

デジタル放送のためのインデックス情報の断片化方式に関する検討

大和田俊和* 浅田一繁* 古瀬一隆†

†株式会社次世代情報放送システム研究所

*株式会社リコー

†茨城大学工学部情報工学科

要旨

近年、衛星を利用したデジタル放送の実用化により、放送の多チャンネル化が進み、膨大で多種多様なコンテンツが配信されるようになった。また、コンテンツを構造化したり、通信ネットワークを併用したりすることで、より高度な放送サービスが実現されようとしている。こうした状況に伴って、不特定多数の利用者が目的のコンテンツを効率よく検索する技術が求められている。本稿では、断片化されたインデックス情報を放送し、受信機でそれを再構築することにより、放送波という公共性の高い伝送路を有効利用しながら、部分的に受信されたインデックス情報を使って検索を行う方式について述べる。

Decomposing Index Information for Digital Broadcasting

Toshikazu Ohwada* Kazushige Asada* Kazutaka Furuse†

†Information Broadcasting Laboratories, Inc.

*RICOH, Co., Ltd.

†Department of Computer and Information Sciences, Ibaraki University

Abstract

Recent launching of digital satellite broadcasting services has significantly widened a selection of broadcasting programming both in quality and quantity. Still more advanced services using structured contents and communication networks are about to be started. These changes have given rise to a need to provide ordinary TV viewers with an efficient way to search and retrieve the broadcast contents of their interest. In this paper, we propose a contents retrieval method using decomposed index information, in which index information is broadcasted to TV receivers and re-constructed as received at the receivers. This method takes advantage of broadcast waves as a means of mass distribution.

1. はじめに

近年、我が国でも衛星を利用したデジタル放送の実用化が始まり、それに伴い様々なサービスが開始されている。さらに近い将来には、地上波デジタル放送も開始されようとしている。

デジタル放送の特色の一つとして、チャンネル数の多さ

があげられる。また、インターネットによる情報サービスの進展もあって、放送波と通信ネットワークを通じて、膨大で多種多様なコンテンツが配信されるようになった。

こうした状況においては、利用者にとって、質のよい有効なコンテンツを効率よく検索するためのインフラが必要となる。このような検索機能を実現する上で、利用者が目的のコンテンツにたどり着くまでのインデックス情報を適切に

† 株式会社リコーより株式会社次世代情報放送システム研究所へ兼任出向中。

The authors are partly on loan from Ricoh Company, Ltd. to Information Broadcasting Laboratories, Inc.

提供することが重要である。インデックス情報の例としては、コンテンツの書誌情報や検索の高速化のために構築される B-tree のようなインデックス構造に関する情報が挙げられる。

しかしながら、コンテンツの構造化や著作権保護、課金制御、利用者の嗜好を踏まえた高度なサービスを実現するには、インデックス情報を豊富に提供することが求められる。そのため、インデックス情報を配信すること自体が、放送波または通信ネットワークの伝送容量を圧迫しかねない。特に、放送波は公共性の高い伝送路なので、その影響は甚大である。一方、インデックス情報が不足してコンテンツが十分に絞り込めないと、無駄なコンテンツの取得要求と配信が多く発生する。映像や音声といった巨大なコンテンツを配信の対象とする場合には、この問題は致命的である。

こうした問題を解決するため、筆者らは、インデックス情報を断片化して放送する方式を考案した。この方式では、配信側で構築したインデックスを小さな単位で断片化し、それを放送するという方法を採用している。受信機側では、放送波より受け取った断片からインデックスを再構築し、それを用いて検索する。本方式では、受信機側ですべての断片が揃わなくてもインデックスを部分的に再構築することができるようになっている。このため、遅延のない(部分的な)検索が可能であるなどの利点を持つ。

また、インデックス情報を断片化して放送することにより、

断片ごとの優先度などを考慮したストリーム化や、各受信機での必要性に応じた断片のキャッシングなどの機構を構築することも可能になる。

以降では、本稿で想定するデジタル放送を使った放送サービスのシステム構成と利用形態、断片化の利点と要件、断片化方式について述べ、最後に、本稿での議論と今後の課題についてまとめる。

