

Linked Data Structures のための記憶構成—MOLDS—

5 K - 1

前川博俊 實藤隆則

ソニー (株) 情報通信研究所

1. はじめに

ポインタでリンクされた構造を持つデータ (Linked Data Structures, 図2, 以下リストデータ) は, 記号処理の分野などで重要で不可欠である。リストデータは, ポインタの付け換えによる構造の柔軟性を持ち, データ構造の変化しやすい問題の記述や処理に向いている。しかし, リストデータをページング方式のような仮想記憶方式 [9] によって管理したとき (図1), そのデータ構造の柔軟性ゆえ, 仮想化の処理と仮想空間上でのガーベージコレクション (GC) において, 主記憶・二次記憶間のデータ転送が多くなり, 全体の処理効率の低下する傾向がある。記憶空間におけるデータセル間の参照の局所性を大きくすれば, 無駄なデータ転送は少なくなる。そのための種々の手法が報告されているが, リストデータの性質を活かした本質的な解決策にはなっていない [4]。

この参照の局所性の問題は, 構造的に柔軟なデータを一次元的にアドレス付けされた空間で管理することによって生じる。我々は, リストデータをその構造に基づいて記憶空間上に表現し効率の良い記憶管理を実現する記憶構成方式 MOLDS (Memory Organization for Linked Data Structures) を提案している [1-4]。

2. Linked Data Structures のための記憶構成

MOLDSでは, 現在処理中のデータおよびそれに論理的に近いデータを主記憶上に表現し, 残りのデータを二次記憶上で管理する (図2)。二次記憶上では, データを論理的に関係の大きい集まり (二次記憶ストラクチャと呼ぶ) にわ

け, 主記憶・二次記憶間のデータ転送はこのストラクチャを単位に行なう (図3)。主記憶から二次記憶へのデータ転送をストラクチャ・アウト, 逆をストラクチャ・インと呼ぶ。ストラクチャ・アウトするデータは, 現在の処理から制御上時間的に遠いところでのアクセスや制御の環境などを二次記憶ストラクチャ間で参照が少なくなるように抽出して得ることができる (例: [5, 6])。

MOLDSでは, 二次記憶を参照の局所性の大きいストラクチャを単位に管理する。GCを含めたリスト処理において, 不要なデータ転送が減少し, ページング方式のような従来の記憶システムに比べて二次記憶アクセスの減少することが期待できる。ポインタは, 主記憶上では主記憶空間, 二次記憶上では二次記憶ストラクチャ空間において表現する。ポインタのための記憶消費量は従来の方式に比べて少なく済み, 主記憶と二次記憶の空間はそれぞれ独立に拡張できる。

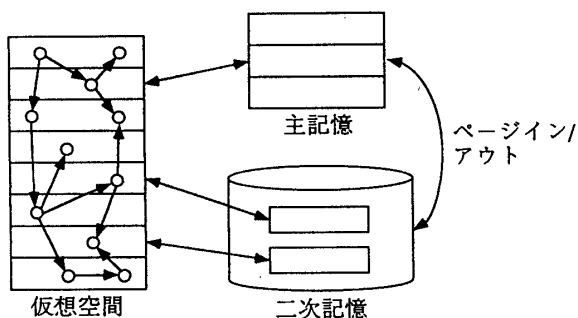


図1. ページングによる記憶管理

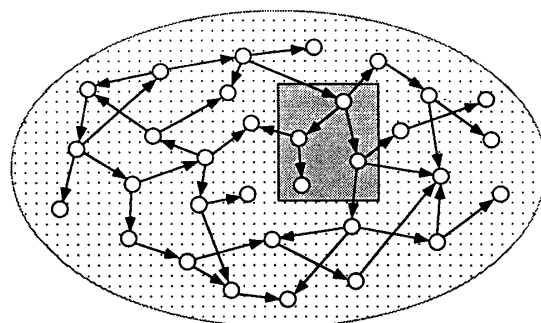


図2. Linked Data Structures の空間

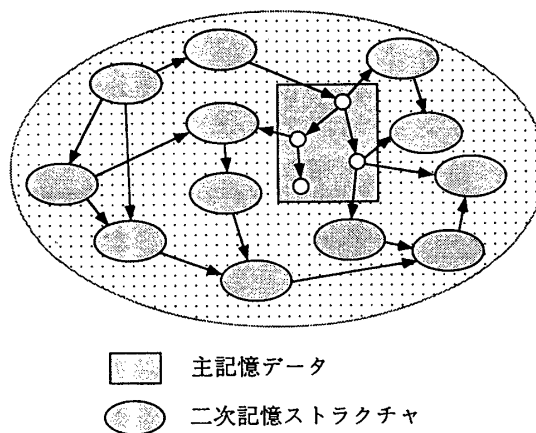


図3. Linked Data Structures のための記憶構成

MOLDS: Memory Organization for Linked Data Structures
 Hirotoishi MAEGAWA and Takanori SANETO
 Telecommunication & Information Systems Research Laboratory
 Sony Corporation

