



HOKKAIDO UNIVERSITY

Title	ERATO湊離散構造処理系プロジェクト：キックオフシンポジウム講演資料
Author(s)	湊，真一
Description	ERATO 湊離散構造処理系プロジェクト春のワークショップ（キックオフシンポジウム）．2010年5月28日（金）～29日（土）．ERATO湊プロジェクト研究室．
Relation	2010年度科学技術振興機構ERATO湊離散構造処理系プロジェクト講究録．p.316-323.
Issue Date	2011-06
Doc URL	https://hdl.handle.net/2115/48396
Type	conference presentation
File Information	16.minato_06.pdf





ERATO 湊離散構造処理系プロジェクト

キックオフシンポジウム講演資料

平成22年5月28日

研究総括 湊 真一
(北海道大学 大学院 情報科学研究科 准教授)



ERATOについて

- JSTの戦略的創造研究推進事業
 - さきがけ(牧場型)、CREST(ハケ岳型)、**ERATO(富士山型)**
- ERATO型の特徴
 - 新しい科学技術の源流を作るような研究を支援。
 - 昭和56年発足。過去に98プロジェクトを採択。
(科学技術の全分野で年4~5件)
 - プロジェクト期間: 5年半 研究費総額: 10~15億円(テーマによる)
 - メンバ規模: 10~15人(人件費に依存)
 - 研究総括に自由裁量を与え、分野・組織にとらわれずに機動的なプロジェクトを構成。
 - 大学への補助金ではなく、JST直轄事業。
本務研究室とは別に独自のオフィスを設置。
 - ERATO専任の事務スタッフをJSTが雇用。

3



ERATO総括の選考過程

- 他薦およびJST独自調査により、母集団リスト(1000人規模)を構成。
- JSTで分野ごとに審査委員長(PO: プログラムオフィサー)を1名選定。
POは審査員4名(うち外国人1名)を指名し、審査パネルを構成。
- 審査パネルで母集団を絞り込み、各分野で10名程度の一次候補者をノミネート。
- この段階で初めて一次候補者に通知し、研究構想の応募を依頼。
 - 候補者は突然の連絡で驚き悩む。(数名は応募辞退)
 - 研究テーマは完全に自由 (国の重点分野のどこかに該当すれば良い)
(といっても、何を期待されて推薦されたのか、空気を読む必要あり)
 - 提案書は、ほぼ無地の様式で英語10ページ、和訳も添付する。
 - 研究分担者を指名・事前協議できない。個人の構想だけで勝負。
- 一次面接(または書類選考)、最終面接で、各分野1人だけ採択
 - 最終面接は英語で1時間以上
 - 審査パネルの意見を参考に、POが最終決定(合議制ではない。)

4



過去の情報系のERATOプロジェクト

- 北野共生システム (1998~2003)
- 今井量子計算機構 (2000~2005)
- 合原複雑数理モデル (2003~2008)
- 浅田共創知能システム (2005~2010)
- 五十嵐デザインインタフェース (2007~2012)
- 岡ノ谷情動情報 (2008~2013)

ERATO発足当初は、物理、化学、医学生命のみ(→ノーベル賞分野)
情報科学系のプロジェクトが採択されるようになったのは、最近10年。
しかも最初は他分野との境界領域テーマでないと取れなかった
(3年前までは分野別の審査ではなく、全分野の審査員で合議だった)