2. システム構成と利用形態

筆者らは、国民規模の利用者に放送波および通信ネットワークを使って、放送サービスを提供するシステムに関する研究を行っている[1]。

従来のインターネットでの検索エンジンに代表される情報検索サービスでは、利用者の検索要求に応じて、配信側で検索処理を行う。しかし、このような集中管理的な形態では、国民規模の利用者が検索要求を行った場合、通信ネットワーク及び配信側の計算機はその負荷に耐えられない。

そこで筆者らは、検索処理の分散化を図るために、インデックス情報を放送波で配信することで、受信機側で検索処理を可能とするシステム構成を考案した。筆者らが想定するシステムの具体的な構成要素と規模は、以下のとおりである(図 1)。

・ 放送局…全国各地に数 100 箇所に分散配置される。コ

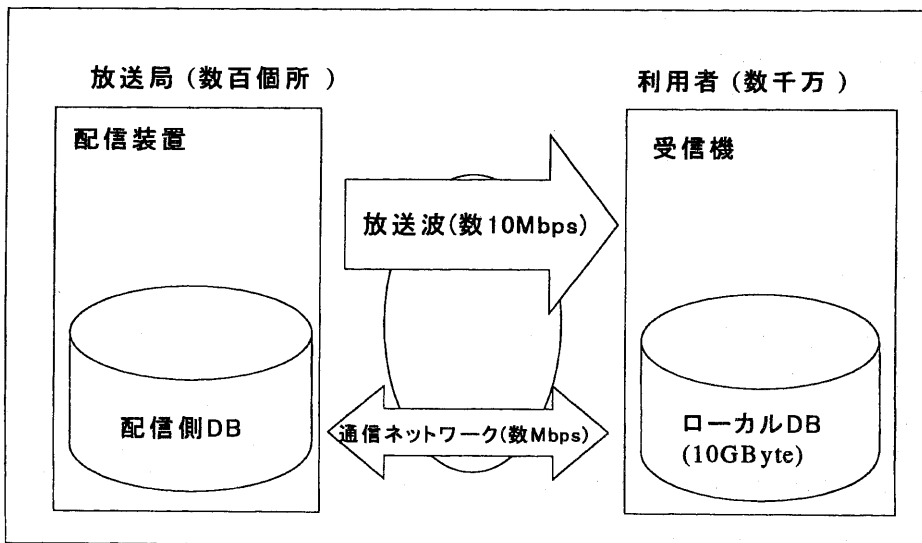


図1 システム構成

コンテンツとインデックス情報を収めた配信側 DB と、デジタル放送による配信機能と通信ネットワークアクセス機能を備えた配信装置を持つ。

- ・利用者…現行放送の視聴者と同程度の数千万規模を想定する。デジタル放送の受信機能と通信ネットワークへのアクセス機能、およびコンテンツとインデックス情報を受取る容量 10Gbyte 程度のローカル DB を備える受信機を持つ。
- ・放送波…数 10Mbps の伝送速度を持つ。
- ・通信ネットワーク…数 Mbps の伝送速度を持つ。

利用者は受信機を通じてコンテンツの利用と検索を行う。コンテンツの利用方法は以下のようなパターンが考えられる。

- ・デジタル放送で配信されてくるものを利用する。
- ・通信ネットワークを通じて取得したものを利用する。
- ・既に受信機に蓄積されているコンテンツを利用する。

