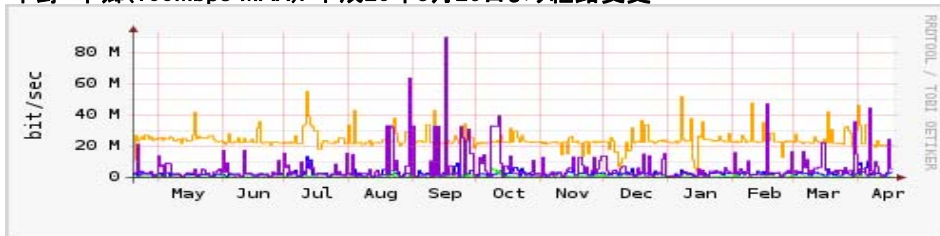
柏-本郷(1Gbps MAX:SINET MPLS 経由): 平成19年5月10日よりSINET3接続に移行



白金-本郷(1Gbps MAX:SINET MPLS 経由): 平成19年5月10日よりSINET3接続に移行



中野-本郷(100Mbps MAX): 平成20年3月26日より経路変更



- GREEN ### 1日平均トラフィック(本郷←各キャンパス)
- BLUE ### 1日平均トラフィック(本郷→各キャンパス)
- VIOLET### 1日最大トラフィック(本郷←各キャンパス)
- ORANGE### 1日最大トラフィック(本郷→各キャンパス)

図2 UTnet のキャンパス間トラフィック

表 1 基幹ネットワークの構成変更およびサブネットワークの割当

項番	申請部局名	設置場所	変更内容
1	情報基盤センター	工学部新2号館	建物間 VLAN の申請
2	新領域創成科学研究科	理学部プレハブA棟	基幹ネットワークの構成変更及びサブネットアドレスの申請
3	情報理工学系研究科	秋葉原	サブネットアドレスの申請
4	本部事務組織	情報基盤センター	建物間 VLAN の申請
5	先端科学技術研究センター	先端研13号館、14号館	建物間 VLAN の申請
6	史料編纂所	史料編纂所-赤門総合棟	建物間 VLAN の申請
7	情報理工学系研究科	工学部新2号館	建物間 VLAN の申請
8	情報理工学系研究科	情報基盤センター	建物間 VLAN の申請
9	本部事務組織	情報基盤センター	基幹ネットワークの構成変更
10	医学系研究科	医学部1, 2, 3号館、総合中央館、動物実験施設	建物間 VLAN の申請
11	地震研究所	地震研1号館、理学部1号館	建物間 VLAN の申請
12	工学系研究科	工学部8号館	建物間 VLAN の申請
13	本部事務組織	アントレプレナープラザ、山上会館別館、工学部列品館	建物間 VLAN の申請
14	情報基盤センター	情報基盤センター	建物間 VLAN の申請
15	医学系研究科	医学部1号館	基幹ネットワークの構成変更
16	保健センター	柏福利厚生新棟	基幹ネットワークの構成変更およびサブネットアドレスの申請
17	インテリジェント・モデリング・ラボラトリー	山上会館	基幹ネットワークの構成変更およびサブネットアドレスの申請
18	素粒子物理国際研究センター	情報基盤センター	基幹ネットワークの構成変更
19	情報理工学系研究科	工学部8号館	基幹ネットワークの構成変更およびサブネットアドレスの申請
20	工学系研究科	武田先端知ビル	サブネットアドレスの申請
21	情報基盤センター	情報基盤センター3F	建物間 VLAN の申請
22	情報基盤センター	情報基盤センター5F	サブネットの返却
23	情報基盤センター	情報基盤センター3F	建物間 VLAN の申請
24	本部事務組織	柏Ⅱキャンパス管理棟	基幹ネットワークの構成変更およびサブネットアドレスの申請
25	本部事務組織	工学部10号館	建物間 VLAN の申請
26	先端科学技術研究センター	先端研4、56号館	建物間 VLAN の申請
27	本部事務組織	産学連携プラザ	基幹ネットワークの構成変更
28	情報基盤センター	駒1情報教育棟	建物間 VLAN の申請
29	情報基盤センター	医科学研究所1号館	建物間 VLAN の申請
30	情報基盤センター	経済学部新棟	建物間 VLAN の申請
31	附属病院	附属病院-IML	光専用利用の廃止
32	本部事務組織	本部管理棟	基幹ネットワークの構成変更
33	医科学研究所	医科研総合研究棟	建物間 VLAN の申請
34	柏地区事務部	柏地区	サブネットアドレスの申請
35	生産技術研究所		光専用利用の廃止
36	情報基盤センター	情報基盤センター5F	建物間 VLAN の申請
37	情報基盤センター	情報教育新棟	基幹ネットワークの構成変更及びサブネットアドレスの申請
38	本部事務組織	産学連携プラザ	基幹ネットワークの構成変更およびサブネットアドレスの申請

39	総合教育研究センター	薬学部本館	建物間 VLAN の申請
40	情報基盤センター	工学部船舶試験水槽	建物間 VLAN の申請
41	生産技術研究所		光専用利用の廃止
42	情報基盤センター	情報基盤センター	サブネットアドレスの申請
43	情報基盤センター	情報教育新棟	建物間 VLAN の申請
44	保健センター	柏基盤科学棟	建物間 VLAN の申請
45	工学系研究科	工学部8, 9号館	基幹ネットワークの構成変更
46	本部事務組織	アントレプレナープラザ	基幹ネットワークの構成変更及びサブネットアドレスの申請
47	本部事務組織	情報基盤センター、山上会館龍岡門別館	建物間 VLAN の申請
48	本部事務組織	情報基盤センター、法文1号館(法学部)、教養学部アドミニストレーション棟、数理科学研究科棟	建物間 VLAN の申請及びサブネットアドレスの申請
49	総合文化研究科	教養学部9号館	基幹ネットワークの構成変更
50	史料編纂所	史料編纂所、福武ホール	建物間 VLAN の申請
51	情報理工学系研究科	工学部2, 9号館	建物間 VLAN の申請
52	新領域創成科学研究科	工学部12号館	基幹ネットワークの構成変更及びサブネットアドレスの申請
53	工学系研究科	工学部12号館	サブネットアドレスの申請
54	分子細胞生物学研究所	インテリジェント・モデリング・ラボラトリー棟	基幹ネットワークの構成変更及びサブネットアドレスの申請
55	本部事務組織	東大本郷けやき保育園	建物間 VLAN の申請

表2 ドメイン名の割当

項番	ドメイン名	申請部局名	備考
1	cbri.u-tokyo.ac.jp	本部事務組織	生物機能制御化合物ライブラリー機構
2	oa.u-tokyo.ac.jp	大学院工学系研究科	海洋アライアンス
3	cks.u-tokyo.ac.jp	大学院工学系研究科	知の構造化センター(ネットワーク)
4	cirp.u-tokyo.ac.jp	本部事務組織	総長室統括委員会
5	ipmu.u-tokyo.ac.jp	柏地区事務部	数物連携宇宙研究機構

表3 Utnet 光ファイバケーブル専用利用の変更及び割当

項番	申請部局	利用区間	種類	
1	先端科学技術研究センター	先端研4号館ー先端研13号館、14号館	SM 4 芯	新規
2	情報基盤センター	情報基盤センターー理学部7号館	SM 2 芯	新規
3	生産技術研究所	駒II設備センターー生研本館	SM 2 芯	廃止
4	医学部附属病院	附属病院管理研究棟ーインテリジェント・モデリング・ラボラトリー棟	GI 2 芯	廃止
5	生産技術研究所	駒場リサーチキャンパスー生研本館	GI 2 芯	廃止
6	大学院工学系研究科	工学部6号館ー情報基盤センター	SM 2 芯	新規
7	先端科学技術研究センター	先端研4号館ー国際・産学共同研究センター	GI 2 芯	新規
8	生産技術研究所	生研電算室ー先端研4号館	SM 2 芯	廃止
9	地震研究所	情報基盤センターー地震研究所1号館	GI 2 芯	新規
10	情報基盤センター	教養情報教育棟ー教養15号館	SM 2 芯	新規
11	大学院新領域創成科学研究科	情報基盤センターー工学部新2号館	SM 2 芯	新規

表 4 ネットワーク利用の申請部局・組織別申請件数

項番	申請部局・組織	山上会館	武田ホール	弥生講堂	柏図書館
1	サスティナビリティ学連携研究機構		1		
2	地球観測データ統融合連携研究機構			1	
3	数物連携宇宙研究機構				1
4	GCOE 化学イノベーション			1	
5	国際連携本部		1	1	
6	産学連携系	1			
7	国際系	1			
8	人事・労務系			1	
9	生物生産工学研究センター			1	
10	情報基盤センター		1	1	
11	医学系研究科	1			
12	工学系研究科	11	8		
13	理学系研究科	1			
14	農学生命科学研究科			5	
15	新領域創成科学研究科	3	1		
16	情報理工学系研究科	2	1		
17	情報学環	1	4		
18	地震研究所			2	
19	東洋文化研究所	1			
20	生産技術研究所		1		
21	史料編纂所	1			
22	物性研究所	1			
23	宇宙航空研究開発機構	1			
24	日本公衆衛生学会		1		
25	理化学研究所		1		
26	日本医科大学			1	
27	京都大学			1	
28	NPO 法人太陽光発電所ネットワーク			1	
29	日本農学図書館協議会			1	
30	池谷コンファレンス組織委員会			1	
31	日本繁殖生物学会			1	
32	東京学芸大学養護教育講座			1	
33	科学技術振興機構			1	

セキュリティ対応

早野裕士

椿山惣一郎

佐野徹

坂井朱美

1 運用報告

最近のネットワークにおいては、高速な接続性の実現とともに、セキュリティへの対応も重要となっている。本学においてもセキュリティインシデントは日常茶飯事と言っても過言ではないほどに頻繁に発生している。特にコンピュータへの不正侵入の試み、電子メールや Web ページを通じてのウイルス感染は問題で、感染を受けた側の被害はもとより、学内や学外への不正アクセスの踏台となることも珍しくない。しかし、一方で、セキュリティを厳重にすることは利用者の利便性の低下にもつながる。一般論としていえば、セキュリティ対策の要点は安全性と利便性のトレードオフといえるが、本学のように多種多様な部局が存在している環境では、このトレードオフのバランス点を一つに収束させることは容易ではない。このような配慮のもと、本学のセキュリティの維持確保に向け、以下の通り、セキュリティ対応業務を実施した。

1.1 不正侵入の監視

本学内のコンピュータに対して不正なアクセスを試みている可能性を検出し、必要ならば当該コンピュータが収容されているネットワークの管理者と連絡をとって不正侵入を防ぐべく対応を行った。このような不正アクセスの検知には、以下の2システムを活用した。

- 不正侵入検知システム (IDS :Intrusion Detection System)

これは、通常のネットワークアクセスでは発生しないようなアクセスパターン (例えば多数のホストを順にスキャンしていく動作など)を対象として検知し、警告を発するシステムである。実際には、マカフィー社の IntruShield 4000 を導入し、学内と学外との通信をタップし IDS に接続して監視した。また、5月より P2P 通信の監視を追加して行うようになり、P2P 通信情報を各部局のセキュリティ担当者宛に電子メールでの連絡を開始した。

- 異常トラフィック監視システム

これは、サービス妨害攻撃またはサービス不能攻撃などと呼ばれる、インターネット経由での不正アクセスの 1 つで、大量のデータや不正パケットを送りつける行為などの不正な攻撃を検知できるシステムである。実際には、沖電気工業社製 Secure Traffic Probe (ソフトウェア)を導入し、学内と学外との通信をタップし異常トラフィック監視システムに接続して、ポート単位での異常なトラフィックの変化について監視した。

1.2 ウイルス対策

Windows パソコンやサーバマシン等における、ウイルスの脅威と感染被害を未然に防ぐ有効な手段として、コンピュータウイルス対策ソフトウェアがある。情報基盤センターでは、昨年度に引き続きトレンドマイクロ社の製品についてサイトライセンス契約を拡大継続することによって、学内への安価な配布サービスを推し進めた。また、4月より、Mac OS (Intel 版 Mac) 対応として Sophos Anti-Virus for Mac、及び多言語対応 (基本的には留学生向け)として、Windows OS の簡体中国語と繁体中国語用の Sophos Anti-Virus for Windows の正式運用をあらたに開始した。コンピュータウイルス対策ソフトウェア配付の 2007 年度の申請状況は表1の通りである。