アルゴリズム技術を中心とするERATOプロジェクトは画期的

5

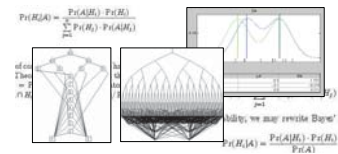


本プロジェクトの概要



離散構造とその処理

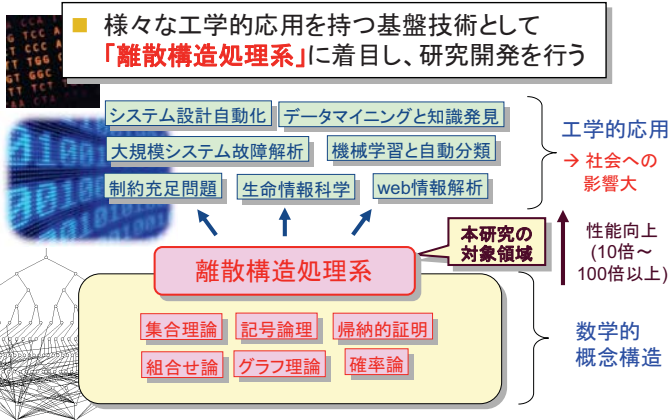
- 離散構造とは
 - **離散数学および計算機科学の基礎をなす数学的構造**
 - 集合理論、記号論理、帰納的証明、グラフ理論、組合せ論、確率論などを含む
- およそ計算機が扱うあらゆる問題は、単純な基本演算を要素とする離散構造の処理に帰着される。
 - 最終的には、**膨大な数の場合分け処理を要することが多い。**
- 離散構造の処理は、計算機の様々な応用分野に共通する基盤技術
 - 典型的な応用分野: システム設計自動化、大規模システム故障解析、制約充足問題、データマイニングと知識発見、機械学習と自動分類、生命情報科学、web情報解析、etc.
 - **現代情報化社会に対する波及効果は極めて大。**



6

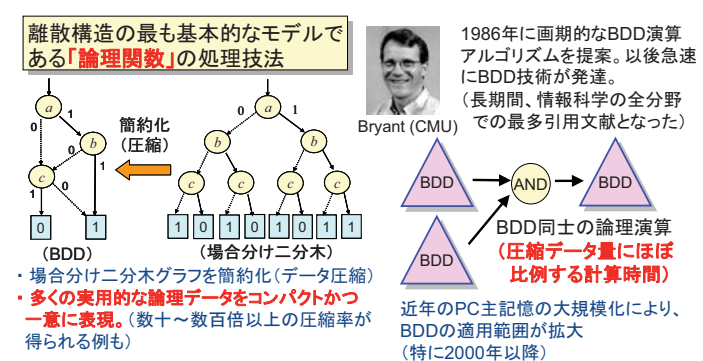
7

本プロジェクトの基本構想



8

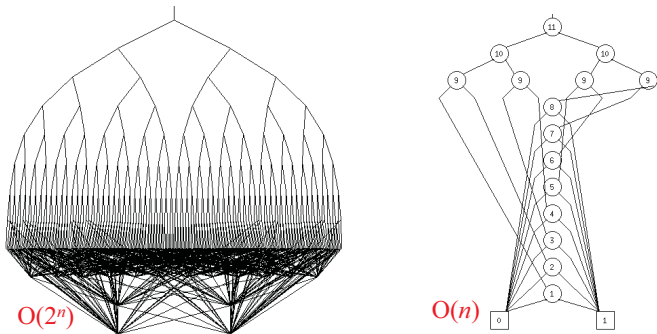
BDD(二分決定グラフ)



9

BDD簡約化の効果

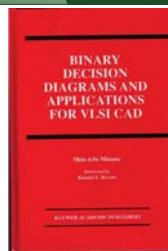
- 特定の問題では、指数関数的な圧縮効果が得られる。
- 例題に依存するが、多くの実用的な問題で効果がある。



10

研究総括の本分野での業績

- 学生時代(1988年)よりBDD研究を開始。
 - 黎明期より一貫してBDD研究に従事。
- 研究成果を単行本として米国で出版
S. Minato: "Binary Decision Diagrams and Applications for VLSI CAD" (Kluwer, 1996).
 - Bryant先生に前書きを寄稿して頂いた
 - 欧米の理工系大学院で副教材として使用された実績を持つ
- 組合せ集合を表現するデータ構造「ZDD」を考案(1993).
 - VLSI論理最適化システムのために開発。(1993～2003)
 - データマイニング・知識処理にも応用。(2004～)
 - Knuthの著名な教科書にも掲載、執筆作業に協力 (2008～2009)



11

論理関数と組合せ集合

a	b	c	F
0	0	0	0
1	0	0	0
0	1	0	0
1	1	0	1
0	0	1	1
1	0	1	1
0	1	1	0
1	1	1	0

論理関数:
 $F = (a \wedge b \wedge \neg c) \vee (\neg b \wedge c)$

組合せ集合:
 $F = \{ab, ac, c\}$

→ ab (買い物客の購入品)

→ c

→ ac

■ 組合せ集合と論理関数の演算は対応関係がある。

- Union of sets → logical OR
- Intersection of sets → logical AND
- Complement set → logical NOT

12

ZDD(ゼロサプレス型BDD)による集合の表現

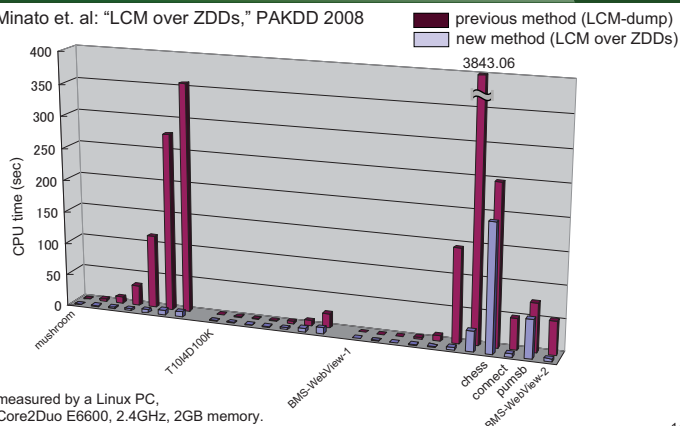
- 「組合せ集合」を効率的に表現するためのBDDの改良
 - 浅が世界で初めて考案し命名(1993年)
 - 通常と異なる簡約化規則を考案。
 - 疎な集合の族を扱う場合に著しい効果が得られる。
(例: 商店の陳列アイテム数に比べて1顧客の購入点数は極めて少ない。)
-
- 通常のBDD
- ZDDの簡約化
- ZDDはBDDの改良技術として現在、世界的に広く使われている。
 - 最近では、データマイニング分野に应用されて、画期的な有効性が示されている。(数百倍のデータ圧縮率・数十倍の処理高速化)
 - 他にも応用例は増えつつある。

13



Frequent itemset mining using ZDDs

Minato et. al: "LCM over ZDDs," PAKDD 2008



measured by a Linux PC,
Core2Duo E6600, 2.4GHz, 2GB memory.

14



BDD / ZDD に関する最近の話題

- Knuthの名著「The Art of Computer Programming」の最新巻 (Vol.4, Fascicle 1, 2009)で、BDDが詳しく解説された。

- BDDだけで140ページ、演習264問の大作

- 種々のBDD改良技法の中で、特にZDDを重要視

- 章の後半30ページ以上にわたり、ZDDについて記述

- ZBDDだった呼び名をZDDに短縮。

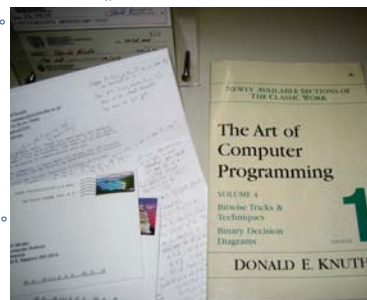
- (BDDと同格扱いに)

- ZDD演算体系を再構築し、「Family Algebra」と命名。

- 日本人の研究成果が、このシリーズに項目として詳細に掲載されるのは初めて。

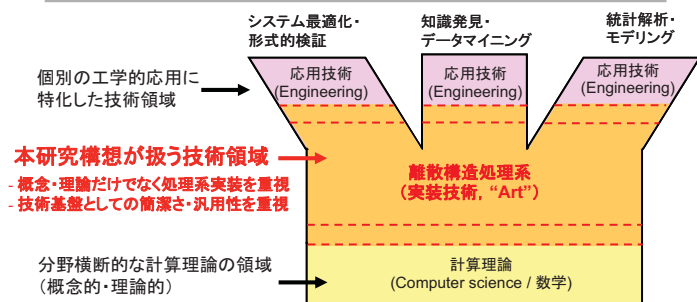
- Knuth氏本人から、ZDD考案者として校正作業への協力を依頼する長文のメールと直筆の手紙を受領。

- Knuth氏は、オックスフォードとスタンフォードで「Fun with ZDDs」というタイトルで特別講義。



本研究構想の対象領域

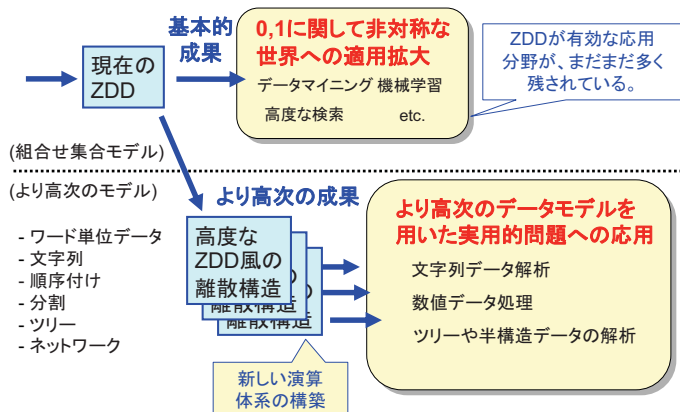
BDD/ZDD技術の新しい切り口として、様々な離散構造を統合的に演算処理する技法を体系化し、分野横断的かつ大規模な実問題を高速に処理するための技術基盤を構築する。



16



本研究プロジェクトの技術的戦略



17



本プロジェクトの成果目標

- BDD/ZDD研究を新しい視点でとらえ直し、汎用的な離散構造処理系として再構築を行う
 - 社会的重要性が高い複数の応用に適用し、従来技法に比べて、10倍~100倍以上の性能向上を達成する。
- 長年に渡って活用される基盤ソフトウェアを提供する。および、それを使いこなせる人材を育成し、産業界に還元する。
 - 形態素解析システムChasen, 連想検索エンジンGetaのように、10~15年が経過しても使われる基盤ソフトウェア (→ 直接的・間接的に、過去の国家プロジェクトの成果が形になったもの)
- 離散構造処理系という技術分野を深く追究し、長年に渡って国内外から参照されるような学術専門書にまとめること
 - 教科書が持つ長期的な影響力は極めて大きい。

18



プロジェクトの立ち上げ状況

19



ERATOの雇用制度

- 研究総括(大学教員兼業)
- 事務スタッフ
 - 技術参事、事務参事、事務員、各1名(JST直接雇用)
- サブグループ(2~4グループ)
 - グループリーダー各1名(大学教員兼業またはJST直接雇用)
 - 博士研究員、研究員、技術員、研究補助員、合計10~15名(兼業不可・JST直接雇用)

JST直接雇用者は、大学のポストを得ていない若手ポスドク、または、民間企業からの出向者(JSTが所属企業に給与を補填)

大学教員・国研から正式メンバに入れるのは、総括とGLだけ

JST ERATOと大学との共同研究の形で、大学教員に研究費を出すことは可能。(研究費を伴わない共同研究も可)

20



研究グループ体制

- 離散構造基盤グループ (GL: 研究総括兼任)
 - 札幌メインオフィス(北大・情報科学研究科)
 - ポスドク研究員 現在日本人5名(白井・櫻井・吉仲・川原・齋藤)
(若手ポスドクは原則2年で次のポストを獲得するよう奨励)
 - 今後、企業出向者・海外ポスドクの勧誘を本格化
 - 事務スタッフ3名
・ 技術参事(白井、ポスドク兼務)、事務参事(齋藤)、事務員(渡辺)
- 機械学習・制約充足応用グループ(GL: 産総研・津田宏治氏)
 - 東京・大岡山(東工大GCOE「計算世界観」)と共同研究)
 - ポスドク研究員1名(田部井) + 東工大・東京地区研究者と連携
- 統計・マイニング応用グループ (GL: 阪大・鷲尾隆先生)
 - 大阪・梅田・民間オフィス(阪急梅田駅北口すぐ)
 - 技術員1名(中元) 補助員1名(金光) + 関西地区研究者と連携

21



ERATOオフィス設置場所

- 札幌にメインオフィスを設置
 - 北大キャンパス内に設置
 - 空港・駅からの交通至便
 - 北大の関連研究者とも交流が容易
→ 研究に集中できる優れた立地環境
- 東京・関西地区にサブオフィスを設置
 - 東工大と大阪梅田に設置
 - 福岡(九大)にも連絡拠点(共同研究)を計画中
 - 高品質TV会議システムを導入し、十分な意思疎通を図る。



フルタイムの若手研究者は極力、札幌に常駐させて、活性化・相乗効果を図る。



札幌オフィス

- 情報科学研究科に隣接する校舎内(工学研究科)
 - 約300m²の独立スペース(建物の端・廊下突き当り)
 - 北大情報系の研究者との交流が容易。(工学部食堂のほぼ真上の3階)
 - 耐震補強工事のための玉突き移動により、このタイミングで偶然出現したまとまったスペース
 - 部局長の特別の措置でERATOに貸出し

情報科学研究科(本務研究室)

ERATOオフィス予定地(C301~306)

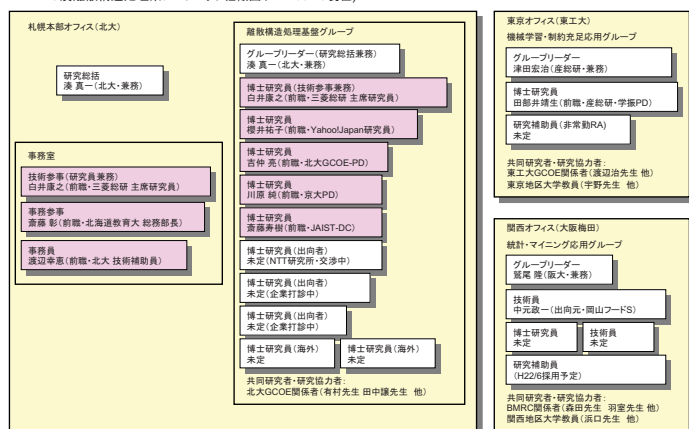


当初の候補地域



北大オフィスレイアウトイメージ

ERATO 離散構造処理系プロジェクト組織図 (H22.05.28現在)



25

他機関との共同研究の立ち上げ状況

- サブグループ設置拠点
 - 北大(湊) 東工大(渡辺治、津田) 阪大(鷲尾)
- 東京地区
 - 早稲田大学(林、計画中) 電通大(植野、計画中) 他？
- 関西地区
 - 大阪府大(森田) 関学(羽室) 京大(加藤先生、山本先生)他？
- 九州地区
 - 九大(竹田、計画中)
- 企業
 - 米国 Cadence Berkeley Lab.(渡部)
 - NTT(交渉中) 日本IBM(打診中) 富士通(打診中) 日立(打診中) 他？

1件ごとに契約条件について個別交渉が必要なため、結構手間がかかる。

研究総括が契約に深く関わるのは10件くらいが限界かも...

26

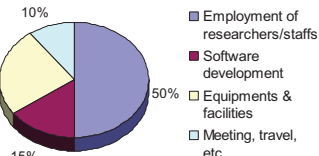
3段階の連携体制

- 「ERATO 세미나会員」を募集
 - ERATOメーリングリストへの登録・イベント案内の送付
 - 定例のERATOセミナーへの自由参加
 - 好きな時間に北大オフィスに立ち寄り、ERATOメンバとお茶を飲みながら議論して頂く
 - 北大教員(希望者)には北大オフィスの夜間休日カードキー貸与
- ERATOメンバとの具体的な共同研究テーマが定まって来たら、数人単位で「ERATOミニプロジェクト」を登録
 - 連携研究者として共著論文執筆や基盤ソフト開発等に参加。
 - 必要な研究費(旅費・謝金・物品等)はJST直轄予算からサポート。
- 知財や雇用が生じる場合は正式な共同研究契約
 - 組織間の契約手続きが必要

27

研究費の主な使途

- 総額11億円(5年半の合計) 年平均2億円
- 人件費(50%強)
 - 大学から1/3、企業から1/3、海外から1/3 を目標にバランスよく集めたい
- ソフトウェア開発費(15%)
 - 核心部分は内製するが、本質でないGUIやデモ機能は外注を検討。
 - 内製プログラムをそのまま仕様として、同機能の実用版を外注し、ドキュメント作成や保守要員確保を行うことを検討。
- 設備費(25%)
 - 超メモリサーバ(超大規模・単一実メモリを搭載した汎用高性能PC) 初年度 50~100GB規模。2年後に1TB級のメモリ容量
 - 離散構造処理系に特化した専用ハードウェアのプロトタイプ開発費用
 - オフィスレンタル費用、机椅子
 - その他の什器
- 会議費・通信費・旅費(10%)
 - 国際ワークショップを年1回、国内研究会を年4回程度企画。(共催含む)
 - 高精細TV会議システムの導入・運用