実際にはこれら三つの方法すべてがコンテンツ利用方法として併用されることになる。

コンテンツの検索には、システム上に用意されたインデックス情報を使用する。インデックス情報は、その内容に応じて、

- ・インデックス構造に関する情報
 - ・コンテンツの内容に関する付加情報
- の2つに大別できる。前者は、B-tree やハッシュといった、検索の高速化を図るために配信側で構築された情報である。後者は、コンテンツの制作時に付与される以下のような情報である。
- ・所在情報
受信機でコンテンツを取得するときに指定される。放送される帯域および時間帯、または通信ネットワーク上のアドレス(例えば URL)など。
 - ・構造情報
構造化されたコンテンツの構成要素にアクセスするときに利用される。映像におけるシーンやカットの構成、文書の章立てなど。
 - ・書誌情報
タイトル、作成者、作成日など
 - ・分類情報
ある特定のルールによる分類名、分類番号を表す。
 - ・要約情報
コンテンツの内容を簡略に表現する。ダイジェスト映像、要約文、紹介記事など。
 - ・アクセス制御情報

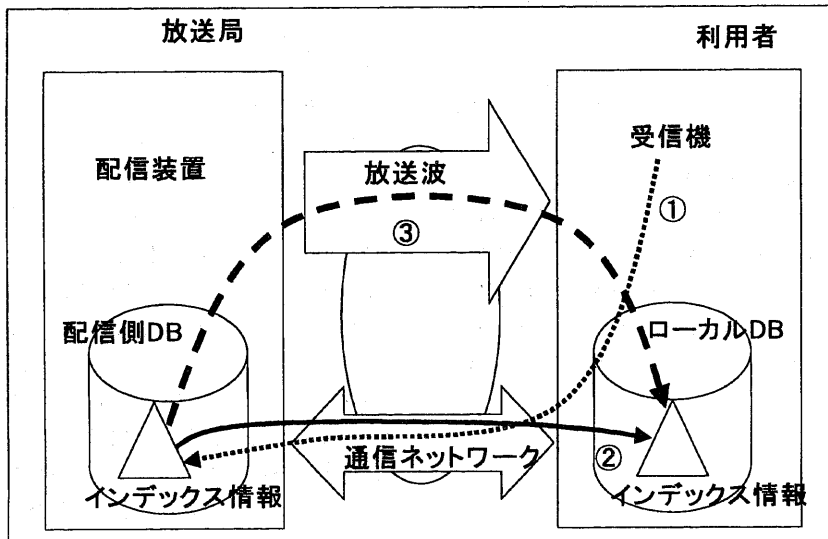


図2 インデックス情報の利用形態

コンテンツを取得するときに検査されるべき利用者に対する要件を表す。年齢制限、利用可能範囲、パスワード設定、課金設定など。

・ 関連情報

関連するコンテンツまたはそのインデックス情報の所在を表す。

利用者が行うインデックス情報の利用形態としては、以下のようなパターンが考えられる(図2)。

- ① 通信ネットワーク上に置かれた検索エンジンのインデックス情報を使用する。
- ② 通信ネットワークにより受信機側に配信されたインデックス情報を使用する。
- ③ デジタル放送波により受信機側に配信されたインデックス情報を使用する。

これら3種のうち、インデックス情報の利用形態としてもっとも優れているのは、③の形態である。その理由は、放送および通信の特性にある。放送は、一方通行ではあるが、同報的配信には非常に優れている。一方の通信は、双方向性は持っているものの、アクセスの過度の集中には弱い。

第2節で述べたように、システムとして数千万人規模のユーザ数にも対応しなければいけないことを鑑みれば、①や②の方式は採用できない。また、インデックス情報は、個々のコンテンツに比べてはるかに公共性の高い情報であるので、放送によって同報的に配信するのは有効である。

以上により、筆者らは③の方式を主要方式として選択する。

なお、インデックス情報を放送によって配送する方式の研究については、[2]がある。しかし、これは携帯端末の電力消費を軽減するための配送方式についての研究であるので、これを本稿のシステム形態に適用するのは意味がない。本稿のようなシステム形態でのインデックス情報の配送方式は、筆者らの知る限り、これまでに研究の例がない。

3. 断片化の利点と要件

インデックス情報のような種類の情報を放送によって配送する場合、同じ情報を一定の周期で放送し続けるデー

タカルーセル方式が採用されるのが一般的である(デジタル衛星放送におけるEPGなど)。しかし、大量のインデックス情報を配送しなければならない本稿が想定する規模のシステムでは、以下に述べるようないくつかの問題が発生する。

まず第一に、すべてのインデックス情報を放送するのにかかる時間(カルーセルの周期)が長くなってしまいう問題がある。デジタル放送波で使用できる伝送速度には限りがあるため、インデックス情報の量が多いからといって簡単に伝送速度を上げることはできない。周期が長くなれば、利用者がインデックス情報にアクセスして必要な情報を得るまでにかかる時間も長くなるため、実用上の大きな支障となる。また、周期が長いと、何らかのトラブルで配送中の情報が一部欠けてしまったり、周期の途中で利用者が検索を開始したときなどの、その次の周期の放送が始まるまでの待ち時間も長大になる。

第二の問題は、インデックス情報全体を受信機側に保持できないということである。受信機側に設けられる記憶装置の容量には限りがあるため、1周期で送られてくるインデックス情報をすべて保持して検索を行うという方法が採用できない。受信機側に十分な記憶容量を用意してすべてのインデックス情報を保持させるという方法も考えられるが、個々の利用者にとってすべてのインデックス情報が必要になるとは限らないことを考えると、記憶容量のオーバーヘッドが大きすぎる。

以上のような問題を解決するため、本研究では、インデックス情報を断片化して放送する方式を考案した。この方式では、配信側で構築したインデックスを後述するような方法で断片化し、それを放送波によって放送する。受信機側では、放送された断片を受信してインデックス情報を再構築し、それを用いて検索を行う。

ここで重要なのは、受信した断片だけを用いて、部分的なインデックス情報を受信機側で構築できるようにすることである。このようにすることにより、受信機側ではすべてのインデックス情報を受信し終わるのを待たずとも検索を始めることができる。また、利用者の好みに合わせてそれぞれの受信機で必要性の高い断片だけを保持することにより、少ない記憶容量でも高品質の検索を実現できるようになる。

配信側でも、断片化によって配信する大きさ、順序を状況に応じて決定できるようになる。これにより、ある特定の

部分のみ繰り返し配信の周期を短くしたり、欠損部分や更新部分のみを配信するといったことが可能になる。また、帯域が狭い場合には断片を小さく、広い場合には断片を大きく取る、といったような柔軟な操作も可能となる、という利点もある。

ところで、これまでに説明したような方式によって配信側で断片化してインデックス情報を受信機側で再構築するためには、個々の断片が以下のような性質を有している必要がある。

- その断片が必要かどうかの判断を受信機が行えるだけの情報を持っていること。
- 再構成のための情報を持っていること。
- 欠損部分や、未受信部分に対して、既受信部分からの推定が可能かつ信頼性を計算できる情報を持っていること。
- 陳腐化が検出できる情報を持っていること。

上記の性質を、本稿では断片の自立性と呼ぶこととする。ただし、上記の性質すべてを満たすような完全な自立性を断片が持つべきかは、後述する断片の分類と断片化方式に依存し、場合によっては部分的な自立性を持つだけでよいこともある。

自立性を持つように断片化を行う場合、断片は以下の種類に分類できる。

A) コンテンツ構造に対応した断片

コンテンツは従来のような単純なデータの連なりではなく、その形態に応じた複雑な構造を持つようになると考えられる。このような情報に対応した断片化を行うことにより、より細かな検索が行えるようになる。

B) 概念構造に対応した断片

コンテンツを分類する上で、上位、下位などの概念構造による分類が行われる場合がある。この概念レベルに対応した断片化を行うことにより、検索時のブラウジングなどをスムーズに行えるようになる。

C) インデックス構造に対応した断片

インデックス構造によって、断片化を行いやすい単位は違ってくる。例えば、ハッシュ表であればテーブル単位、B-tree であればノード単位といった断片化が考えられる。

D) 自立可能な単位の断片

構造を問わず、自立性を持った最小単位に分割する。

本研究では、インデックス情報の構造としてはもっとも一般的である木構造インデックスを対象とし、上記 C) の構造による断片化方式を検討した。ただし、現段階ではインデックス情報の更新やインデックス構造の変化については考慮していない。