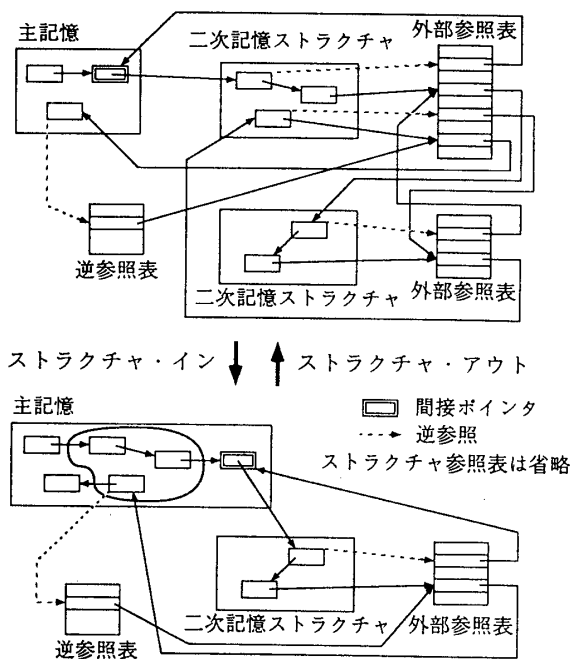


図4. MOLDSにおける記憶管理

MOLDSでは一方、次のようなオーバーヘッドを持つ。我々は、逆ポインタのないデータ表現を考えているため、ストラクチャ・アウトするデータに主記憶上の他のデータから参照のあったとき、それを「間接ポインタ」として残さなければならない。そのため主記憶上のデータセルに、被参照の認識のための多重参照マークまたは参照カウンタが必要である。主記憶と二次記憶でポインタの表現が異なるため、ストラクチャ・イン/アウトの際その変換が必要である。しかし、これらのオーバーヘッドがあっても、二次記憶アクセスが減少すれば、システム全体の処理効率を上げることができる。

3. MOLDSにおける記憶管理

主記憶上で、二次記憶のデータは「間接ポインタ」によって参照する。二次記憶ストラクチャ間および二次記憶から主記憶への参照・被参照は、ストラクチャ・イン/アウトの際の二次記憶アクセスを少なくするため次の参照表で管理する。

- ・外部参照表: ストラクチャ毎に持ち、二次記憶ストラクチャデータの外部に対する参照・被参照を管理する。
- ・逆参照表: 系全体でひとつ持ち、主記憶データの二次記憶に対する被参照を管理する。
- ・ストラクチャ参照表: 系全体でひとつ持ち、外部参照表を管理する。外部参照表のコンパクション、二次記憶ストラクチャの参照カウントに使用する。

ストラクチャ・イン/アウトを実行するとき、これらの参照表、間接ポインタを用いて、主記憶・二次記憶間、二次記憶ストラクチャ間の参照の変更を処理する(図4)。ポインタ表現の変換は、二次記憶制御部などで効率良く実行で

きる。ストラクチャ・アウトは、GCによっても十分な量のフリーセルが主記憶上に得られないとき起動する。

GCは、主記憶上ではデータセル単位に、二次記憶上では参照表を使ってストラクチャ単位に処理する。二次記憶の実際のデータセルをアクセスする必要がなく、高速処理が可能である。また、主記憶セルの被参照を管理しているので、インクリメンタルなGCを余分なオーバーヘッドなしで実現できる。MOLDSでは、GCは簡単なアルゴリズムで動作可能である。

4. まとめ

主記憶と二次記憶からなる階層的記憶構成において、リストデータとその構造に基づいて記憶空間上に表現し効率の良い記憶管理を実現する記憶構成方式 MOLDSを報告した。我々の行なったシミュレーションでは、MOLDSにおけるリスト処理はページング方式に比べて、基本性能において10倍から40倍高速であるという結果を得た[5, 6]。また、記憶空間の使用量も減少させることができた。

今後さらに詳細な検証と実機[7, 8]上での実験を行う予定である。

謝辞

本研究の遂行にあたって、福田譲治部長をはじめとする研究部のメンバーから有益な助言を得た。本研究の機会を与えて頂いた総合研究所宮岡所長、情報通信研究所松田所長に感謝する。

参考文献

- [1]前川博俊 他, Pointer-Linked Data における仮想記憶管理の一手法, 情報処理学会研究報告 SYM50-1 (1989).
- [2]前川博俊 他, Linked Data Structures の記憶管理, 情報処理学会第39回全国大会講演論文集, 1Q-4 (1989).
- [3]前川博俊 他, Linked Data のその構造に基づく記憶空間の構成, 情報処理学会研究報告 ARC80-5 (1990).
- [4]Maegawa, H., Memory Organization and Management for Linked Data Structures, Proc. of the ACM 1991 Computer Science Conference (1991).
- [5]實藤隆則 他, Linked Data Structures のための記憶管理とその動特性, 情報処理学会研究報告 ARC85-10 (1990).
- [6]實藤隆則 他, 記憶構成方式MOLDSの動特性, 情報処理学会第42回全国大会講演論文集 (1991).
- [7]安田弘幸 他, 計算資源指向型並列分散処理システム-Lilac-, 情報処理学会研究報告 SYM55-2 (1990).
- [8]安田弘幸 他, 並列分散システムLilacの構成, 情報処理学会第41回全国大会講演論文集(6), 5P-4 (1990).
- [9]Denning, P. J., Virtual Memory, Computing Surveys 2, 3 (1970), 153-189.