1.3 迷惑メール(SPAMメール)対策

学外から大量の迷惑メール(SPAMメール)が送信され、学内に設置されている複数のメールサーバに負荷がかかり、研究・業務等に支障をきたすような状況が発生してきた。そのため、全学的な迷惑メール対策を行うための事前の性能評価を行うことにし、10月より複数の業者から迷惑メール対策装置を借用し、さまざまな評価を行い次年度から試験運用を実施するための準備を行った。

表1 サイトライセンスソフトウェアの申請状況(2007年度)

ソフトウェア名	申請件数	部局数	申請台数
ウイルスバスターコーポレートエディション	11	8	662
ウイルスバスター2008(日本語版)	523	43	23,348
ウイルスバスター2008(英語版)	82	24	851
Sophos Anti-Virus for Windows	10	8	59
Sophos Anti-Virus for Mac OSX	21	13	193
Server Protect	51	20	233
InterScan VirusWall	15	11	33
InterScan VirusWall エンタープライズエディション	4	3	10

2 講習会・研究会開催報告

(東京大学キャンパスネットワークシステム(UTnet3)の運用管理の2.を参照)

東京大学情報システム緊急対応チーム(UT-CERT)との連携

早野裕士

椿山惣一郎

佐野徹

坂井朱美

1 運用報告

情報基盤センターでは、UTnet3 の基幹ネットワークでセキュリティ対策を実施していることから、ネットワークを介したセキュリティインシデントについて、東京大学情報システム緊急対応チーム UT-CERT(University of Tokyo-Computer Emergency Response Team)から以下の委託業務を受け、UT-CERT と連携協力しながら、全学的なセキュリティ対策を実施した。

- セキュリティ関連情報の収集と学内への注意喚起等
 メーカーやセキュリティ情報サイト等から、セキュリティの脆弱性やウイルスに係わる情報を収集し、情報基盤センターの web ページへの掲載や部局 CERT へのメールによって、セキュリティ対応に関し注意喚起を行った。
- 学内と学外との通信におけるネットワークを介したインシデントの監視
 IDS や異常トラフィックシステムで学内と学外との通信を監視し、セキュリティインシデントの検知を行った。
- セキュリティインシデントの部局 CERT への連絡
 学内と学外との通信の監視や UT-CERT への学内外からの連絡等によって判明したセキュリティインシデントについて、該当する部局 CERT 担当者へメールで対処を依頼した。また、インシデントレポートシステムでの報告を求めた。
- P2P 通信情報の部局 CERT への連絡
 P2P の通信の監視や通信情報を、該当する部局 CERT 担当者へメールで連絡を行った。また、部局 CERT からの P2P 通信に対する詳細情報提供等の依頼に関する対応を行った
- インシデントレポートシステムの運用管理
 部局 CERT から、発生したセキュリティインシデントの内容や対処等を定型的な書式で報告できるように、インシデントレポートシステムを開発し、運用管理した。
- インシデントレポートの集計
 部局 CERT から報告されたインシデントレポートについて、毎月集計し UT-CERT のホームページ (<http://park.itc.u-tokyo.ac.jp/ut-cert/>) に掲載した。(表1参照)
- セキュリティインシデントの当該 IP アドレスのネットワーク遮断
 部局 CERT に連絡したセキュリティインシデントのうち、発見時点から事象が連続している場合は、緊急措置として、当該 IP アドレスのホストによる通信を UTnet 機器でネットワーク遮断した。
- セキュリティインシデントの部局 CERT の対策への連携協力
 部局 CERT に連絡したセキュリティインシデントのうち、具体的な対処方法等について、部局 CERT から UT-CERT に協力の依頼があった場合、UT-CERT 及び部局 CERT と連携協力して対応した。

2 講習会・研究会開催報告

(東京大学キャンパスネットワークシステム(UTnet3)の運用管理の 2. を参照)

3 サービス統計

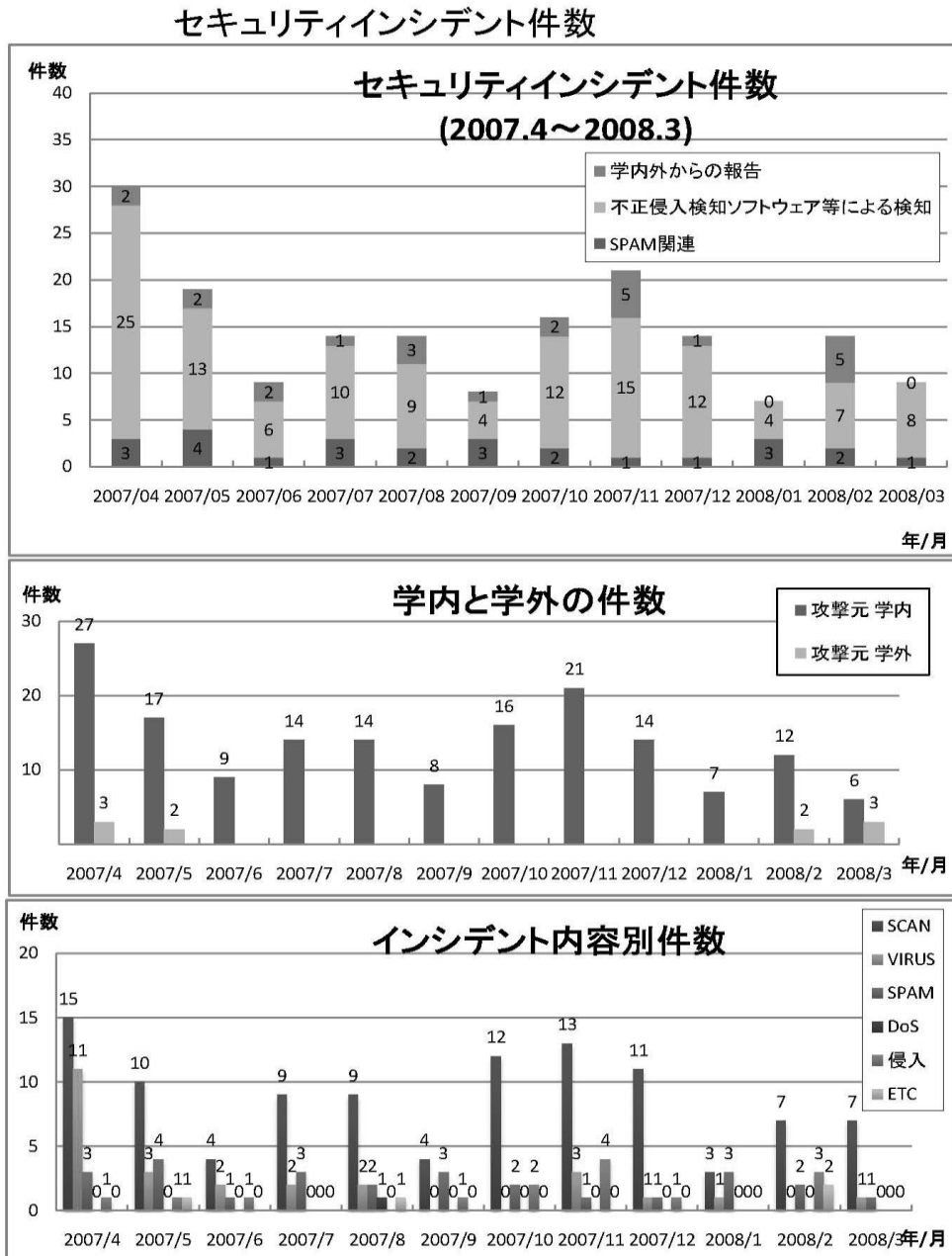


表 1 2007 年度の学内のセキュリティインシデント件数

学内ソフトウェアライセンス

早野裕士

椿山惣一郎

佐野徹

坂井朱美

1 運用報告

学内での研究教育の円滑な推進を達成することを目的として、利用が多いソフトウェアのうち、可能な範囲でサイトライセンス契約を提供業者との間で結び、安価でかつ容易に利用できるよう図っており、具体的には、以下の各種ソフトウェアの配布サービスとライセンスの管理を行った。

(1)ウィルス対策ソフトウェア

Windows パソコン、ファイルサーバ、メールサーバ等のコンピュータウィルス対策ソフトウェア。

(2)Pro/ENGINEER

3次元のCADのソフトウェア。

(3)JMP

統計解析ソフトウェア。

(4)SAS9

統計解析ソフトウェア。

(5)MATHEMATICA

数値計算や数式処理用のソフトウェア。

(6)ChemOffice

化学・バイオ分野で必要とされる様々なツールを1つにまとめた統合化学ソフトウェア。

(7)Buzzsaw

設計レビューと建設プロセスを自動化・標準化しながら、プロジェクトに関連するドキュメントや設計情報を一元管理することができるソフトウェア。

(8)Adobe CLP

Adobe社のAcrobat、Photoshop等の製品が安価に購入できるライセンス。

スーパーコンピューティング部門



ベクトル並列型スーパーコンピューター SR11000モデルJ1



導入作業中の新スーパーコンピューター HA8000

スーパーコンピューティング部門

概要

スーパーコンピューティング部門では、学術研究および教育に供することを目的として、全国の大学・研究機関等に在籍する大学教員、大学院学生、および卒業研究や授業を目的とした学生に対して、スーパーコンピューターシステムを用いた高度かつ大規模な計算サービスを提供している。ここでは2007年度中に行った本部門の業務について報告する。

2007年3月に同一契約において中間性能向上を行い、同年4月からこれまでの性能を約3倍に増強したベクトル並列型スーパーコンピューターシステム SR11000/J2(以下、SR11000 と記述)のサービスを開始した。また、同時に利用者の利便性を考慮し、システムリソースの拡大、利用負担金コースの新たな追加等を行った。

また、2007年3月をもってサービスを終了した超並列型スーパーコンピューターシステム SR8000/MPP(以下、SR8000/MPP と記述)の後継機の調達を行い、入札の結果、HA8000-tc/RS425(以下、HA8000 と記述)952 ノードを導入することになった。このシステムは東京大学、筑波大学、京都大学の3大学で定められた「T2K オープンスパコン仕様」に基づいたクラスタ型のスーパーコンピューターシステムで、現在、2008年6月稼働に向けて準備中である。なお、本システム導入に向け、施設および設備の整備を2007年9月から2008年2月にかけて行った。

また、文部科学省が2007年度から開始した委託事業「先端研究施設共用イノベーション創出事業(産業戦略利用)」から支援を受け、2007年7月から、東京大学情報基盤センターが代表として7大学の全国共同利用施設(北海道大学、東北大学、東京大学、名古屋大学、京都大学、大阪大学、九州大学)のスーパーコンピューターシステム及び応用ソフトウェア(シミュレーションソフトウェア等)を社会貢献の一環として民間企業へ提供する事業を新たに開始した。

スーパーコンピューティング業務

システム管理係

システム運用係

1 2007年度のシステム整備状況

本部門で提供しているスーパーコンピューターシステムのサービスについて、2007年度中に行ったシステム整備状況を以下に述べる。

利用者の利便性を考慮し、新たな利用環境等の整備、従来から継続してきた大規模(超並列)ジョブ実行サービスや教育利用への支援等を行った。また、長年サービスしてきたオープンフロアサービスは、ネットワーク利用の拡大に伴い近年その利用が激減したことで、施設の有効利用の観点からスーパーコンピューティング専門委員会の承認を得て2006年度末でサービスを終了したため、オープンフロアサービス終了の後処理や、プログラム相談室の常設の終了後の相談体制の整備等も行った。

1.1 SR11000/J2のサービス開始

SR11000/J1について、2007年3月25日に44ノードモデルから128ノードモデル(いずれも16CPU構成)へ増強を行い、4月2日(月)からサービスを開始した。(増強後(128ノード)のシステムをSR11000/J2と表記する。)性能を表1にまとめる。

表1. システム性能比較

	SR8000/MPP	SR11000/J1	SR11000/J2
ノード数	144 (8CPU)	44 (16CPU)	128 (16CPU)
1ノードあたりの理論演算性能	14.4 GFLOPS	121.6 GFLOPS (16CPU)	147.2 GFLOPS (16CPU)
総理論演算性能	2,073.6 GFLOPS	5,350.4 GFLOPS	18,841.6 GFLOPS
1ノードあたりの主記憶容量	16 GB	128 GB	128 GB
総主記憶容量	2,304 GB	5,632 GB	16,384 GB
ノード間ネットワーク性能	(片方向) 1.6 GB/秒 (双方向) 3.2 GB/秒	(片方向) 12 GB/秒 (双方向) 24 GB/秒	(片方向) 12 GB/秒 (双方向) 24 GB/秒
磁気ディスク装置	4.7 TB	32.3 TB	94.2 TB
サービス終了・開始時期	2007年3月7日 8:30 サービス終了	2007年3月25日 8:30 サービス終了	2007年4月2日 9:30 サービス開始

SR11000/J1では物理ノードを論理分割し、8CPUで構成するSMPを1ノードとして運用していた。これにより、1ノードあたりの理論演算性能、主記憶容量は73.6GFLOPS、64GBとなる。

1.2 2007年度のサービス変更

1.2.1 主な変更点

2007年4月以降のSR11000の利用に関する主な変更点は以下のとおり。詳細については次項以降を参照のこと。

- ・提供システムリソースの拡大
- ・追加オプション「計算リソース利用」の新設
- ・「研究室コース」の新設
- ・「利用負担金」額の改正

なお、2007 年度からのサービスについては、本センターのスーパーコンピューターシステムのサービスを審議するスーパーコンピューティング専門委員会での承認を得ている。

1.2.2 サービス変更点

SR11000 システムのサービスの変更点について以下に記載する。なお、システムのホスト名、IP アドレス、ファイルシステム名 (/batch、/para-io) は変更ない。

16CPU/node による運用

2006 年度までのシステムは 8CPU/node で運用してきたが、2007 年 4 月以降はシステムの大部分を 16CPU/node で運用した。これにより 16CPU/node で利用するノードにおいては、112GB/node のメモリーが利用できることとなった。複数ノード利用のキューについては、この 16CPU/node を充てた。なお、1 ノード利用のキューとしては P001(16CPU/node)、H1(8CPU/node) 両方用意した。

P016 キューの新設

2006 年度までは並列用ジョブキューとして P008(8CPU/node) が最大だったが、P016(16CPU/node) キューを新設した。これにより演算量・メモリー量とも 4 倍のリソースが利用できるようになった。なお、2006 年度まであった P002、open、personal キューは廃止した。

debug キューの新設

全ての利用者(研究室コース(後述)・教育利用(試行運用、無料)を除く)が利用可能 debug キューを新設した。4 ノード(16CPU/node)まで利用可能である。

single キューの再編

single キューは 2006 年度まで 6 段階(A~F)だったが、これを 4 段階(A~D)に変更した。また、ES キューは廃止した。

計算リソース追加オプション

計算リソース追加オプションとして、1・4・8 ノード専用キューを新設した。詳しくは「1.2.4 計算リソース追加オプション」を参照のこと。なお、SR8000/MPP のグループコースで利用可能だったグループ専用キューについては用意しないこととした。

規模ジョブサービス

大規模ジョブサービスとして 64 ノード利用可能な P064 キューサービスを月 1 回(原則として最終金曜日午後～翌月曜日朝)行った。このサービスは 5 月から開始した。

ファイルシステム

- ・ /batch の上限値は 10GB とした。グループコースは 1 グループ(利用者番号 10 個まで)で 200GB とした。追加オプションによりディスク容量を増やすことができる。
- ・ /short/batch は廃止した。
- ・ /para-io はインタラクティブから参照可能とした。(書き込みは NQS からのみとした。)
保存期間は 5 日間のままとした。ファイル上限値設定は行わない。

ジョブクラス制限値は表 2 のとおり。

1.2.3 利用負担金

利用負担金について、基本的には 2006 年度までの定額制を継続した。主な変更点は以下のとおり。

- ・ それまでのパーソナル・グループコースに加え、研究室コースを新設した。
- ・ パーソナルコースでは 4 コースを、グループコースでは 2 コースを用意した。
- ・ 追加オプションによる計算リソース (1・4・8 ノード専用キュー) を新設した。

利用負担金表は表 3 のとおり。

1.2.4 計算リソース追加オプション

追加オプションにより Q001(48 時間)、Q004(24 時間)、Q008(24 時間) キューが利用できるサービスを 2007 年 5 月から開始した。制限時間が一般のキュー P001(24 時間)、P004(12 時間)、P008(8 時間)より長く設定されている。また、人数制限されているためジョブが実行しやすくなっている。

利用見込み額

- ・ 利用申し込みのためには、利用見込み額の設定が必要。
- ・ 新規および継続利用申込書には、利用見込み額の記入欄がある。
- ・ 利用見込み額の追加は、Web ページまたは専用申込書(書面)により行える。
- ・ 利用見込み額として記入された金額が課金されるのではなく、に示す Web による利用申し込みをした時点で課金対象となる。
- ・ 利用見込み額を超えた利用申し込みはできないが、利用見込み額を追加することで利用申し込みが可能になる。

利用申し込み

- ・ 利用申し込みは、本センタースーパーコンピューティング部門の Web ページ (<http://www.cc.u-tokyo.ac.jp/>) から行う。
- ・ 利用申し込みは、月単位とする。3 ヶ月先まで予約可能とする。
- ・ それぞれのキューについて各月の上限を 10 名とする。(複数のキューを利用したい場合はそれぞれ申し込みが必要。Q008 を申し込めば Q001、Q004 も利用できるわけではない。)

ジョブの実行

- ・ NQS でのキューの指定は“#@\$-q Q001”のようにキュー名を指定する。
- ・ 1 利用者番号あたり投入可能なジョブは 1 とする。
- ・ ジョブの実行は FIFO(First In First Out)とする。つまり、待ちキューに並んでいる順番に実行される。
- ・ ジョブスケジューリングシステムにおける CPU 使用量の積算は行わない。

1.2.5 研究室コース

グループコースに登録するためには、10 人程度の利用者を集める必要があり、かつ 100 万円以上の利用負担金が発生していた。このため数人からなる小規模研究室では利用しにくい状況にあった。これらを解消するため、「研究室コース」を新設した。利用負担金表(表 3)の注意事項にある要件を満たすことにより 3 名以上で安価な利用負担金で利用できる。利用可能なキューについてはジョブクラス制限値(表 2)を参照。

表2. ジョブクラス制限値

キュー名	CPU数	最大ノード数	制限時間	メモリー (GB) ¹	パーソナル				グループ		研究室コース
					コア1	コア2	コア3	コア4	コア1	コア2	
(single)											
A	1	-	1 時間	2(56)							
B	"	-	10 時間	"							
C	"	-	50 時間	"							
D	"	-	225 時間	"							
S1	8	1	2 時間	56(56)							
S4	"	4	2 時間	"	×	×	×	×	×	×	
H1	"	1	24 時間	"	×						×
(parallel)											
P001	16	1	24 時間	112(112)	×						×
P004	"	4	12 時間	"	×	×					×
P008	"	8	8 時間	"	×	×	×		×		×
P016	"	16	6 時間	"	×	×	×		×		×
P064 ²	"	64	2 時間	"	×	×	×		×		×
debug	"	4	5 分	"							×
Q001	"	1	48 時間	"	×						×
Q004	"	4	24 時間	"	×	×					×
Q008	"	8	24 時間	"	×	×	×		×		×
インタラクティブ	1	-	18 時間	512MB (1GB)							

追加オプションにより申込可

1 1ノードあたりの標準値(括弧内は最大値)

2 P064 キューは月1回のみサービス(原則として最終金曜日午後～翌月曜日朝)

NQS でのキューの指定方法について

- ・single(A～D)、parallel(P001～P064) 利用については、従来どおり “#@\$-q single”、“#@\$-q parallel” と指定する。
- ・それ以外のキュー(S1、S4、H1、debug、Q001、Q004、Q008) については、キュー名をそのまま指定する。(例：“#@\$-q S1”、“#@\$-q debug”)
- ・キュー名は、大文字・小文字を区別するので、注意が必要。

表3. 利用負担金表

コース		負担金額(税込)	/batch	備考
パーソナルコース	コース1	20,000 円 / 年	10GB	計算リソース追加オプションは不可
	コース2	80,000 円 / 年		
	コース3	200,000 円 / 年		
	コース4	400,000 円 / 年		
グループコース	コース1	1,500,000 円 / 年	200GB	利用者番号 10 個まで
	コース2	3,000,000 円 / 年	+20GB	利用者番号 10 個まで
	追加オプション (1利用者番号につき)	150,000 円 / 年		コース1用利用者番号追加
		300,000 円 / 年	コース2用利用者番号追加	
研究室コース (1利用者番号につき)		50,000 円 / 年	10GB	計算リソース追加オプションは不可
追加オプション	計算リソース	1ノード専用キュー	20,000 円 / 月	パーソナル2・3・4, グループ1・2 利用者が申込可
		4ノード専用キュー	50,000 円 / 月	パーソナル3・4, グループ1・2 利用者が申込可
		8ノード専用キュー	80,000 円 / 月	パーソナル4, グループ2 利用者が申込可
	ディスク	600 円 / (GB*年)		GB 単位で申込可

- 利用期間は4月から当該年度末サービス終了日までの1年間を基本とし、年度途中で利用登録を取り消した場合であっても利用負担金額の変更は行わない。
- 利用負担金は、原則として以下のとおりそれぞれ初旬に一括して請求する。
 - パーソナルコースは、利用開始月が4月から5月までは7月、6月から8月までは10月、9月から11月までは1月、12月から3月までは3月。
 - グループコース・研究室コースは、利用開始月が4月から6月までは7月、7月から9月までは10月、10月から12月までは1月、1月から3月までは3月。
 - 計算リソース追加オプションについては、申し込み月直後の7月、10月、1月、3月。
 - 上記以外の月に請求を希望する場合は、別途相談に応じる。
- パーソナルコース(ただし、本センターのスーパーコンピューターシステムに初めて登録された利用者)においてのみ、利用開始月の翌月末日までに利用を中止することができる。利用負担金はパーソナルコースの利用開始月3月の金額を適用し、請求する。
- 利用負担金が減額となる変更はできない。

- コース間の変更については、利用負担金が増額になる場合のみ別途相談に応じる。(ただし、利用者番号変更の場合がある。)
- グループコースのディスク量は、グループ全体の上限値である。
- 研究室コースの申し込みは、以下の要件を満たす必要がある。
 - 同一研究室内にてグループを構成する。
 - グループは3名以上(教員・学生各1名以上)で構成される。(教員がパーソナルコース2以上に登録している場合は、学生2名以上のみの登録を可能とする。)
 - 登録人数の上限はない。
 - 教員がパーソナルコース2以上に登録している場合は、当該教員に対してS4キューの実行を可能とする。
- 計算リソース追加オプションの申し込みは、月単位とする。それぞれのキューについて、各月の上限を10名とする。申し込み者の人数により、お断りする場合がある。利用申し込みは、スーパーコンピューティング部門のWebページから行う。

1.3 サービス状況

SR11000 の 16CPU/node のノード稼働率は、年平均 89.0%、最大月(11月)で 96.2%にも達し、ほぼ飽和状態であった。

1.4 教育利用

前年度に引き続き、教育利用のためのシステム提供(試行、無料)を行い、SR11000 を利用した並列プログラミング教育として 4 件の利用があった。

- ・東京大学理学系研究科地球惑星科学専攻
- ・東京大学工学部・工学系研究科(前期)
- ・東京大学工学部・工学系研究科(後期)
- ・東京大学教養学部 全学ゼミ(1、2年生)

1.5 施設等の改修工事について

超並列型スーパーコンピューターシステムの 2008 年 6 月稼働開始に伴い、2008 年 3 月末までに、施設及び設備について、本センター本館及び別館の床等の施設改修、空調及び電源等の設備増強を行った。

1.6 参考

システム構成図を図 1 に、ソフトウェア一覧を表 4 に示す。

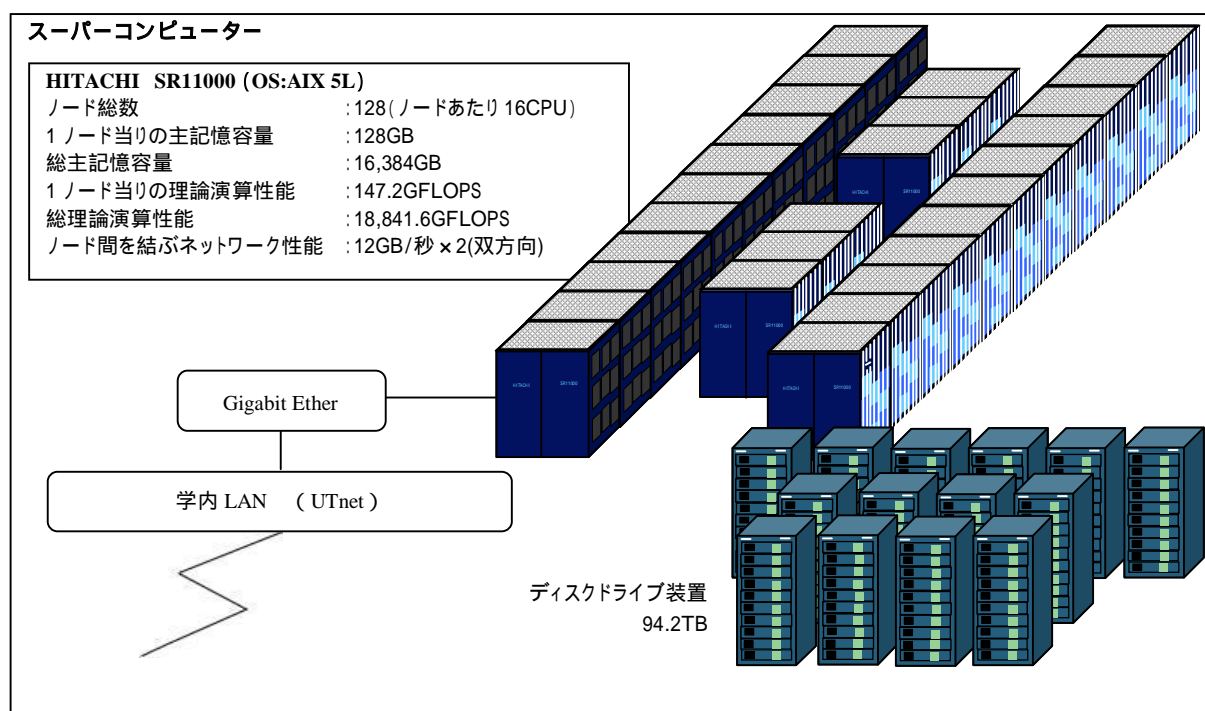


図 1. システム構成図(2007 年 4 月)

表 4. ソフトウェア一覧

ベクトル並列型スーパーコンピューター	: HITACHI SR11000
オペレーティングシステム	: AIX 5L システム
シェル (コマンドインタプリター)	: sh, csh
言語プロセッサ	: 最適化 FORTRAN77, 最適化 FORTRAN90, 最適化 C, 最適化標準 C++ , 他 国際標準規格 ISO/IEC 14882:1998 準拠 及び 旧言語仕様 (ARM 仕様) 対応
数値計算	: MATRIX/MPP, MATRIX/MPP/SSS, MSL2, BLAS, LAPACK, ScaLAPACK
分子計算	: GAUSSIAN03
並列化支援	: MPI (MPI-2)

2 2008 年度のシステム整備計画

HA8000 については、6 月 1 日サービス開始予定として作業を進めている。また、SR11000 については、2007 年度のサービス形態を継続する。

HA8000 のサービス形態については、パーソナルコース、専用キュー・ノード固定の各コースを用意する。プロジェクト等からの申し込みについては、現有の SR11000 であるグループコースは用意せず、専用キュー・ノード固定での申し込みを受け付ける。SR11000 で行っている教育利用・若手利用については、同様の試行サービス(当面無料)を行う予定である。また、月に1回数日程度、大規模ジョブ実行環境を提供する予定である。

HA8000 は、大規模ジョブを優先的に実行するというコンセプトから、ジョブクラス制限値を、パーソナルコースの利用では最低でも 4 ノードから、専用キュー・ノード固定の申し込みは 8 ノードからとする。(参考:1 ノード = 4 コア*4 = 16 コア、147.2GFLOPS/ノード、147.2*4=588.8GFLOPS(4 ノード)、147.2*8=1177.6GFLOPS(8 ノード))

また、2008 年 6 月から 9 月末までは、試行利用期間として無料で提供し、正式サービス(有料)は 10 月からを予定している。なお、利用負担金については SR11000 とは別の体系とするが、SR11000 からの移行のための措置をとる予定である。

3 超並列型スーパーコンピューターシステムの調達

2007 年 3 月にサービスを終了した SR8000/MPP の後継機の選定を行った。

3.1 概要・日程

2007 年度中の日程は以下のとおりである。

- (1) 仕様書案に対する、照会及び提案の申出・修正期限(2007.5.14)
- (2) 入札説明書(仕様書、総合評価基準等を含む)の策定(2007.5.14)
- (3) 第 8 回スーパーコンピューティング専門委員会(2007.9.11)
- (4) 第 3 回超並列型スーパーコンピュータシステム仕様策定委員会の開催(2007.9.27)
- (5) 官報公示掲載日、技術審査委員会の設置・運営(2007.10.1)
- (6) 入札説明会の開催(2007.10.9)
- (7) 応札締切日(2007.11.29)
- (8) 第 1 回超並列型スーパーコンピュータシステム技術審査委員会の開催(2007.12.4)
- (9) 第 2 回超並列型スーパーコンピュータシステム技術審査委員会の開催(2007.12.18)
- (10) 開札日(2007.12.25)
- (11) 第 9 回スーパーコンピューティング専門委員会(2008.3.14)

3.2 仕様策定

9月27日に、平尾公彦(大学院工学系研究科教授)委員長のもと、第3回仕様策定委員会を開催した。

3.3 技術審査

12月4日および12月18日に、森下真一(大学院新領域創成科学研究科教授)委員長のもと、技術審査委員会を開催した。

3.4 開札結果

12月25日に開札が行われ、日立製作所の「HA8000-tc/RS425」システムに決定した。

3.5 システム概要

3.5.1 ハードウェア

HA8000(管理サーバを除く)のハードウェア諸元は、表5のとおりである。

表5. HA8000 ハードウェア諸元

項目		仕様
システム全体	総理論演算性能	140.1344 TFLOPS
	総主記憶容量	31.25 TB
	総ノード数	952
	ノード間ネットワーク性能	5 GB/s×双方向(計算ノード群A) 2.5 GB/s×双方向(計算ノード群B)
	ストレージ装置容量	1 PB
ノード	理論演算性能	147.2 GFLOPS
	プロセッサ数(コア数)	4(16)
	主記憶容量	32 GB(936ノード) 128 GB(16ノード)
	ローカルディスク容量	250 GB
プロセッサ	プロセッサ(周波数)	AMD社製 Quad Core Opteron プロセッサ (2.3GHz)
	キャッシュメモリ	L2: 512 kB/コア L3: 2 MB/プロセッサ
	プロセッサコア 理論演算性能	9.2 GFLOPS

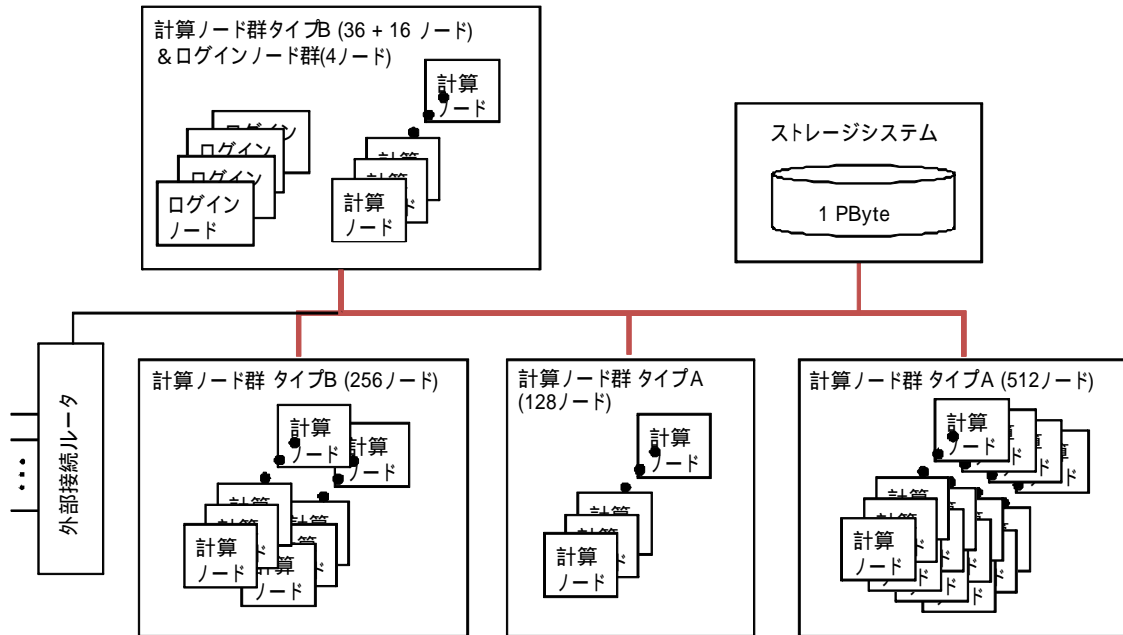


図 2 . HA8000 新システムの概略

3.5.2 ソフトウェア

HA8000（管理サーバを除く）のソフトウェア諸元は、表 6 のとおりである。

表 6 . HA8000 ソフトウェア諸元

項目	仕様
OS	RedHat Enterprise Linux 5
バッチシステム	NQS（現有の SR11000 と同等の機能）
コンパイラ	日立製作所製 最適化 Fortran(77/90/95)、 最適化 C、最適化標準 C++ （全て OpenMP 2.0 を含む）
並列化支援	MPI1.2 通信ライブラリ (MPICH・MX)
数値計算ライブラリ	MSL2、MATRIX/MPP、MATRIX/MPP/SSS BLAS、LAPACK、ScaLAPACK
分子計算アプリケーション	Gaussian03

3.5.3 SR11000 との比較

表 7 に、現有の SR11000 とシステム全体の比較を示す。総理論演算性能値では、現有スパコンの約 7 倍、総主記憶容量は約 2 倍となっている。ノード単体での性能は、理論演算性能値は同じで、主記憶容量は 1/4 となっている。

表7. 現有スパコンとの比較

	HA8000	SR11000
総理論演算性能	140.1344 TFLOPS	18.841 TFLOPS
総主記憶容量	32000 GB	16384 GB
総ノード数	952	128
ノード間ネットワーク性能	タイプ A : 5 GB/s × 双方向 タイプ B : 2.5 GB/s × 双方向	24 GB/s
単体ノード理論演算性能	147.2 GFLOPS	147.2 GFLOPS
単体ノード主記憶容量	32 GB (128 GB)	128 GB

4 専門委員会報告

本センター運営委員会のもとにスーパーコンピューティング専門委員会を設置し、全国共同利用スーパーコンピューターシステムおよびその運用に関する事項について、企画、立案及び審議を行っている。2007年度に開催した同専門委員会の主な議事内容は以下のとおりである。

第8回 スーパーコンピューティング専門委員会

日 時：平成19年9月11日(火)13:00～

議 題：イノベーション創出事業について

- ・東京大学情報基盤センタースーパーコンピューターシステム利用規程の改定等について
- ・超並列型スーパーコンピューターシステムの導入について
- ・国立環境研との共同研究について
- ・スーパーコンピューター若手利用者推薦(試行)について
- ・平成18年度スーパーコンピューティング部門決算報告について
- ・プログラム相談状況について
- ・システム稼動状況報告について
- ・その他

第9回 スーパーコンピューティング専門委員会

日 時：平成20年3月14日(金)13:00～

議 題：超並列型スーパーコンピューターシステムの調達状況および今後のスケジュールについて

- ・超並列型スーパーコンピューターシステムの概要について
- ・超並列型スーパーコンピューターシステムの運用(案)および利用負担金(案)について
- ・東京大学情報基盤センター規則の改定について
- ・東京大学情報基盤センタースーパーコンピューターシステム利用期間の変更について
- ・東京大学情報基盤センタースーパーコンピューター利用資格者審査委員会内規の改定について
- ・T2Kの間における連携・協力の推進に関する協定書について
- ・イノベーション創出事業について
- ・プログラム相談状況について
- ・システム稼動状況報告について
- ・その他

5 システムの利用状況

2007 年度におけるジョブ統計は表8のとおりで、スーパーコンピューターの利用状況については実利用者が 550 名を超え、計算件数及び演算時間共に昨年度を大幅に上回っている。

5.1 SR11000/J2 (AIX 5L) のジョブ処理状況

表 8 . ジョブ処理状況

年月	登録者数	実利用者数	TSS 件数	NQS 件数	接続時間 (時間)	CPU 時間		ノード稼働率 (%)	/batch ファイル 使用量(MB)
						sr11000-s (時間)	NQS (時間)		
0704	574	247	6,503	5,406	8,945.68	85.73	922,938.78	70.1	6,916,701
0705	632	274	6,915	10,134	6,960.01	165.57	814,087.27	69.2	7,476,493
0706	651	281	9,214	11,245	11,433.71	195.63	1,126,228.71	79.2	6,653,725
0707	657	276	8,887	10,122	11,320.12	172.67	1,087,884.20	75.3	7,074,926
0708	633	250	5,991	8,997	8,328.28	157.71	791,696.08	79.0	7,490,724
0709	691	232	7,841	7,295	9,704.18	163.23	1,030,975.49	80.4	9,382,304
0710	755	307	8,917	8,990	10,184.65	177.83	990,756.08	80.3	11,474,240
0711	774	300	10,832	9,776	11,141.78	274.29	1,122,100.69	84.7	9,262,011
0712	791	299	10,401	8,926	11,137.62	233.98	1,105,665.86	82.1	9,662,213
0801	796	314	12,105	10,450	12,788.45	239.17	1,136,868.03	85.2	11,364,957
0802	812	293	10,332	10,687	12,603.05	271.91	1,109,448.96	86.6	11,917,289
0803	813	277	7,748	9,075	9,658.02	230.52	1,186,970.61	83.3	12,158,862
合計	816	556	105,686	111,103	124,205.57	2,368.57	12,425,620.76		

