

# 半永続データの自己組織化管理に関する研究

成 凱  
Kai CHENG

九州産業大学 情報科学部 社会情報システム学科  
Faculty of Information Science, Kyushu Sangyo University  
chengk@is.kyusan-u.ac.jp, <http://www.is.kyusan-u.ac.jp/~chengk/>

## 1. 研究目的

Web とデータベース技術を融合することにより情報配信のコストが大幅に低くなり新しい情報流通革命を起こしつつある。大量の情報を容易に入手できると同時に情報の新陳代謝も激しくなり効率よく管理することが困難となった。従来のデータ管理技術では永続データ (Persistent Data: 明示的に変更・削除しない限り、永遠に保存し続けるデータ) とストリームデータ (Stream Data: 大量で高速に到来し続けるデータの流れ) という極端なケースしか考えておらず、新陳代謝の激しいデータに対応できない。本研究では、明示的或は暗黙的に寿命が付けられ、期限内のみ保存されるデータを半永続データ (Semi-Persistent Data) と捉え、そのようなデータの性質の解明とモデル化を図る。

## 2. 主な研究結果

本研究 (研究代表者: 成, 研究分担者: 牛島和夫教授) は萌芽研究\*1として平成 17 年に採用され平成 17 年度～18 年度二年間にわたって研究を行ってきた。

17 年度は、半永続データのモデル化を中心に研究を進めた。まず、データの構造化と永続性の関係に基づいて従来のデータモデルとデータ管理システムの分類方法を提案した。半永続構造化データと半永続半構造化データに着目するとの方向性を決めた。次に、半永続データの寿命のモデル化について固定型寿命と適応型寿命を考察した。最後に、適応型寿命を維持するために、データの利用パターン (アクセス頻度、最近のアクセス時間) に関する情報を効率よく管理するデータ構造 TBF (Time-Decaying Bloom Filters) を考案した。TBF はメンバーシップを管理するデータ構造 Bloom Filters の拡張として、複数のカウンタを用いている。カウンタ値は増える一方ではなく、与えられた Decaying 関数に従い定期的に減量する。学部ウェブサイトの利用状況を解析する実

験を行い、TBF の有効性を評価した。

18 年度は、次世代 Web を代表とする Web 2.0 における半永続データ管理方式の研究及び並列計算モデル BSR における半永続データを効率よく扱う原理の解明をした。まず、半永続データの応用として、次の特徴を有するいわゆるソーシャル Web アプリケーションを考えた、1. 複数のユーザが同時にサーバと接続し、共通のスペースにおいてインタラクションを行い、データの変更を行う。2. ユーザはほかのユーザの動きを意識しながら行動する。3. サーバ側は全体の状況を把握できる。各々のユーザの状況は把握できない。ソーシャル Web アプリケーションでは 1. サーバ側でデータの更新状況を管理する。そのために空間的に効率よい TBF を採用する。2. サーバ側の持つデータの更新状況をクライアント側に伝える。これに基づいて、クライアント側でよりスマートにサーバとの同期頻度を調整できる。また、ハードウェア技術の進歩に伴うマルチプロセッサの普及のなか、並列計算が重要になってきた。並列計算モデル BSR における半永続データの効率よい処理についての研究を行った。

2 年間発表論文 8 通り: 論文誌論文 2 通 [1][6], 査読付き国際会議論文 6 通である。詳しくは参考文献をご覧ください。

## ◇ 参考文献 ◇

- [1] K. Cheng, M. Iwaihara, L. Xiang and K. Ushijima. Efficient Web Profiling by Time-Decaying Bloom Filters, DBSJ Letters, vol.4. no.1 June, 2005. pp.137-140
- [2] K. Cheng, L. Xiang, M. Iwaihara, H. Xu, and M. Mohania. Time-Decaying Bloom Filters for Data Streams with Skewed Distributions, RIDE-SDMA 2005, in conjunction with ICDE 2005, IEEE Computer Society Press, Tokyo. Apr. 2005. pp.63-69
- [3] K. Cheng, L. Xiang, T. Hirota, and K. Ushijima. A Web-based Classroom Environment for Enhanced Residential College Education, ICWL 2005, Hong Kong, August, 2005. (LNCS 3583) Springer. pp.56-65
- [4] K. Cheng and H. Wang. A Web Warehouse System for Semi-Automatically Gathering and Managing Online News, The Third International Conference on Active Media Technology (AMT 2005), 2005. pp. 343-344
- [5] L. Xiang, K. Cheng, and K. Ushijima. Encoding a Binary Search Tree in Constant Time on BSR, Proc. PDCAT 2005, IEEE Computer Society Press, Dalian, China, Dec.

\*1 萌芽研究とは「研究者が一人または、同一の研究機関に所属する複数の研究者が共同して行う研究であって、独創的な発想、特に意外性のある着想に基づく芽生え期の研究、例えば新しい研究分野の発展につながるような成果が生まれること、またはその契機となることが期待されるものを対象とする」研究種目

- 5-8, 2005, pp.740-744
- [6] L. Xiang, K. Ushijima, K. Cheng, et al. Constant Time Algorithm to Generate Plus and Minus Symbol Sequences, INFORMATION, vol.9, no.2, 2006, pp.331-337.
- [7] L. Xiang, K. Cheng, and K. Ushijima, Sequence analysis in constant time on BSR, PDCAT '06, Taipei, December 4-7, 2006, pp.431-436
- [8] L. Xiang, K. Cheng, and K. Ushijima, On Gray Codes for Set Partitions, Proc. Information-MFCSIT'06, Cork, Ireland, August 1-5, 2006, pp.180-183
- [9] 成, 横田, 牛島, クライアント・サーバー協調による効率的なアクティブ Ajax の実現, 第7回 Web インテリジェンスとインタラクション研究会, 平成 18 年 11 月, 東京: