

情報放送のための 超大規模分散データベースシステム

飯沢 篤志*, 浅田 一繁*, 白田 由香利*
株式会社 次世代情報放送システム研究所
{izw, asada, shirota} @ ibl.co.jp

近年テレビ放送は現在のアナログ放送からデジタル放送に移行しようとしている。我々はデジタル放送の将来あるべき形態を研究しているが、その統合化されたデジタル放送を利用した情報配信基盤を情報放送と呼んでいる。本稿では、まず情報放送の機能モデル及び利用イメージを説明する。情報放送は広いエリアに渡り配信可能であり、膨大な量を扱うことが予想されるので、情報放送では超多地点大規模分散データベースシステムが必要になる。そのデータベースシステムにおいては、データ複製、ディレクトリサービス、分散最適配置、情報フィルタリング、索引づけ、データの自動分類、情報検索の Quality of Service(QoS) 制御などの機能が必要になる。本稿では、情報放送システム上のデータベースシステムに必要とされるこれらの要件について論じる。

Very Large Distributed Database Systems for Information Broadcasting

Atsushi Iizawa[†], Kazushige Asada[†], and Yukari Shirota[†]
Information Broadcasting Laboratories, Inc.
{izw, asada, shirota} @ ibl.co.jp

Recently, the existing analog television broadcasting has been changed to be digitalized broadcasting. We are researching the future digitalized broadcasted infrastructure which we call "information broadcasting." In the paper, first we explain a functional model of the information broadcasting and how an end user uses the information broadcasting system. The information broadcasting can send a lot of data to a very large area. Therefore, the system will have to build a very large distributed database system. The database system requires many functions: data replication, directory service, optimization of distributed site allocation, information filtering, indexing, automatic classification of data, and Quality of Service (QoS) control of information retrieval. The paper describes requirements of these database system functions for information broadcasting.

*株式会社リコーより兼任出向中。リコーでの所属は研究開発本部ソフトウェア研究所。

[†]The authors are partly on loan from Software Research Center, R & D Group, Ricoh Company, Ltd.

1 まえがき

近年、テレビ放送は現在のアナログ方式からデジタル方式に移行しようとしている。またインターネットを中心に発達してきた、WWW 及びデータベースシステムにおける通信形態として、放送型サービスが注目をあびてきている。従来の通信形態が、自分の必要な情報を取り出してくる PULL 型であるのに対し、放送型 (PUSH 型) では予め設定したジャンルの情報を周期的にまとめて配信する。インターネットと比較したときに、放送型通信形態は以下のような特徴をもつ [1]:

(a) 同報性をもつ、(b) ユーザは listen only である、(c) 問合せをすることは、放送された情報をフィルタリングすることにほかならない、(d) 周期的にデータを流すことができる、(e) ユーザ数が増加しても放送コストは一定である。よって頻繁にアクセスされる情報ほど、放送することで高い効果が得られる。

デジタルテレビ放送技術の発展により、情報システムの放送型配信のために、デジタルテレビ放送を活用しようという研究が今後盛んになると考えられる。デジタルテレビ放送の形態としては、次の 3 種類が考えられる [2]。

1. 衛星デジタル放送:

衛星による放送。日本では放送衛星 (BS) を利用した放送に限っていたが、1992 年からは通信衛星 (CS) を利用した衛星放送も開始した。CS は電波中継器 (トランスポンダ) 数が 20~30 個と多く、多くのチャンネルサービスが提供できるが、チャンネルあたりの送信出力が BS に比べて小さいので、受信には大型のアンテナが必要となる。現在、国内では PerfectTV! が CS によるデジタル放送を行っている。1998 年には、CS サービスとして、さらに DIRECTV JAPAN、JskyB が利用できる予定。次期放送衛星 BS-4 の後発機のデジタル化も決定している¹。

2. 地上デジタルテレビ放送 (DTTB: Digital Terrestrial Television Broadcasting):

地上放送によるデジタルテレビ放送。周波数帯域としては、アナログテレビ放送用チャンネルの空いているチャンネルを使用することが考えられる。郵政省は 2000 年以前に地上波テレビ放送をデジタル化する体制を整備すると表明している。

¹郵政省は 2000 年にサービスを開始する BS-4 後発機をデジタル化することを宣言した。

3. デジタル CATV:

有線テレビ放送 (Cable Television) によるデジタル放送。

デジタルテレビ放送の特長は、データを速く、安く、大量に送ることができることである。

- 高速性: NHK は BS-4 後発機で利用可能なデジタル伝送方式を提案したが、これによると、電波中継器 1 本当たりの信号周波数帯域を 33MHz としたときの情報速度は最大 53Mbps と算出されている [3]。
- 低コスト性: [4] によると、企業自ら放送チャンネルをもって情報を発信した場合、ホームページ 1 ページあたり 100K バイト換算で 10 ページを配信すると、コストは 5 円程度となる²。電話回線で 20Kbps、3 分 10 円とした場合、およそ 20 円かかるのに比較して、これは非常に安いと言える。
- 大量配信: 放送型の同報性は大量配信を可能とする。通信衛星 (CS) では、衛星の中継器 1 本で、2Mbps のチャンネルを 12 チャンネルまで設けることが可能である。CS を使ったデジタル衛星放送だけで、98 年にはテレビチャンネル数は 350 を超えると予想される [5]。こうしたバンド幅が放送型で提供される場合、多数の受信者に配信可能となる。

我々は、デジタルテレビ放送の将来あるべき形態を研究することを目的としている⁴。その統合化されたデジタルテレビ放送を **情報放送**と呼ぶことにする。

次節では、**情報放送**のイメージを説明する。**情報放送**システムにおいてはデータベースシステムが重要なファクタとなるが、そのデータベースシステムの満たすべき要件は何か、3 節で論じる。

2 情報放送とは何か

情報放送とは、デジタル放送を利用した情報配信のインフラである。

2.1 機能モデル

情報放送の機能モデルは図 1 のように表すことができる。

²ここでは 5Mbps のチャンネルの利用料金を年間 1 億円で単純計算した場合を仮定している。

³2010 年を想定している。

⁴(株) 次世代情報放送システム研究所は、1997 年 2 月、基盤技術研究促進センター、ソニー、松下電器産業、リコー、日本放送協会 (NHK)、日本テレビ放送網、フジテレビジョンが出資して設立した。

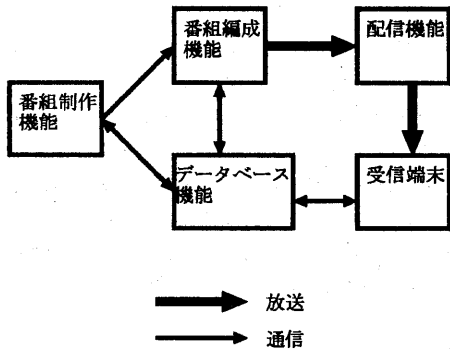


図 1: 情報放送機能モデル

1. 番組制作機能: 番組の制作を行う。コンテンツ⁵の制作も含む。
2. 番組編成機能: 番組の編成、調達及び利用者への共有を行う。
3. 配信機能: 放送伝送路を保有し、伝送施設の整備、運用管理を行う。
4. 受信端末: 放送伝送路を介して番組を受信する。また、必要に応じて通信伝送路を介してコンテンツを検索する。受信端末の機能としては、選択受信、映像再生、検索インタフェース、検索結果の視覚化、コンディショナルアクセス、著作権保護などがある。
5. データベース機能: コンテンツ及びデータベースを運用管理する。また、そのデータベースを受信端末から検索するためのサービスを提供する。

想定するシステム規模を表 1 に示す。

2.2 利用イメージ

以下に 情報放送の利用イメージを示す。

<<情報パッケージ>>

番組の一形態として、情報パッケージを放送する。利用者は情報パッケージを手がかりに、必要なコンテンツを受信または検索する。情報パッケージは、雑誌や CD-ROM のようなまとまりで、100M バイト程度の容量を想定する。情報パッケージの内容として、主として、以下のものが考えられる。

⁵我々は番組の素材をコンテンツと呼び、視聴する利用者、放送する時間帯などを考慮して、コンテンツの集まりを意図をもって編成したものを番組と呼ぶ。

表 1: 想定するシステム規模

項目	数量
番組編成者数	数 100
伝送 (配信) 事業者数	数 100
受信端末数	数 1000 万
データベース事業者数	数 1000
データベース総量	数 P(ペタ) バイト
受信端末蓄積量	10G バイト
放送路伝送帯域	100Mbps
通信路伝送帯域 (事業者～受信端末間)	数 Mbps
通信路伝送帯域 (事業者～事業者間)	数 Gbps

- 放送番組表 (EPG): 現行放送⁶の番組表に、情報放送の放送スケジュールを加えたもの。
- 現行放送番組の付加情報: 映像コンテンツに関するより詳細な情報を視聴者に与える。
- 選択性のある情報番組: 利用者が選択的かつ対話的にほしい情報を得ることができる番組。全国の気象情報及び、タウン情報誌的なイベント情報など。
- インデックス情報: 情報 (コンテンツ) の所在と放送時刻に関する索引。