(注) ・ノード稼働率とは1ノードが100%稼働したと仮定した場合のサービスノードに対する稼働率である。

6 プログラム相談体制と刊行物

6.1 プログラム相談体制

サービスの一環として、利用者からのプログラム相談を受け付けている。しかし、オープンフロアサービスを 2006 年度をもって終了したことにより、これまでのようにプログラム相談員を常設せず、事前予約により本センター教職員が対応するものに改められた。また、メールによる相談もメール指導員は置かず、本センターの教職員が回答することになった。変更点は表9のとおりである。

表 9 . プログラム相談体制の変更点

	2006 年度 (旧)	2007 年度 (現行)	備考
指導員室 (センター 1 階)	指導員室での面談	事前予約による面談、センター教職員が対応	2007 年度からオープンサービスを廃止することに伴う
メール指導員	メールによる質問の回答	soudan@cc.u-tokyo.ac.jp は継続し、センター教職員が回答	
連絡所付指導員	各連絡所での面談	(継続)	

6.2 刊行物

2007 年度に刊行したものは以下のとおりである。

- スーパーコンピューティングニュース Vol.9 No.3 ~ Vol.10 No.2 (2007 年度)
- スーパーコンピューティングニュース Vol.9 No. Special Issue 1 2008.2
- 東大センターにおけるスーパーコンピューター利用入門 第 3 版 (2007.4)

また従来から発行しているものには、以下のものがある。

- ベクトル並列型スーパーコンピューターSR11000 チューニングガイド(2005.5)

7 職員による研究開発活動

業務系の職員は、日々利用者サービスの向上やシステム運用の効率化・高度化を図りながら業務に携わっており、その成果を発表することで更なる動機付けに期待される。このような成果の発表の場として、全国共同利用情報基盤センター長会議のもとに研究開発連合発表会講演会を毎年開催している。その他、各大学等の機関で行っている総合技術研究会にも参加している。

7.1 第 29 回全国共同利用情報基盤センター 研究開発連合発表講演会

北海道大学情報基盤センターの当番で、第 29 回全国共同利用情報基盤センター研究開発連合発表講演会が 2007 年 10 月 18 日に開催され、本センターからも職員が参加し発表を行った。

本講演会は日々急激に進展するスーパーコンピューターを中心とした計算機、ソフトウェア及びネットワークに関する最新の技術について知識・サービスの発表・討議を行い、情報基盤センター群相互の情報交換を図るものである。

< 本センターの研究開発論文 >
(東京大学情報基盤センター)

- ・「計算リソース追加オプション予約システム」
- 井爪健雄(発表者)、松葉浩也、有賀浩、石崎勉、田川善教、滝本友彦 -

本システムは、スーパーコンピューターSR1100 サービスにおいて、通常のジョブ実行サービスとは別にジョブ実行を優先する専用キューを新たに設け、その申込みを Web インターフェース化することにより高速化、公平化、簡略化を図ったシステムである。

詳しくは、「研究開発論文集 No.29」平成 19 年 10 月全国共同利用情報基盤センター (P28 ~ P33) をご覧願いたい。

先端的大規模計算シミュレーションプログラム利用サービス

- 文部科学省先端研究施設共用イノベーション創出事業【産業戦略利用】 -

共同利用係

1 事業内容

1.1 事業の趣旨と概要

「先端研究施設共用イノベーション創出事業」は、大学、独立行政法人等の研究機関が有する先端的な研究施設・機器について、広範な分野における幅広い利用(共用)を促進し、イノベーションにつながる成果を創出するために、平成 19 年度から文部科学省が新たに開始した委託事業である。

本事業を通じて、産学官の研究者による戦略的かつ効率的な研究開発や、研究機関・研究分野を越えた横断的な研究開発活動を推進することにより、継続的に産学官の知の融合によるイノベーションを加速していくことを目指している。

全国共同利用施設であるスーパーコンピューターを有する7大学(北海道大学、東北大学、東京大学、名古屋大学、京都大学、大阪大学、九州大学)は、社会貢献の一環として、大学で開発された応用ソフトウェアとスーパーコンピューターの、民間企業への提供を実施している。7大学が有するスーパーコンピューターの総計算能力(理論性能値)は、平成 20 年度には約 330TFlops となった。

本事業の特徴は、各大学が開発した応用ソフトウェア(シミュレーションソフトウェア等)と各大学が持つ国内有数のスーパーコンピューターを、それらの利用支援サービスと一体化して民間企業に提供することにある。このようにサービス、ソフトウェア、ハードウェアの3つを同時に支援する試みは、国内外を通じて初めてのものである。この事業の目的は、大学の持つ知と施設によって我が国の経済発展に貢献することである。

- 本事業は、東京大学が代表して、文部科学省「先端研究施設共用イノベーション創出事業」プログラムから支援を受けている。
- 国立大学法人 東京工業大学において同プログラムからの支援を受けた「『みんなのスパコン』TSUBAME によるペタスケールへの飛翔」プログラムと連携し、民間企業へのスーパーコンピューター提供事業を行っている。

1.2 支援内容

7 大学は、ASP(Application Service Provider)事業を行う NPO 等の団体あるいは企業と協力して戦略分野利用推進と新規利用拡大の2つのカテゴリに分けて支援する。

1.2.1 戦略分野利用推進

ライフサイエンス、もの作り技術、情報通信、環境エネルギー、社会基盤分野を対象に、大学で開発された先端的シミュレーションソフトウェアとスーパーコンピューターの利用課題を民間企業から募り、企業によるイノベーション創出を支援する。民間企業が有するアプリケーションを利用した課題あるいは新規アプリケーションを開発する課題も募っていく。

- タンパク質機能シミュレータによる創薬開発
- ナノマテリアルシミュレータによる新材料開発
- 音・熱・力の解析シミュレータによるものづくり
- 機器・システムレベルのシミュレータ開発・利用実験による次世代情報通信技術開発
- 地盤と拡散物質・構造物解析シミュレーション技術の開発

1.2.2 新規利用拡大

新しいニーズを掘り起こしイノベーション創出につながる利用課題も民間企業から募っている。分野は特定していない。産業界の潜在的高性能計算需要を掘り起こし、大学とASPとの連携によるサービス、ソフトウェア、ハードウェアを有償で利用する下地を作ることが目的となる。

1.2.3 利用可能なソフトウェア

- 革新的シミュレーションソフトウェアの研究開発(RSS21)で開発されたソフトウェア群
 - タンパク質構造解析ソフトウェア myPresto / AMOSS
AMOSSは日本電気株式会社の製品
 - 分子動力学法に基づくソフトウェア OCTA
 - 全電子第一原理計算ソフトウェア TOMBO
 - 数値タービンシミュレーション
 - 超臨界流体シミュレータ
 - TAS-code(Tohoku University Aerodynamic Simulation Code)
 - 時間領域差分法に基づく大規模電磁界解析ソフトウェア JetFDTD
 - ワイヤレスコミュニケーション可視化システム(RWV) Ver.2
 - 可視化システム KVS
- ◇ 上記以外で企業がすでに利用しているアプリケーションをスパコンで実行することも可能。また、企業によるソフトウェアの新規開発も可能。

1.3 公募日程

1.3.1 平成19年度第1期公募

- 8月22日(水) 第1回公募 受付開始
- 9月14日(金) 第1回公募 受付締め切り
- 9月28日(金) 第1回公募 採択通知
- 10月1日(月) 第1回公募 利用手続き開始

東京大学情報基盤センターでは8月22日に公募説明会を実施した。

1.3.2 平成19年度第2期公募

- 10月15日(月) 第2回公募 受付開始
- 11月30日(金) 第2回公募 受付締め切り
- 12月14日(金) 第2回公募 採択通知
- 12月17日(月) 第2回公募 利用手続き開始

東京大学情報基盤センターでは10月15日に公募説明会を実施した。

1.4 採択結果

1.4.1 平成19年度第1期公募

- 戦略分野利用推進
 - 申請 10件
 - 採択 10件(うち東京大学1件)
 - ◇ 三菱自動車工業株式会社 技術開発本部 機能実験部 流体技術グループ「自動車部品から発生する空力騒音予測手法の開発」

- ◇ 特定非営利活動法人バイオグリッドセンター関西「タンパク質機能シミュレータによる創薬開発」
- ◇ (株)豊田中央研究所 先端研究センター 倉本研究グループ「第一原理計算による貴金属/担体の相互作用と触媒活性についての解析」
- ◇ 三菱重工業株式会社 原動機事業本部蒸気タービン統括技術部「三次元蒸気タービン静動翼列多段流動の大規模数値シミュレーション」
- ◇ 松下電器産業株式会社 システムエンジニアリングセンター「大規模 EMC シミュレーション」
- ◇ 日本電気株式会社 ナノエレクトロニクス研究所「固体高分子型燃料電池材料の in silico 設計」
- ◇ 積水化学工業株式会社 R&D センター NBO 開発推進センター「熱硬化性樹脂原料の構造・物性シミュレーション」
- ◇ SRI 研究開発株式会社 材料プロセス研究部「高性能タイヤの開発を目的とした大規模 MD のための初期構造作成方法の検討」
- ◇ SRI 研究開発株式会社 情報研究部「ゴルフボール空力大規模計算による乱流遷移の研究」
- ◇ ソニー株式会社 モノ造り技術センター 技術開発室「大規模電磁界解析の高速化/高精度化の検討」
- 新規利用拡大
 - 申請 10 件
 - 採択 9 件(うち東京大学 4 件)
 - ◇ 国際電気通信基礎技術研究所 音声言語コミュニケーション研究所「音声翻訳のための音声言語データの収集とモデル化の並列処理」
 - ◇ 鹿島建設株式会社 ITソリューション部「建設分野構造解析における大規模外部リソース利用と RSS21 ソフトウェア活用に関する研究」
 - ◇ SRI 研究開発株式会社 商品研究部「大規模/フルアトムックモデルを用いた分子シミュレーションによる分子構造からの高分子化合物の物性予測」
 - ◇ 株式会社神戸製鋼所 電子技術研究所「有機分子/金属界面構造のシミュレーション」
 - ◇ 株式会社情報数理研究所「環境対応型全体最適指向生産スケジューラ ASP のための分散・並列計算基盤の構築」
 - ◇ 川崎重工業株式会社 航空宇宙カンパニー研究部空力技術課「航空宇宙分野における大規模空力解析の応用」
 - ◇ みずほ情報総研株式会社 サイエンスソリューション部「大規模流体解析を対象とした ASP 事業のフェージビリティスタディ」
 - ◇ 日本アイ・ビー・エム株式会社 東京基礎研究所「大規模エージェントベースシミュレーションの性能評価」
 - ◇ ビヨンド・コンピューティング株式会社「グリッド ASP 利用モデルの研究」

1.4.2 平成 19 年度第 2 期公募

- 戦略分野利用推進
 - 申請 2 件
 - 採択 1 件(うち東京大学 0 件)
 - ◇ 株式会社ケイ・ジー・ティー ビジュアライゼーション事業部「大規模計算結果の遠隔可視化システムに関する技術開発と実証試験」
- 新規利用拡大
 - 申請 1 件

- 採択1件(うち東京大学0件)
 - ◇ ダイハツ工業株式会社 先端技術開発部「脱貴金属を目指すナノ粒子自己形成触媒の新規発掘」

1.5 諸会議

1.5.1 イノベーション事業委員会

北海道大学、東北大学、東京大学、名古屋大学、京都大学、大阪大学、九州大学の各委員が参加し、民間からの応募課題について審査・採択等をおこなった。委員会は合計4回開催した。

第1回イノベーション事業委員会 平成19年9月26日(水)

議題:

1. 臨時全国共同利用情報基盤センター長会議報告について
2. イノベーション事業委員会委員長の選出について
3. 応用課題の審査・採択について
4. 第2回公募スケジュールについて
5. 平成20年度公募スケジュールについて
6. その他

第2回イノベーション事業委員会 平成19年12月12日(水)

議題:

1. 平成19年度第2期応募課題の審査・採択について
2. 平成20年度第1期公募について
 - ・公募公告について
 - ・パンフレットの作成・配布について
 - ・予算額と課題採択の関係について
 - ・昨年前半利用の場合の利用負担金について
 - ・グリッドについて(SINETの企業利用についての見込等)
3. 次回イノベーション事業委員会開催日について
4. その他

第3回イノベーション事業委員会 平成20年2月27日(水)

議題:

1. 第2回イノベーション事業委員会議事要旨確認
2. 平成19年度報告書改定について
3. グリッド運用について
4. 文部科学省ヒヤリング報告
5. シンポジウム準備状況について
6. 平成20年度計画について
7. その他

第4回イノベーション事業委員会 平成20年3月24日(月)

議題:

1. 第3回事業委員会(2/27)議事要旨(案)について
2. 平成20年度第1期応募課題の審査・採択について
3. 平成19年度採択課題の継続利用について
4. 平成20年度業務計画(案)

5. 平成20年度の計算機使用に係る予算方針・執行について
6. グリッドについて
7. その他

1.5.2 イノベーション業務連絡会

北海道大学、東北大学、東京大学、名古屋大学、京都大学、大阪大学、九州大学の各業務レベルの担当者が業務に関して検討を行った。会は合計3回開催した。

第1回イノベーション業務連絡会 平成19年8月2日(木)

議題:

1. 事業の概要説明
2. 利用者番号について
3. 民間企業の利用について
4. 運転経費(計算機利用負担金相当額)の東京大学からの支出方法について
5. 利用状況把握及び実績報告について
6. セキュリティ及びソフトウェアライセンスについて
7. その他

第2回イノベーション業務連絡会 平成19年12月20日(木)

議題:

1. 平成19年度第2期採択結果について
2. 平成19年度利用負担金の請求について
3. 平成20年度への継続利用申請について
4. 平成20年度第1期公募について ~ 予算額と課題採択の関係 ~
5. 利用統計について
6. 各大学のRSS21インストール及びライセンスについて
7. Gaussianのライセンスについて
8. シンポジウムの開催について
9. その他

第3回イノベーション業務連絡会 平成20年3月24日(月)

(第4回イノベーション事業委員会と同時開催)

議題:

1. 第2回業務連絡会(12/20)議事要録(案)について
2. 採択期間と計算機利用負担金について
3. RSS21インストール状況及びライセンスについて
4. Gaussianのライセンスについて
5. 利用統計について
6. その他
 - ・シンポジウムについて
 - ・平成20年度の連絡体制について(名簿とメーリングリスト)
 - ・利用規程/負担金表の送付について
 - ・広報について

1.6 広報活動

事業発足について、平成 19 年 7 月 24 日(火)に記者会見を行った。その後、新聞・雑誌に公募広告を掲載した。また 4 ページのリーフレットを 1700 部作成し、各種セミナー、展示会などで配布を行った。その他、企業などでの講演会で講演を行い、本事業の認知度を高める努力を行った。関係機関のメーリングリストへ公募説明会情報などの送信を依頼し配布した。共用ナビへの公募情報、シンポジウム情報を掲載した。

東京工業大学と連携して広報活動を行い、雑誌広告などは重複しないように担当を分けて掲載を行った。

1.6.1 新聞・雑誌記事

- 朝日新聞 7 月 25 日朝刊
- 毎日新聞 7 月 25 日朝刊
- 日刊工業新聞 7 月 25 日朝刊
- フジサンケイビジネスアイ 7 月 25 日
- 工業調査会「M&E 2 月号」(1 月 30 日)
- 日経 BP 社「日経コンピュータ 2 月 1 日号」(1 月 31 日)

1.6.2 新聞・雑誌広告

- 日本経済新聞 全国版 2 月 18 日朝刊
- 日経産業新聞 2 月 22 日
- 工業調査会「M&E 2 月号」(1 月 30 日)
- 日本計算工学会学会誌「2008 年第 1 号」(1 月 31 日)
- 日本化学会「化学と工業 2 月号」(2 月 1 日)
- 日刊工業新聞「機械設計 3 月号」(2 月 10 日)
- 日経 BP 社「日経コンピュータ 3 月 1 日号」(2 月 29 日)

1.6.3 企業などでの講演会

- 日本自動車工業会:作業部会での講演(7 月 19 日)
- みずほ情報総研:大規模流体解析 ASP サービス紹介セミナー(1 月 18 日)
- 豊田中央研究所:講演会(1 月 25 日)
- PC クラスタコンソーシアム:PC クラスタワークショップ in 名古屋(3 月 7 日)

1.7 シンポジウム開催

3 月 28 日に本プログラムの現状報告、採択企業からの本年度の成果報告、計算機シミュレーションによるイノベーション創出に向けての大学の役割についてのパネル討論などを実施するシンポジウムを行った。当日は 151 名の来聴者があった。

先端的大規模計算シミュレーションプログラム利用シンポジウム

開催日時: 2008 年 3 月 28 日(金)9:50 ~ 18:30

開催場所: 東京ステーションコンファレンス 5 階

定員: 200 名

主催: 東京大学情報基盤センター、
北海道大学情報基盤センター、
東北大学情報シナジーセンター、
名古屋大学情報連携基盤センター、

京都大学学術情報メディアセンター、
大阪大学サイバーメディアセンター、
九州大学情報基盤研究開発センター
後援： 文部科学省

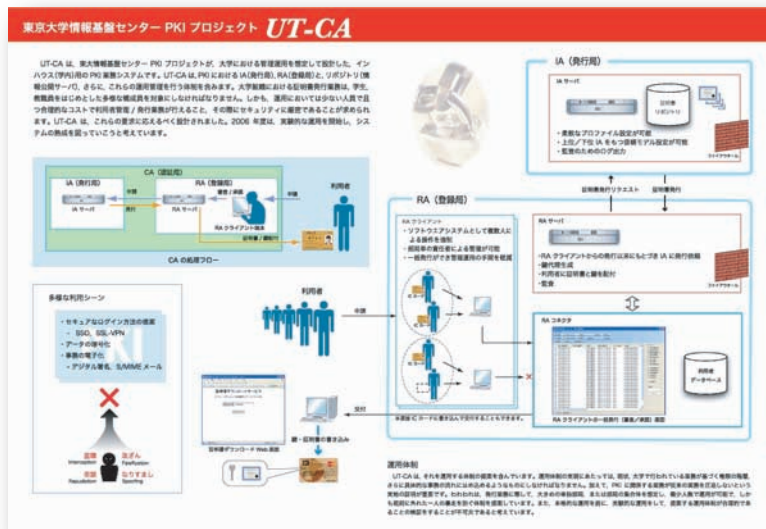
プログラム：

- 9:20 ~ 受付開始
- 9:50 ~ 10:00 開会挨拶
米澤 明憲(東京大学 情報基盤センター センター長)
- 10:00 ~ 10:30 基調講演「先端研究施設の共用 ~イノベーション創出のために~」(仮題)
田口 康(文部科学省研究振興局 研究環境・産業連携課 課長)
- 10:30 ~ 11:00 基調講演「次世代スーパーコンピュータプロジェクトの概要と状況」
渡辺 貞(理化学研究所 次世代スーパーコンピュータ開発実施本部
プロジェクトリーダー)
- 11:00 ~ 11:45 「7 大学スパコンセンターによる民間利用支援の現状と今後」
石川 裕(東京大学 情報基盤センター)
- 12:00 ~ 13:30 休憩
- 13:30 ~ 15:00 平成 19 年度第 1 期採択企業からの報告
- 「第一原理計算による貴金属/担体の相互作用と触媒活性についての解析」
倉本 圭(豊田中央研究所)
 - 「自動車部品から発生する空力騒音予測手法の開発」濱本 直樹(三菱自動車工業株式会社)
 - 「音声翻訳のための音声言語データの収集とモデル化の並列処理」安田 圭志(国際電気通信基礎技術研究所)
 - 「大規模流体解析を対象とした ASP 事業のフィジビリティスタディ」大谷 泰昭(みずほ情報総研株式会社)
 - 「環境対応型全体最適指向生産スケジューラ ASP のための分散・並列計算基盤の構築」磯野 彬(株式会社 情報数理研究所 社長)
 - 「グリッド ASP 利用モデルの研究」西 克也(ピヨンド・コンピューティング株式会社)
- 15:00 ~ 15:15 休憩
- 15:15 ~ 17:00 パネル討論「計算シミュレーションによるイノベーション創出に向けて大学の役割」
- 司会： 平尾 公彦(東京大学 副学長)
- パネリスト： 川添 良幸(東北大学)、米澤 明憲(東京大学)、
村上 和彰(九州大学)、高田 俊和(理化学研究所)、
善甫 康成(住友化学筑波研究所)
- 17:00 ~ 18:30 ポスターセッション & 懇親会
- 「タンパク質機能シミュレータによる創薬開発」((財)千里国際情報事業財団(NPO 法人バイオグリッドセンター関西))
 - 「建設分野構造解析における大規模外部リソース利用と RSS21 ソフトウェア活用に関する研究」(鹿島建設株式会社 IT ソリューション部)
 - 「第一原理計算による貴金属/担体の相互作用と触媒活性についての解析」((株)豊田中央研究所 先端研究センター 倉本研究グループ)
 - 「大規模 EMC シミュレーション」(松下電器産業株式会社 システムエンジニアリングセンター)

- 「有機分子/金属界面構造のシミュレーション」(株式会社神戸製鋼所 電子技術研究所 薄膜材料研究室)
- 「高性能タイヤの開発を目的とした大規模 MD のための初期構造作成方法の検討」(SRI 研究開発株式会社 材料プロセス研究部)
- 「ゴルフボール空力大規模計算による乱流遷移の研究」(SRI 研究開発株式会社 情報研究部)
- 「大規模エージェントベースシミュレーションの性能評価」(日本アイ・ピー・エム株式会社 東京基礎研究所)

順不同

PKI プロジェクト



UT-CA(東大認証局のパイロット版)リーフレット

「オープンドメイン証明書発行実験」に伴う東大シールとサーバ検証結果の表示



サイト検証結果 - www.pki.itc.u-tokyo.ac.jp

https://seal.pki.itc.u-tokyo.ac.jp/cgi-bin/verifypage.cgi?lang=ja

サイト検証結果: 有効

検証日時: 2008年6月2日 18:03 (日本時間)

下記のサイトに対する証明書は、東大登録局(TRA)による審査の上で発行されています。

サイト名: www.pki.itc.u-tokyo.ac.jp
 部署: 情報基盤センター
 有効期間: 2007年11月13日 から 2009年3月31日 まで

アクセス先のサイトが上記と相違ないことをご確認ください。
 このページのアドレスがhttps://seal.pki.itc.u-tokyo.ac.jp/で始まることをご確認ください。

東京大学 THE UNIVERSITY OF TOKYO

完了 seal.pki.itc.u-tokyo.ac.jp

PKIプロジェクト

佐藤周行

プロジェクト概要

PKI プロジェクトは、CA（認証局）を安定的に運用するための運用技術を、発足以来重要項目として検討してきたが、いくつかの試行錯誤を経て、UT-CA としてキャンパス PKI のための CA を実際に構築し、運用技術を蓄積するために、インソースでの運用でのプロトタイプを構築した。現在では、UT-CA の実験協力部局として情報学環、情報基盤センター・本部合同を得て実験開始以来 1 年 5 ヶ月の間安定して運用できている。このことで当プロジェクトが提唱している分散 RA の運用モデルの有効性が明らかになった。

NII が主導している CSI UPKI プロジェクトが開始した「オープンドメイン証明書発行実験」に参加し、パブリックサーバ証明書発行のための学内の体制を構築した。ここにおいても分散 RA 方式を取っている。2008 年 3 月末で参加部局は 6 部局を数え、サーバ証明書を発行した部局は 11 部署、67 件に及ぶ。

PKI プロジェクトのもうひとつの目標は PKI のアプリケーション調査である。特に有力なアプリケーションとして、SSL-VPN を取り上げ、そのテスト運用を行った。特に図書館電子化部門・分散教育メディア研究部門と協力して図書館アクセス認証のゲートウェイとして SSL-VPN を採用し、テスト的に運用を行っている。

1 運用報告

1.1 UT-CA

http://www.pki.itc.u-tokyo.ac.jp/UT_CA-top.html

現在では、UT-CA の実験協力部局として情報学環、情報基盤センター・本部合同を得て実験開始以来 1 年 5 ヶ月の間安定して運用できている。このことで当プロジェクトが提唱している分散 RA の運用モデルの有効性が明らかになった。