4. 断片化方式

インデックス構造に対応した断片化の方式には、配信側での断片化から受信機側での再構築までに4つの段階がある。配信側ではまず断片化が行われる。次にストリーム化が行われ、断片が配信される。受信機側では、受信した断片に対してキャッシングを行う。そして、残った断片を使用して再構築が行われる。以下ではこの作業の流れを順に検討していく。

4-1 配信側におけるインデックス情報の断片化

断片化の単位は、木構造のインデックスを構成する各ノードとする(図 3)。これは後述するように、断片化の単位がそのまま受信機側におけるキャッシングの単位になるからである。また、受信機側で再構築に必要な情報を確保しつつ、再構築をできるだけ単純に行うためにも、ノード単位の断片化が有効である。

各断片に埋め込む情報は、再構築およびキャッシングのため、以下のものが必要である。

・ノード ID

各ノードを一意に識別するための ID。受信機でインデックス構造を再構築する際に、ノードとノードを接続するために使用する。

・位置情報

ノードの、ルートノードからの位置を表す情報。再構築の際に絶対的なノードの位置を決めるために使用する。

・タイムスタンプ

このノードを配信側で生成した時刻。同じ ID を持つノードの置き換えに使用する。

・有効期限

このノードが受信機側において利用できる期限。検索時に検索の有効性を判断するためや、新しいノードの受信が必要かどうかを判断するために使用する。

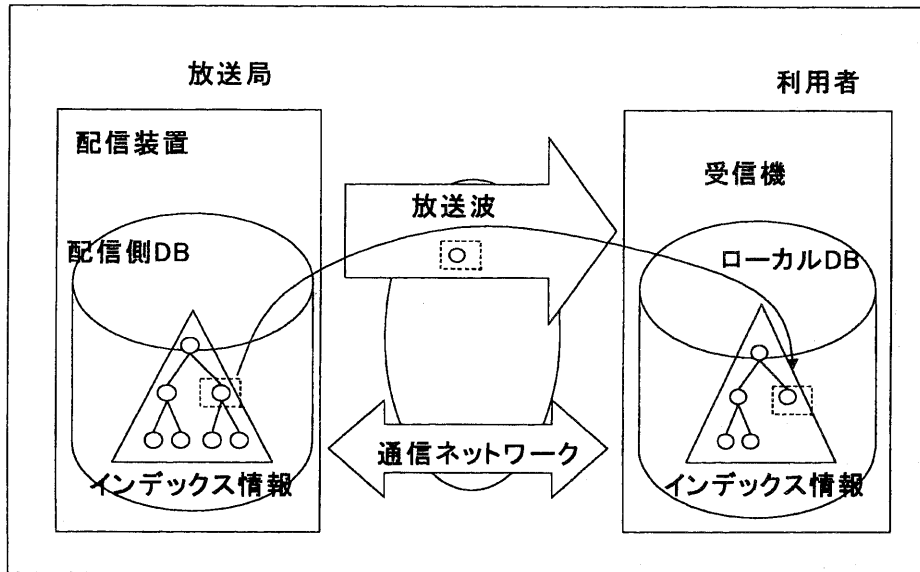


図3 ノード単位の断片化

・次回放送予定時刻

このノードの情報を次に送る予定の時刻。キャッシングの際に、蓄積すべきかどうかを判断するために利用する。

・構造情報

他ノードとの構造上の関連情報。ただし、子ノードへのリンクは、それぞれのノードのノードIDによって表現する。また、リンクしている子ノードの次回放送予定時刻も埋め込む。

4-2 インデックス情報のストリーム化及び配信

断片化したインデックス情報は、放送するにあたってストリーム化する必要がある。

木構造に基づくインデックス構造のストリーム化の方式は、最も単純なものとしては、各ノードを深さ優先か幅優先で繰り返し送る方式が考えられる。しかし、より効率のよい放送をするためには、それぞれのノードの放送頻度を非一様にするのが有効である。すなわち、より重要性の高いノードやより頻繁に更新されるノードを、より高い頻度で放送する(図4)。