28

研究の話題

29

Knuth邸訪問

- 4月末:
国際会議SDM2010(SIAM Data Mining Conference) 発表のため米国出張(NII・宇野先生と共著)
- 往路にStanford大学・Knuth氏のオフィス(自宅)を訪問
 - 湊から電子メールにて訪問を依頼
 - 教科書掲載の挨拶と学生への小切手受領のお礼
 - 最近のBDD/ZDD関連技術に関するディスカッション
 - 宇野先生も同行



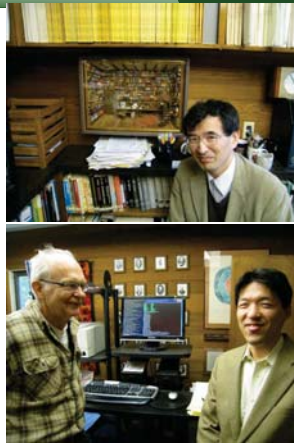
Knuth邸外観

- Knuth邸は、Stanford大学キャンパス内の教職員居住地区に所在。大学との最長80年(1968~2048)のリース契約



大学に教員室はあるが、基本的に仕事は自宅の書斎
名誉教授のため教務は無し
年1回の特別講義と週1回の教員昼食会だけ出勤

Knuth邸内 (パイプオルガンと書斎)



書斎の本棚



教科書執筆の保存資料



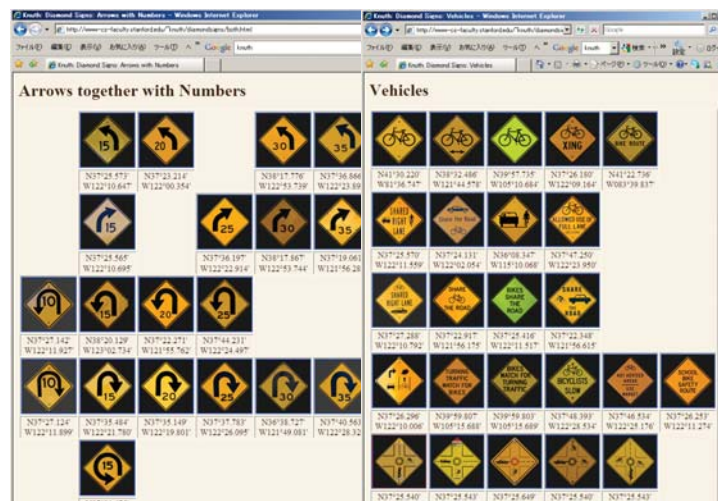
京都賞メダル、小物ディスプレイ棚



手土産と記念の品



Knuth邸の庭の
「松ぼっくり」





Knuth邸での議論

- 有限系列集合を表現する「Sequence BDD」とその応用について議論
 - 非常に興味を持っておられた。
- Knuth本にZDDが載ったことがきっかけで、私が大変なことになってしまっていることを報告
 - 「自分が教科書に何かを書くと、いろいろな人に影響を与えることは認識している。だから私にも少しは責任がある」
 - 「じゃあプロジェクトの今後の研究テーマについて、少し語ろう」
 - Knuth先生が考えるFuture workをいくつか挙げて下さった。基本的には、私がERATOの面接資料に書いた内容とほぼ一致。「ZDDをベースとした、より高度な離散構造のAlgebraの構築」

プロジェクトの方向性は間違っていないことを確認し、意を強くした

38



ZDDの代数系

- ZDDのデータ構造としての側面だけではなく、代数系(algebra)を重視。

$\emptyset, \{1\}$	Empty and singleton set. (0/1-terminal)	基本演算集合 (ブール代数に対応)
$P.top$	Returns the <i>item-ID</i> at the top node of P .	
$P.onset(v)$ $P.offset(v)$	Selects the subset of itemsets including or excluding v .	
$P.change(v)$	Switching v (add / delete) on each itemset.	
$U, \cap, \setminus$	Returns <i>union</i> , <i>intersection</i> , and <i>difference set</i> .	
$P.count$	Counts <i>number</i> of combinations in P .	湊が導入した 拡張演算集合
$P * Q$	Cartesian product set of P and Q .	
P / Q	Quotient set of P divided by Q .	
$P \% Q$	Reminder set of P divided by Q .	多くの実用的な 応用がある。

当初は“Unate cube set algebra”と呼んでいたが、Knuth 先生が“Family algebra”という名前で見直し。

39



北大版BDD/ZDDパッケージ

- 北大版BDD/ZDDパッケージ
 - 湊がNTTから北大に移籍した最初の年(2004年)に開発
 - NTT時代のパッケージを作り直した(特許・著作権問題なし)
- 現在、パッケージのソースは一般公開していない。
 - 法的には公開して問題ないが、質問や問合せに対応できない。
 - バージョンが一元管理できる体制が作れていない
 - まだドキュメントが日本語のみ
 - 現在でも開発が進んでいて、ときどき内部をいじっている
 - 個人的に付き合いのある人にだけソースを開示
 - BEM-II、VSOPというBDD/ZDD処理インタプリタは公開済み
- ERATO期間内に整備してオープンソースとして公開したい。
 - 現在のコードを仕様書として、再コーディングを外注するか
 - ドキュメントの整備、一定期間のサポート体制
 - Rubyのようなスクリプト言語版も整備したい。(関西グループ)
 - 他のBDDパッケージとの性能比較もたぶんやった方がよい。

40



超メモリーサーバ

- 昨年度予算で、512GBの単一実メモリ空間を専有できるLinuxサーバを購入
 - 北大オフィスの空調・防音サーバ室に設置
 - AMD opteron 64bit SuSE Linux
 - 価格: 1500万円強 (ほとんどがメモリ代です)
 - 数年前よりは相当安くなった。クラウドサービス等の市場が広がって来たためと思われる
- 100億ノードまでの大規模BDD/ZDDをメモリ上に構築可能
 - BDDパッケージを今週インストールして、昨日、正常動作を確認。
 - 約85億ノードのBDDを構築するのに約300分かかる。
- 他にも応用の可能性
 - 超巨大Suffix-treeの構築
 - In-memory Databaseの処理
- ハードディスクが足りなかった(1TB?)ので追加を計画中
 - メモリに全財産(?)を投入したため

41



ERATO 세미나

- ERATOの日常行事として、定例のセミナーを開催します。
 - 原則ほぼ毎週金曜 15:30より 北大ERATOオフィスにて
 - 専用TV会議サービスにより、東京・大阪のサブオフィスと接続
- セミナ内容
 - ポスドク研究員が持ちまわりで最近の研究成果発表
 - 北大教員・大学院生の招待講演
 - 北大訪問研究者による招待講演 (皆さま、年に1, 2回程度は、北大にお越し下さい。)
- ERATOセミナー会員
 - メールリストでセミナーの案内を送付
 - 興味あるテーマのときに自由参加(お近くの拠点へ)
 - 立ち寄りついでにERATOメンバと研究打合せも

42



まとめ

- ソフトウェア系のERATOは、特殊な実験装置はないので、優れた研究者集団を形成するための環境作りが最も重要。
 - 日本における研究者の人材集積地は東京と関西であるが、そこから離れた札幌で実施するということが、非常に特徴的である。
 - プログラムオフィサー(西尾章治郎先生)からは「札幌を、優秀な人材が集まる梁山泊のようにして欲しい」との激励を頂いている。
 - 東京・関西のサブグループ拠点と活発な研究交流を行いながら、優秀な若手人材を見出し、札幌に引っ張って来て、高いレベルの「梁山泊」を形成 (→ コアとなる研究はじっくり札幌で行う。)
 - 将来、育った人材を中央に還流させるときにも、サブグループ拠点は重要な機能を果たす。

43



今後のイベント

- 6/18(金) 西尾章治郎先生 北大オフィス来訪
 - 学内関係者向けプロジェクト説明会・見学会を開催
- 7/30(金) OR学会NEO研究会を北大オフィスで開催
 - 幹事: 京大・伊藤大雄先生
 - ERATOセミナーと共同開催
- 9/8(水)午前 FIT-2010 企画セッション(九州大学にて)
 - 研究総括によるプロジェクト概要紹介(30分)
 - ERATOプロジェクト関係者によるポスターセッション
各発表者ポジショントーク45分+ポスターセッション75分
- その他、共催イベント募集中