1.2 TRA

<http://www.pki.itc.u-tokyo.ac.jp/cerpj/>

NII が主導している CSI UPKI プロジェクトが開始した「オープンドメイン証明書発行実験」に参加し、パブリックサーバ証明書発行のための学内の体制を構築した。ここにおいても分散 RA 方式を取っている。2008 年 3 月末で参加部局は 6 部局を数え、サーバ証明書を発行した部局は 11 部署、67 件に及ぶ。

東大は、学内体制として TRA を立ち上げ、CP/CPS を公開した上で可能ならば発行にともなう審査を部局に委譲するモデルを取っている。これは従来からあった UTnet が安定して運用されていることに大きく依存するモデルであった。この運用モデルは成功し、参加部局の広がりが見られるようになった。

東大シールと一部オンライン化

TRA の審査は、部局のもつサブドメインまでを保証する形になっている点で通常のパブリックサーバ証明書と異なる。これらを保証する CP/CPS を公開していることを可視化する意味で東大シールを運用している。

審査過程は紙メディアで保存するが、それをのぞいた CSR や証明書の受け渡し、TLRA のサービスのために一部オンラインでの運用を開始した。

1.3 SSL-VPN 試行サービス

図書館のもつ電子ジャーナルを ECC の運用管理する ID 体系でサービスしている。理工系にとどまらず、広い分野でアクセスがなされている。新年度からはサービス可能な電子ジャーナルの数が大幅に増え、主だったところがサービス対象になった。

1.4 その他のサービス

SSO やシンクライアントへの対応を検討中である。

1.5 CSI UPKI プロジェクトへの係わり

NII が主導している CSI UPKI プロジェクトが開始した「オープンドメイン証明書発行実験」に参加し、パブリックサーバ証明書発行のための学内の体制を構築した。

さらに、インソースで CA を運用する場合の CP/CPS のガイドラインの策定に協力した。

1.6 教育活動

「パブリックサーバ証明書の発行及び運用に関するワークショップ」を開催し、サーバ証明書に関係するトピックを概観した。

さらに日本ペリサインと協力して情報セキュリティ基盤論(大学院新領域創成科学研究科)を開講した。さらに佐藤, 中山, 西村が「Web を支える情報通信技術」(全学ゼミナール)を開講した。

2 講習会・研究会開催報告

- UT-CA 報告会, 2007 年 4 月 27 日.
- パブリックサーバ証明書の発行及び運用に関するワークショップ(第 58 回コンピュータ・ネットワーク利用セミナー), 2007 年 9 月 3 日.

サーバ証明書導入にあたって～運用管理と統制のポイント～ / 監査を行う側から見た情報システム運用のポイント/東大におけるパブリックサーバ証明書発行体制の運用

授業提供

- 佐藤, 日本ベリサイン: 情報セキュリティ基盤論 (大学院新領域創成科学研究科)
- 佐藤, 中山, 西村: Web を支える情報通信技術 (全学ゼミナール)

3 付録論文

学外での発表

教員によるもの

- 西村, 佐藤: “東京大学における認証局 UT-CA の全学展開に向けた取り組み,” 第 21 回インターネット技術第 163 委員会研究会, 2007.
- 佐藤, 西村: “東京大学におけるキャンパス PKI の配備に向けて,” JNSA PKI Day 2007, 2007.
- 西村, 佐藤: “東京大学におけるサーバ証明書発行体制の構築と課題,” IPSJ DSM-48, 2008.

職員によるもの

- 大島: “東京大学におけるパブリックサーバ証明書発行プロジェクトについて,” 平成 19 年度実験・実習技術研究会, 2008.
- 大島: “東京大学におけるパブリックサーバ証明書発行プロジェクトについて,” 平成 19 年度核融合科学研究所技術研究会, 2008.

雑誌等取材

- MacFan, 2007 年 11 月号, p. 47, (No. 295).
- 平成 19 年度アクセス制御機能に関する技術の研究開発の状況等に関する調査 (警察庁) (<http://www.npa.go.jp/cyber/research/index.html>)

学内での発表

その他講演

- 西村: “パブリックサーバ証明書発行プロジェクトのご案内,” 第 5 回 UTnet ミーティング, 2007.
- 佐藤: “PKI システム – Digital ID とその信用,” 平成 19 年度技術職員研修 (コンピュータ関係), 2007.

広報

- 佐藤: “情報セキュリティ基盤論,” Digital Life, Vol. 9, p. 36, 2007.
- 西村: “パブリックサーバ証明書発行プロジェクトのご案内,” Digital Life, Vol. 9, pp.37–38, 2007.

- PKI プロジェクト: “UT-CA 報告会,” Digital Life, Vol. 9, pp.39–40, 2007.
- 大島: “デジタル世界の身分証明書その 4 – 情報基盤センター PKI プロジェクト ,” Digital Life, Vol. 9, pp. 41–45, 2007. –
- 西村: “PKI プロジェクトでのサーバ証明書の発行状況,” Digital Life, Vol. 10, p. 31, 2008.
- 大島: “デジタル世界の身分証明書その 5 – 情報基盤センター PKI プロジェクト ,” Digital Life, Vol. 10, pp. 32–36, 2008.

PART 4

その他

委員会委員等

講習会・セミナー

平成19年度委員会委員等

教員名	委員会委員等名	任 期
田中 哲朗	社団法人 情報処理学会 会誌編集委員会委員	18.4.1～20.3.31
	社団法人 情報処理学会 論文誌編集委員会編集委員	18.6.1～20.5.31
	社団法人 情報処理学会 論文賞委員会委員	18.9.13～19.5.30
	社団法人 情報処理学会 論文賞委員会委員	19.8.30～20.5.30
関谷 貴之	社団法人 情報処理学会 会誌編集委員会委員	18.4.1～20.3.31
丸山 一貴	早稲田大学 非常勤講師	19.10.1～20.3.31
中川 裕志	東京大学 キャンパス計画委員会本郷地区部会委員	18.4.1～20.3.31
	東京大学駒場図書館 運営委員会委員	18.4.1～20.3.31
	東京大学柏図書館 運営委員会委員	18.4.1～20.3.31
	北海道大学情報基盤センター 全国共同利用委員会委員	19.4.1～21.3.31
	三菱総合研究所 テキストインフォメーションハイディング研究会委員	19.4.18～20.3.31
	東京大学 学術諮問委員会委員	19.6.18～21.3.31
	日立コンサルティング 共通技術委員	19.8.3～20.3.31
	横浜国立大学学際プロジェクト研究センター 人事協議会情報学分野分科会委員	19.9.1～20.3.31
	日本学術振興会 科学研究費委員会専門委員	20.1.1～20.12.31
	清田 陽司	言語処理学会 年次大会プログラム委員会委員
	リッテル 上席研究員	19.8.1～20.7.31
若原 恭	社団法人 電子情報通信学会 執行委員会ソサイエティ副会長	17.5.28～19.5.27
	社団法人 電子情報通信学会 国際委員会委員	17.5.28～19.5.26
	東京大学 キャンパス計画委員会柏地区部会委員	18.4.1～20.3.31
	名古屋大学 情報連携基盤センター運営委員会委員	18.4.1～20.3.31
	独立行政法人 情報処理推進機構 情報処理技術者試験試験委員会委員	18.4.1～20.3.31
	東京大学 国際交流委員会専門委員	19.4.1～20.3.31
	東京大学 バリアフリー支援室室員	19.4.1～20.3.31
	東京大学 評価支援室室員	19.4.1～20.3.31
	東京大学 評価実施委員会委員	19.4.1～20.3.31
	東京大学 情報セキュリティ委員会 情報セキュリティ・ポリシー検討WG委員	19.4.1～20.3.31
	東京大学 技術職員研修臨時委員	19.4.1～20.3.31
	国立情報学研究所 学術情報ネットワーク運営・連携本部委員	19.4.1～20.3.31
	国立情報学研究所 学術情報ネットワーク運営・連携本部 ネットワーク作業部会委員	19.4.1～20.3.31

	社団法人 電子情報通信学会 ネットワークソフトウェア時限研究専門委員会専門委員	19.4.1～20.3.31
	社団法人 電子情報通信学会 ネットワークソフトウェア特集号編集委員会編集委員	19.4.13～21.4.1
中山 雅哉	総務省関東総合通信局 関東 JGNII 懇話会委員	16.11.16～20.3.31
	社団法人 電子情報通信学会 インターネットアーキテクチャ 研究専門委員会専門委員／副委員長	18.5.27～20.5.24
	国立情報学研究所 学術情報ネットワーク運営・連携本部 情報セキュリティポリシー策定作業部会委員	18.8.1～19.10.31
	東京大学 本郷地区事業場衛生委員会委員	19.4.1～20.3.31
	宇宙航空研究開発機構 宇宙利用推進本部衛星利用推進センター 主幹開発員	19.4.1～20.3.31
	総務省情報通信政策局 評価検討会構成員	20.2.15～20.3.31
	情報通信研究機構 新世代ネットワーク研究センター客員研究 員	20.3.1～21.3.31
加藤 朗	独立行政法人 情報通信研究機構 研究開発プロジェクト特別研究員	18.2.13～20.3.31
	独立行政法人 情報通信研究機構 次世代高度ネットワーク推進会議メンバー	18.3.6～20.3.31
	慶應義塾大学 SFC 研究所 上席所員（訪問）	19.4.1～20.3.31
	総務省情報通信政策局 次世代の情報セキュリティ政策に関する研究会構成員	19.10.23～20.6.30
小川 剛史	日本バーチャルリアリティ学会 サイバースペースと仮想都市研究委員会委員	19.4.1～21.3.31
	社団法人 情報処理学会 グループウェアとネットワークサー ビス研究運営委員会運営委員	19.4.1～20.3.31
	社団法人 情報処理学会 論文誌査読委員	19.4.1～21.3.31
	社団法人 情報処理学会 CollabTech 実行委員会委員	19.6.15～20.3.31
	社団法人 情報処理学会 インタラクシオン 2008 実行委員	19.8.28～20.3.4
中村 文隆	法政大学 非常勤講師	19.4.1～20.3.31
	社団法人 電子情報通信学会 ネットワークシステム研究専門委員会専門委員	19.5.25～20.5.24
	社団法人 電子情報通信学会 第7回アジア太平洋情報通信技 術シンポジウム Technical Program Committee 委員	19.9.5～20.7.31
関谷 勇司	慶應義塾大学 SFC 研究所 上席所員（訪問）	19.4.1～20.3.31
金田 康正	東京大学気候システム研究センター 研究協議会委員	17.4.1～20.3.31
	東京大学物性研究所 スーパーコンピュータ共同利用委員会委員	18.4.1～20.3.31

	東京大学物性研究所 スーパーコンピュータ共同利用課題審査委員会委員	18.4.1～20.3.31
	京都大学学術情報メディアセンター 全国共同利用運営委員会委員	18.4.1～20.3.31
	大阪大学サイバーメディアセンター 運営委員会委員	18.4.1～20.3.31
	国立環境研究所地球環境研究センター スーパーコンピュータ研究利用専門委員会委員	19.8.1～20.3.31
石川 裕	全国共同利用情報基盤センター長会議 スーパーコンピュータ研究会委員	18.12.1～20.11.30
佐藤 周行	国立情報学研究所 学術情報ネットワーク運営・連携本部 認証作業部会委員	19.4.1～21.3.31
	社団法人 情報処理学会 コンピューティングシステム論文誌 編集委員会編集委員	19.4.1～21.3.31
	日本ソフトウェア科学会 理事	19.4.1～21.3.31
黒田 久泰	社団法人 情報処理学会 論文誌編集委員会編集委員	19.6.1～21.5.31
片桐 孝洋	社団法人 情報処理学会 コンピューティングシステム論文誌 編集委員会編集委員	19.4.1～20.3.31
	社団法人 情報処理学会 HPCS2008 プログラム委員	19.7.1～20.1.31