<<選択蓄積>>

情報放送では、次のような選択蓄積機能をもつ受信端末を想定している。

- 放送伝送路を使って配信される情報の中から利用者がフィルタリングを行い、必要部分だけを選択し、受信端末データベースに蓄積する。
- 有料の情報 (コンテンツ) であっても、蓄積しただけでは課金されず、利用者が視聴したときに初めて課金される。

<<検索操作>>

情報放送における検索は、広範な利用者による視聴を想定するため、一般的なデータベース検索よりも簡易性が求められるであろう。計算機に向かい検索を積極的に行う利用者とは異なり、より受動的な態度で、情報放送受信端末が提示するコンテンツを選択あるいは破棄する程度の操作ですむようにすべきであると考えられる。

⁶我々は現行の放送を情報放送と区別して用いている。

情報放送の検索操作に対する要件としては以下があげられる。

- 従来のテレビリモコン操作程度の簡易性をもつこと
- 検索操作を可能な限り、利用者に意識させないこと
- 曖昧な条件を元に検索できること
- 記憶を辿るような絞り込み方(曖昧条件の AND による絞り込み)ができること。
- 表示される検索結果は、受動的な利用者にも受け入れられるように、音と映像を駆使した編集者の意図のもとにまとまりをもつこと。
- 受信端末は利用者が明示的に指示した検索条件の他、利用者の操作履歴、登録情報などを活用して、付加的な検索条件を自動作成できること。

<<検索結果の入手方法>>

検索結果の入手方法としては以下がある。

1. 受信端末に蓄積されたデータの中から得る。
2. 通信伝送路を介して、データベース事業者のデータベースから入手する。
検索結果が小容量の場合、あるいは緊急に必要であり、多少費用がかかってもよい、という場合である。
3. 放送スケジュールを調べて、放送されるまで待ち、放送伝送路から入手する。
4. リクエストを出して、放送スケジュールに載るのを待つ。

3 データベースシステムの要件

本節では、情報放送システムにおけるデータベースの位置付け、及び役割、必要となる機能要件について論じる。

3.1 データ複製

同じデータを複製し複数サイトに置くことにより、(1)信頼性向上、(2)アクセス可能性の向上、(3)性能の向上、という利点が生まれる。反面、データ複製間でのデータの一貫性保持のためのコストが増大するが、情報放送システムでのデータ更新の頻度は小さいことが予想されるので、情報放送システムにはデータ複製が適して

いると考える。障害が発生した場合、従来はローカルにバックアップをとっておき、それを使ってデータの修復を行っていたが、データ複製環境においては近隣サイトからデータをコピーすることにより、バックアップ作業を軽減することが可能になる。つまり、データ複製しておけば、どこかのサイトでデータが破壊されても、全体としてはデータを網羅しているので、ローカルなデータ障害は修復可能となる。データ複製間での一貫性保持の問題については[6]で論じる。

3.2 ディレクトリサービス

情報放送の分散環境において、ある番組はどのサイトに格納されているのか、管理するディレクトリサービス機能が必要である。これには以下のようなディレクトリサービスの機能が含まれる。

- ネームサービス: 番組をすべて名前で管理する。
- データ移動/複製に伴うディレクトリ情報の自動更新: 参照側はデータ移動/複製を意識することなく、同じようにディレクトリサービスが受けられる。
- トレーダーサービス: 番組が提供するサービスによって番組を発見できるようにする⁷。

ディレクトリサービスの標準としては ISO X.500、DCE CDS、Sun NIS+やインターネットの LDAP などがあるが、情報放送ではこうした既存標準をベースとして、どのようなディレクトリ情報を放送で配信するかが重要と考える。

3.3 分散最適配置

複製されたデータをどのように配置するとアクセスコスト最小にできるか、その最適な配置を考えなくてはならない。検索時間に異常に時間がかかるのでは利用者がいなくなってしまうので、ある一定のサービスの質を保てる必要がある。その制約を満たす範囲でどのように分散最適配置をするかが問題となる。アクセスコスト評価に関連するパラメータとしては次のようなものが含まれると考えられるが、実際にどのように組み合わせれば良い評価システムが構築できるのかは今後の課題である。

⁷ネームサービスが名前で検索するのに対し、トレーダーサービスは提供するサービスで検索する。CORBA のオブジェクトサービスのひとつとして提供されている [7]。

