

Title	Using Database Technology to Improve Performance of the Web : Caching and Beyond(Abstract_要旨)
Author(s)	Cheng, Kai
Citation	Kyoto University (京都大学)
Issue Date	2002-03-25
URL	http://hdl.handle.net/2433/149745
Right	
Type	Thesis or Dissertation
Textversion	none

氏名	成 凱
学位(専攻分野)	博士(情報学)
学位記番号	情博第50号
学位授与の日付	平成14年3月25日
学位授与の要件	学位規則第4条第1項該当
研究科・専攻	情報学研究科社会情報学専攻
学位論文題目	Using Database Technology to Improve Performance of the Web: Caching and Beyond (データベース技術を用いたウェブパフォーマンス向上: キャッシングとその一般化)
論文調査委員	(主査) 教授 上林 彌彦 教授 湯浅 太一 教授 田中 克己

論 文 内 容 の 要 旨

本論文は、従来のキャッシュを一般化してウェブ効率の向上をはかる新しい手法について述べたものである。ウェブデータのハイパーテキスト性の利用、データベース技術に基づくコンテンツキャッシュ手法の改良、キャッシュコンテンツの非透過的利用の実現などの技術を用いたもので、6章からなっている。

第1章は序論で、本研究の目的と概要、従来技術との比較が述べられている。近年のウェブの急速な普及に伴いインターネット上の通信量が爆発的に増えつつため、ネットワークアクセスのボトルネックの解消が重要となっていることが、本研究の背景になっている。このための有効な対処法としてウェブキャッシュがよく知られているが、従来のキャッシュは物理的データ、正確な意味定義のある論理データにしか対応していなかったために、ウェブコンテンツのキャッシュとその効率的利用のためには新しい方式の開発が必要であることを指摘している。

第2章では本論文のための基本的概念として、ウェブのクライアント・サーバ構造、ハイパーテキストを基礎としたコンテンツデリバリー機構、キャッシュの役割と種別、キャッシュの構造及び評価指標についてまとめている。

第3章では、高度のキャッシュ手法のためのコンテンツ管理モデル及びアルゴリズムの設計について述べている。まずキャッシュ管理に必要なメタデータを列挙している。ウェブデータの異質性に対処するための分類規則及び階層的コンテンツ管理モデルを提案し、このモデルに基づくキャッシュ管理方式「仮想多階層(VML)コンテンツキャッシュ」を提案している。このキャッシュにおけるキャッシュ置換には「ボトムアップ」方式と「トップダウン」方式があり、これらの方式の併用により高度のキャッシュ手法を簡単に設計できる。このモデルの有効性を示すため、新しいキャッシュ手法LRU-SPを提案し、シミュレーション実験による検証でLRU-SPの総合効率(ヒット率とバイトヒット率)がこれまでのキャッシュ手法より優れていることを示した。

第4章では、ウェブのハイパーテキストの特徴及びセマンティック情報をキャッシュ管理に用いる方式について述べている。ウェブデータ利用の効率化のためには、人間の利用形態をキャッシュに把握させる必要がある。このため、ウェブデータのハイパーテキスト構造とキーワードベースのセマンティック情報を抽出し、階層的なコンテンツ管理を行う方式について述べている。この階層構造とVMLキャッシュ方法と統合し「コンテンツを意識する知的キャッシュ手法」を提案している。この方法を具体化したアルゴリズムとしてLRU-SP+を開発し、従来のキャッシュより効率及び効果を大きく改善できることを実験により検証した。ここで、コンテンツを意識するキャッシュ効果を評価するために、コンテンツの重要度を考慮した「有効ヒット率」を用いて評価している。

第5章では、従来の透過的・受動的に利用されるキャッシュコンテンツに対してはキャッシュ手法をいくら改良しても60%以上のデータはあまり利用されないことを実験的に示し、検索推薦など非透過的にキャッシュコンテンツを利用するウェブウェアハウスの概念について述べている。それによってこれまでキャッシュの限界以上の効率向上が得られることを示した。

第6章は結論で、本論文のまとめが示されている。

論文審査の結果の要旨

本論文は、ウェブデータのハイパーテキスト性の利用、データベース技術に基づくコンテンツキャッシュ手法の改良、キャッシュコンテンツの非透過的利用の実現などの技術を用いて、従来のキャッシュを一般化してウェブ効率の向上をはかる新しい手法について述べたものであり、得られた主な成果は以下のとおりである。

1. 従来のキャッシュ手法は非構造的データ（物理的データブロック）あるいは構造的データ（関係データなど）にしか対応していなかったため、ウェブのようなハイパーテキストや半構造データを効率よく扱うのが困難であった。このため、ウェブに適合したキャッシュ手法向けの「仮想多階層（VML）コンテンツキャッシュ」という新しいキャッシュ構造を提案し、多様な特性を取り込むことができ、効率のよい知的キャッシュを開発できるようにした。

2. 仮想多階層コンテンツキャッシュに基づいて、ウェブデータの利用頻度、最近利用性、サイズなど多様な特性を考慮した新しいキャッシュ手法LRU-SPを開発し、その手法の効率および効果をシミュレーションにより評価しその有効性を示した。

3. 半構造データの管理方法を導入することにより、従来のキャッシュでは取り扱うことができなかった文書の内容などの意味的特性を取り込んだ「コンテンツを意識する知的キャッシュ手法」を提案した。その方法を具体化したアルゴリズムLRU-SP+を開発し、従来のキャッシュに比べ効率及び効果を大きく改善できることを実験により検証した。

4. キャッシュコンテンツを透過的・受動的に利用した場合は、キャッシュ手法を改良しても60%以上のデータは利用されないままで捨てられてしまうことを実データから示した。透過的・受動的ではない利用をするためウェブウェアハウスの概念を提案し、検索・推薦などで非透過的にキャッシュコンテンツを利用する方法を開発した。それによりこれまでキャッシュの限界以上の効率を得られることを示した。

以上、本論文は、これまで対応できなかったウェブのような半構造データを効率よくキャッシュし利用するためのキャッシュ構造、コンテンツを意識する知的キャッシュ手法、キャッシュコンテンツの非透過的利用などの導入により、ウェブの利用効率の大きく改善できることを示した独創的な研究で、学術的にも実用性の上でも極めて有意義であるといえる。

よって、本論文は博士（情報学）の学位論文として価値あるものと認める。

また、平成14年2月5日に実施した論文内容とそれに関連した試問の結果合格と認めた。