このような放送頻度の非一様化により、未受信あるいは欠損ノードを原因とする検索不可部分をできる限り減少さ

せ、また、古い情報による不正確な検索を防ぐことが可能となる。

4-3 受信したインデックス情報のキャッシング

放送されるインデックス情報は全体では莫大な量となるため、受信機側ですべてを保存しておくことは困難である。そこで、何らかの方式によってノードごとの(その受信機における)必要度を定め、必要度の低いものを優先的に破棄するような機構が必要になる。

このような状況ではLRUが使用されることが多いが、本研究の場合には有効ではない。その理由は、先に述べたようにノードによって放送頻度が非一様であることにある。キャッシュのミスヒットが起こったときにそのノードを取得するまでにかかる時間はノードによって異なるので、単純に受信機側でヒット率の高いノードを優先的にキャッシュすればよいということにならない。すなわち、各ノードの必要度を以下のパラメータによって決定する。

- ・その受信機における、ノードのアクセス頻度
- ・そのノードの放送頻度

例えば、頻繁に放送されるノードはすぐに次の放送機会がおとずれるので、たとえ受信機で頻繁にアクセスされ

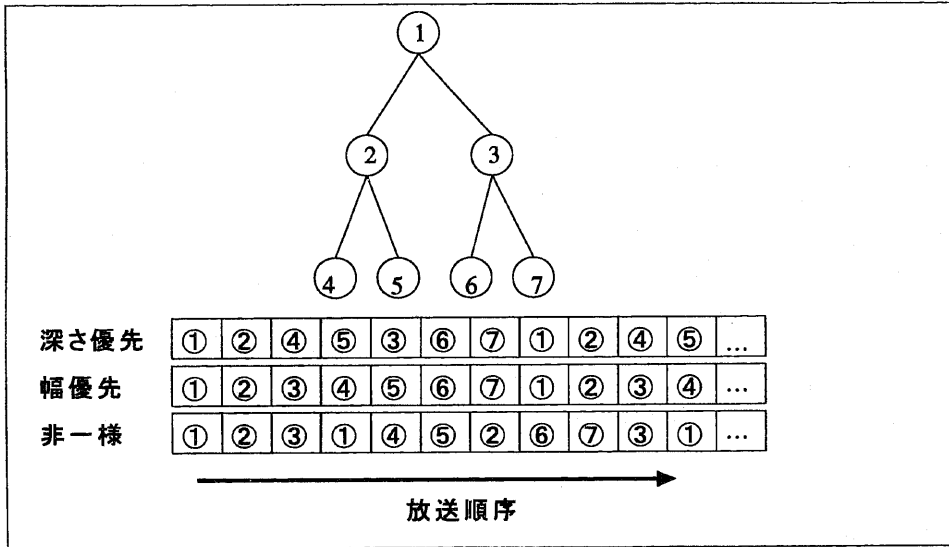


図4 ノードの放送方式の例

たとしても)受信機側でキャッシュしなければならない必要度は低い。また、同じアクセス頻度のノードであれば、放送頻度の高い方を破棄した方がよい。

ただし、受信機側のアクセス頻度とノードの放送頻度をどのように勘案してそのノードの最終的な必要度を定めればよいかは、難しい問題である。例えば、極めて頻繁に放送されるノードは破棄して構わないかもしれないが、そうでない場合には、放送頻度に関係なく、受信機側のアクセス頻度のみによって必要度を決定する方が効率が高いという可能性がある[3]。また、アクセス頻度と放送頻度は、タイムスケールの取り方によって著しく変化する可能性がある。そのため、利用者の興味の変化や、配信側の放送スケジュールを考慮する必要がある。