平成19年度講習会・セミナー

開催日	講習会・セミナー名	担当係名
平成19年04月06日	総合図書館オリエンテーション「OPAC入門」	学術情報リテラシー係
平成19年04月09日	総合図書館オリエンテーション「OPAC入門」	学術情報リテラシー係
平成19年04月10日	データベース出張講習会(学際情報学府)	学術情報リテラシー係
平成19年04月10日	総合図書館オリエンテーション「OPAC入門」	学術情報リテラシー係
平成19年04月12日	総合図書館オリエンテーション「OPAC入門」	学術情報リテラシー係
平成19年04月12日	総合図書館オリエンテーション「OPAC入門」[英語編]	学術情報リテラシー係
平成19年04月13日	総合図書館オリエンテーション「OPAC入門」	学術情報リテラシー係
平成19年04月16日	総合図書館オリエンテーション「OPAC入門」	学術情報リテラシー係
平成19年04月16日	総合図書館オリエンテーション「OPAC入門」[英語編]	学術情報リテラシー係
平成19年04月17日	総合図書館オリエンテーション「OPAC入門」	学術情報リテラシー係
平成19年04月18日	総合図書館オリエンテーション「OPAC入門」[英語編]	学術情報リテラシー係
平成19年04月18日	総合図書館オリエンテーション「OPAC入門」	学術情報リテラシー係
平成19年04月19日	総合図書館オリエンテーション「OPAC入門」	学術情報リテラシー係
平成19年04月20日	データベース出張講習会(教養学部「基礎演習」)	学術情報リテラシー係
平成19年04月20日	データベース出張講習会(教養学部「基礎演習」)	学術情報リテラシー係
平成19年04月20日	総合図書館オリエンテーション「OPAC入門」	学術情報リテラシー係
平成19年04月23日	データベース出張講習会(教養学部「基礎演習」)	学術情報リテラシー係
平成19年04月24日	データベース出張講習会(教養学部「基礎演習」)	学術情報リテラシー係
平成19年04月27日	データベース出張講習会(教養学部「基礎演習」)	学術情報リテラシー係
平成19年05月07日	データベース出張講習会(教養学部「基礎演習」)	学術情報リテラシー係
平成19年05月09日	第56回コンピュータ・ネットワーク利用セミナー「Adobe Flashテクノロジーを使った授業・研究支援ツールについて～FlashベースのWeb会議システム「Acrobat Connect」ほか」	アプリケーション支援係
平成19年05月10日	データベース出張講習会(新規採用職員研修)	学術情報リテラシー係
平成19年05月11日	情報探索ガイダンス「日本の論文を探すには？」	学術情報リテラシー係
平成19年05月17日	情報探索ガイダンス「電子ジャーナルを利用するには？」	学術情報リテラシー係
平成19年05月22日	情報探索ガイダンスElectronic Journals Course [英語編]	学術情報リテラシー係
平成19年05月23日	情報探索ガイダンス「自宅から検索するには？」	学術情報リテラシー係

開催日	講習会・セミナー名	担当係名
平成19年05月28日	第57回コンピュータ・ネットワーク利用セミナー「Autodesk Buzzsawの活用事例とその機能について」	アプリケーション支援係
平成19年05月30日	情報探索ガイダンス「LRC with MLA」	学術情報リテラシー係
平成19年05月30日	情報探索ガイダンス「InfoTrac Custom」	学術情報リテラシー係
平成19年06月04日	情報探索ガイダンス「AGRICOLA」	学術情報リテラシー係
平成19年06月06日	情報探索ガイダンス「Business Source Elite/SocINDEX」	学術情報リテラシー係
平成19年06月06日	情報探索ガイダンス「PsycINFO/PsycARTICLES」	学術情報リテラシー係
平成19年06月07日	情報探索ガイダンス「LexisNexis Academic」	学術情報リテラシー係
平成19年06月07日	情報探索ガイダンス「Lexis.com」	学術情報リテラシー係
平成19年06月08日	情報探索ガイダンス「ASSIA/LLBA」	学術情報リテラシー係
平成19年06月11日	情報探索ガイダンス「ECCO」	学術情報リテラシー係
平成19年06月12日	情報探索ガイダンス「JapanKnowledge」	学術情報リテラシー係
平成19年06月13日	情報探索ガイダンス「eol ESPer」	学術情報リテラシー係
平成19年06月14日	データベース出張講習会(文学部Academic Writing)	学術情報リテラシー係
平成19年06月15日	情報探索ガイダンス「JSTOR」	学術情報リテラシー係
平成19年06月18日	情報探索ガイダンス「ASSIA/LLBA」	学術情報リテラシー係
平成19年06月19日	情報探索ガイダンス「SciFinder Scholar」	学術情報リテラシー係
平成19年06月20日	情報探索ガイダンス「Web of Science人文社会科学系」	学術情報リテラシー係
平成19年06月20日	情報探索ガイダンス「Web of Science自然科学系」	学術情報リテラシー係
平成19年06月20日	情報探索ガイダンス「BIOSIS Previews」	学術情報リテラシー係
平成19年06月20日	情報探索ガイダンス「EndNote Web」	学術情報リテラシー係
平成19年06月21日	情報探索ガイダンス「ECCO」	学術情報リテラシー係
平成19年06月22日	情報探索ガイダンス「SciFinder Scholar」	学術情報リテラシー係
平成19年06月26日	情報探索ガイダンスWeb of Science Course [英語編]	学術情報リテラシー係
平成19年06月27日	データベース出張講習会(理学部生物学科)	学術情報リテラシー係
平成19年06月28日	情報探索ガイダンス「SciFinder Scholar」	学術情報リテラシー係
平成19年06月29日	情報探索ガイダンス「Web of Science自然科学系」	学術情報リテラシー係
平成19年06月29日	情報探索ガイダンス「Web of Science人文社会科学系」	学術情報リテラシー係
平成19年06月29日	情報探索ガイダンス「EndNote Web」	学術情報リテラシー係
平成19年07月03日	情報探索ガイダンス「聞蔵II ビジュアル」	学術情報リテラシー係

開催日	講習会・セミナー名	担当係名
平成19年07月03日	データベース出張講習会(工学系研究科社会基盤学専攻交通研究室)	学術情報リテラシー係
平成19年07月05日	データベース出張講習会(工学系研究科社会基盤学専攻国際プロジェクト)	学術情報リテラシー係
平成19年07月05日	情報探索ガイダンス「MC-統計」	学術情報リテラシー係
平成19年07月06日	情報探索ガイダンスElectronic Journals Course [英語編]	学術情報リテラシー係
平成19年07月09日	情報探索ガイダンス「自宅から検索するには？」	学術情報リテラシー係
平成19年07月24日	JDream2説明会	学術情報リテラシー係
平成19年08月02日	事務職員向け情報探索ガイダンス	学術情報リテラシー係
平成19年08月08日	事務職員向け情報探索ガイダンス	学術情報リテラシー係
平成19年08月22日	Littel Navigator説明会	学術情報リテラシー係
平成19年09月03日	第58回コンピュータ・ネットワーク利用セミナー「パブリックサーバ証明書の発行及び運用に関するワークショップ」	アプリケーション支援係
平成19年09月07日	情報探索ガイダンス「JapanKnowledge」	学術情報リテラシー係
平成19年09月13日	情報探索ガイダンス「新聞記事を探すには？」	学術情報リテラシー係
平成19年09月14日	図書館職員のためのアプリケーション開発講習会(第1回)	図書館情報係
平成19年09月18日	情報探索ガイダンス「図書館の蔵書を検索するには？」	学術情報リテラシー係
平成19年09月26日	情報探索ガイダンス「Web of Science」	学術情報リテラシー係
平成19年09月27日	図書系職員のためのアプリケーション開発講習会(本郷地区第1回欠席者対象)	図書館情報係
平成19年09月28日	情報探索ガイダンスWeb of Science Course [英語編]	学術情報リテラシー係
平成19年10月03日	図書館職員のためのアプリケーション開発講習会(駒場地区第1回欠席者対象)	図書館情報係
平成19年10月04日	データベース出張講習会(工学部システム創成学科)	学術情報リテラシー係
平成19年10月10日	留学生向け総合図書館ガイダンス&データベース講習会	学術情報リテラシー係
平成19年10月12日	図書系職員のためのアプリケーション開発講習会(第2回A)	図書館情報係
平成19年10月12日	情報探索ガイダンス「電子ジャーナルを利用するには？」	学術情報リテラシー係
平成19年10月12日	留学生向け総合図書館ガイダンス&データベース講習会[英語編]	学術情報リテラシー係
平成19年10月16日	留学生向け総合図書館ガイダンス&データベース講習会[英語編]	学術情報リテラシー係
平成19年10月18日	情報探索ガイダンス「日本の論文を探すには？」	学術情報リテラシー係
平成19年10月19日	図書系職員のためのアプリケーション開発講習会(第2回B)	図書館情報係
平成19年10月24日	情報探索ガイダンス「自宅から検索するには？」	学術情報リテラシー係

開催日	講習会・セミナー名	担当係名
平成19年10月25日	第5回UTnetミーティング	キャンパスネットワーク係
平成19年10月30日	情報探索ガイダンス「Web of Science」	学術情報リテラシー係
平成19年11月01日	情報探索ガイダンス「JapanKnowledge」	学術情報リテラシー係
平成19年11月07日	留学生向け情報探索ガイダンス[中国語編]	学術情報リテラシー係
平成19年11月08日	データベース出張講習会(文学部社会心理学研究室)	学術情報リテラシー係
平成19年11月09日	図書系職員のためのアプリケーション開発講習会(第3回A)	図書館情報係
平成19年11月13日	留学生向け情報探索ガイダンス[韓国語編]	学術情報リテラシー係
平成19年11月14日	情報探索ガイダンス「LexisNexis Academic」	学術情報リテラシー係
平成19年11月16日	図書系職員のためのアプリケーション開発講習会(第3回B)	図書館情報係
平成19年11月16日	留学生向け情報探索ガイダンス[英語編]	学術情報リテラシー係
平成19年11月22日	留学生向け情報探索ガイダンス[日本語編]	学術情報リテラシー係
平成19年11月27日 ～ 11月29日	第59回コンピュータ・ネットワーク利用セミナー「平成19年度 東京大学技術職員研修(コンピュータ関係)公開講義」	アプリケーション支援係
平成19年11月29日	データベース出張講習会(文学部Academic Writing)	学術情報リテラシー係
平成19年12月04日	データベース出張講習会(教養学部「英語」)	学術情報リテラシー係
平成19年12月04日	情報探索ガイダンス「電子ジャーナルを利用するには？」	学術情報リテラシー係
平成19年12月07日	図書系職員のためのアプリケーション開発講習会(第4回A)	図書館情報係
平成19年12月07日	情報探索ガイダンスElectronic Journals Course [英語編]	学術情報リテラシー係
平成19年12月12日	情報探索ガイダンス「電子ジャーナルを利用するには？」	学術情報リテラシー係
平成19年12月12日	第60回コンピュータ・ネットワーク利用セミナー 「Mathematica6の活用例とその機能について」	アプリケーション支援係
平成19年12月14日	図書系職員のためのアプリケーション開発講習会(第4回B)	図書館情報係
平成20年01月10日	図書系職員のためのアプリケーション開発講習会(第5回A)	図書館情報係
平成20年01月11日	情報探索ガイダンス「新聞記事を探すには？」	学術情報リテラシー係
平成20年01月17日	図書系職員のためのアプリケーション開発講習会(第5回B)	図書館情報係
平成20年01月17日	情報探索ガイダンス「JapanKnowledge」	学術情報リテラシー係
平成20年01月23日	情報探索ガイダンス「日本の論文を探すには？」	学術情報リテラシー係
平成20年01月29日	情報探索ガイダンス「自宅から検索するには？」	学術情報リテラシー係
平成20年02月04日	情報探索ガイダンス「Web of Science」	学術情報リテラシー係
平成20年02月05日	図書系職員のためのアプリケーション開発講習会(第6回A)	図書館情報係
平成20年02月12日	情報探索ガイダンスWeb of Science Course [英語編]	学術情報リテラシー係

開催日	講習会・セミナー名	担当係名
平成20年02月13日	図書系職員のためのアプリケーション開発講習会(第6回B)	図書館情報係
平成20年02月15日	情報探索ガイダンス「電子ジャーナルを利用するには？」	学術情報リテラシー係
平成20年02月28日	情報探索ガイダンス「Web of Science」	学術情報リテラシー係
平成20年03月07日	情報探索ガイダンス「自宅から検索するには？」	学術情報リテラシー係
平成20年03月13日	情報探索ガイダンス「電子ジャーナルを利用するには？」	学術情報リテラシー係
平成20年03月19日	情報探索ガイダンス「図書館の蔵書を検索するには？」	学術情報リテラシー係

Web 掲載版

(2008年10月 www.itc.u-tokyo.ac.jp 掲載)

東京大学情報基盤センター年報

2007年度(第9号)

東京大学情報基盤センター年報編集スタッフ

編集委員長

小川 剛史

編集委員

田中 哲朗, 二宮 崇, 片桐 孝洋, 西村 健

編集スタッフ

大塚 浩一, 柏 芳美, 本多 玄, 中山 仁史, 平野
光敏, 佐藤 安一郎, 米山 浩, 灰塚 毅弘, 佐々木
一孝, 大島 大輔

編集

東京大学情報基盤センター年報編集委員会

発行

東京大学情報基盤センター

〒113-8658 東京都文京区弥生 2-11-16

電話 03(5841)2710

印刷

よしみ工産株式会社

2008年6月発行



東京大学
THE UNIVERSITY OF TOKYO