ノードのアクセス頻度と放送頻度のほかにも、必要度を判断するためのパラメータとして、ノードの有効期限が利用できる。

アクセス頻度と放送頻度が等しく、アクセス頻度が高い場合には、有効期限が長いノードを破棄する方が有効である。これは、有効期限が長いノードは、有効期限が短いノードよりも、再び取得できる機会の多いことが期待されるからである。逆に有効期限が短いノードを破棄してしまうと、そのノードを取得するまでの間に有効期限が切れ、検索洩れが生じる恐れがある。このような方針を取ることで、より

幅広いコンテンツに利用者がアクセスする機会を与えることになる。一方、アクセス頻度が低い場合には、次回アクセスする前に有効期限が切れる可能性があるので、有効期限が長いノードをキャッシュするのが得策である。

このように、パラメータの値とその組み合わせによって、有効なキャッシング方法が大きく違ってくる。本システムに最適な方法を選定するためには、今後検討を続けていく必要がある。

4-4 受信側におけるインデックス情報の再構築

受信機側では、受信したノードによってインデックス構造を再構築する。再構築にあたっては、ノードIDと位置情報を用いる。

ノードはストリーム化されて順次送られてくるので、同じIDを持つノードが後に送られて来たら、古いノード情報は捨て、新しいものと置き換える。ただし、インデックス構造そのものは変化しないことを前提としているので、この更新でノードIDや位置情報、構造情報が変化することはない。

再構築されるインデックス構造では、次回放送予定時刻なども保存しておく。これらの情報はキャッシング処理に利用する。

利用者が検索を行う際、ノードの一部が欠損していたり、有効期限を過ぎていたりといった原因により、正しく検索で

きない可能性がある。このようなときにどういった方法で解決するかは、求められる検索の質によって違ってくる。例えば、大雑把な検索結果が得られればよい場合には、その時点で受信機に再構築されているインデックス構造のみを用いて検索を行う。より正確な検索結果が必要となる場合には、何らかの方法で不足しているノードを入手する必要がある。入手方法としては、不足しているノードの次回放送予定時刻まで待つという手段が基本であるが、場合によっては、通信ネットワークを用いて配信側 DB とアクセスし、不足しているノードだけを取り寄せるといった手段も考えられる。放送頻度が低いノードを待っている場合には、後者の手段を取ることで待ち時間を減少できる。ただし、多数の利用者が同時に多量のノードを通信ネットワークから得ようとする、アクセスの集中による負荷増大をまねいて結局は放送を待った方が早いという事態も起こりうる。このため、通信ネットワークによる配信側 DB 中のインデックス情報へのアクセスは、不足ノード数がわずかな時のみ行うべきものと言える。

5.まとめ

本稿では、数千万規模の利用者に対してコンテンツの検索と送受信を行うシステムを対象とし、このシステムにおけるインデックス情報の送受信の方式としてデジタル放送を有効利用する機構を提示した。システムにおけるインデックス情報の量と、デジタル放送によるデータ配信の特性、さらには受信機側の記憶装置容量から発生する問題について考察し、これを解決する手段としてインデックス情報の断片化について述べた。そして、断片状態での利用という観点から、配信側と受信機側で行われる配信と再構築の流れに沿って、断片化の方式について検討を行った。

デジタル放送はこれからますます発展が見込める分野であり、当方式を適用することで高度な検索機能を持った放送サービスが実現されることが期待される。

筆者らも、引き続いて当方式のコストモデルを構築して解析を行い、当方式の有効性について検証を行っていく予定である。また、有効なキャッシング方式およびストリーム化方式についてさらに検討を行っていくとともに、今回想定しなかったインデックス構造が変化する場合の方式についても検討を進めていきたい。

参考文献

- [1] 飯沢篤志、浅田一繁、白田由香利:「情報放送のための超大規模分散データベースシステム」、情報処理学会研究報告、97-DBS-113-44、1997。
- [2] T.Imielinski, S.Viswanathan, B.R.Badriath. "Energy Efficient Indexing on Air", Proc. ACM SIGMOD '94.
- [3] Swarup Acharya, Rafael Alonso, Michael Franklin, Stanley Zdonik. "Broadcast Disks: Data Management for Asymmetric Communication Environments", Proc. ACM SIGMOD '95.