- **アクセス頻度:** 利用者からのアクセス頻度の高いものは、可用性を高めるべきである。可用性を高めるには通信コストの低いサイトや、処理能力の高いサイトに複製をおくとよい。可用性向上のためには情報放送で何を放送すべきかという問題も重要である。
- **価値:** データの価値の高いものは、複製を多数作り、障害時にも必ず復旧できるようにすべきである。
- **新鮮度:** データの新鮮度(生成されてから経過した時間)によって、データの価値や、アクセス頻度が変化するデータが存在する。例えば、ニュースや流行に関するものである。そのようなデータはどのアクセス頻度やバックアップの仕方を時間の関数により変える必要がある。

3.4 情報フィルタリング

膨大な量の情報が送られてきた場合、その中から自分の興味／嗜好に応じて有用な情報を抽出及び選別する作業を情報フィルタリングと呼ぶ。情報フィルタリングを行うためには、予め番組制作者側で、番組に対して属性を生成あるいは抽出する必要がある。番組の受信側では自分の興味をそれら属性により表現し、それをプロファイルとして保持する。このプロファイルは固定的なものではなく、受信側の知識獲得につれて高度化したり、利用者履歴により、動的に変化することがある。また、明示的な条件の他、統計的な手法により半自動的あるいは自動的に獲得する場合もある。

情報放送システムにおいては、(1) 利用者の受信端末、及び(2) 各データベースサイトで、情報フィルタリングが行われる。まず(1)の利用者の受信端末では情報放送によって配信された番組をフィルタリングをし、選択された番組のみが受信端末のデータベースに蓄積される。その際の情報抽出の条件としては、利用者が指定した条件の他、それに関連する番組も蓄積することが望まれる⁸。関連番組を発見する手法としては以下が利用可能と考える。

- データマイニングによる相関関係の発見
- 統計情報による規則性の発見

⁸これは、同じアクセスパターンで視聴する利用者でも、それと関連度の高い、換言すると、その番組の周辺番組を提示すると、それら関連番組もアクセスする可能性が高いと、考えられるからである。例えば、プロ野球とニュースしか見ない利用者は、潜在的にプロバスケットの視聴者である、というようなアクセスパターンルールが存在する可能性がある。

WWWのロボット検索の分野においてもキーワードの相関関係を計算することにより、利用者に関連度の高いサイトを提示する機能をもつシステムがある[8]。こうした技術は情報放送の情報フィルタリングにも利用可能である。上記(2)の各データベースサイトでは、そのサイトのデータ収集方針を情報フィルタリングのためのプロファイルとする。データベースサイトの方針としては、映画、スポーツのような分野ごとにデータを収集する方針の他、論理的体系は無視してそのサイトの利用者からのアクセスコストを最小化することを優先してデータ収集をする、という方針もある。後者の場合、上述した関連情報発見手法によりアクセスパターンの規則性を発見することができる。

3.5 索引

索引情報も1次情報と同様に、サイト間で複製されたり、放送で配信されたりする。また、索引情報はサイト間でやり取りされ、分割統合されて利用されると考えられる。そのためには索引はモジュール化できる必要がある。また、受信端末利用者が検索を行う際、引続き関連情報の検索を行う可能性が高いので、検索結果とともに、関連情報のための索引を送ることが考えられる。しかし受信端末はデータベースサイトに比較してデータ容量が小さいので、索引も小さな単位に分割できる必要があるだろう。利用者の受信端末に蓄積された情報は、個人情報であるので、公開しないことを前提とする。その意味では、受信端末はデータベースサイトとは異質である。しかし、検索の効率化のため索引のやり取りが行われることを考えると、受信端末もデータベースサイトとして含めたモデルを想定して、全体のアクセスコスト軽減を考えた方がよいだろう。その場合、対象となるDBSのサイト数は数1000万と推定される。

情報放送システム上の索引の利用に関しては考えるべき項目は次の通り。

- 放送で配信する場合、どういう頻度でどのような索引を配信すると利用者からのアクセスコストが小さくてすむか。
- 索引の粒度はどの程度が適当か。
粒度としては(1) 検索条件で用いる番組属性の粒度と、(2) 検索結果として得られるその番組の粒度を分けて制御できることが望まれる⁹。

⁹前者としては例えば、巨人軍で検索するのか、それとも個々の選手名で検索するのか、という粒度の違いがある。後者は、検索結果としてどの程度の広さをもった情報を返せばいいかというものである。

3.6 データの自動分類

利用者からの検索を容易にするためには、データの組織化が自動的あるいは半自動的に行われることが望ましい。近年マルチメディア情報の組織化に関する研究が数多く行われている。田中ら [9] は Kohonen の自己組織化マップを用いてテキストの自己組織化システム、及び動画の自動分類/検索システムを作成している。また有木ら [10] は、ニュース番組を対象として、意味を表現する音声、内容をまとめる文字、情景や状況を伝える映像を処理し、自己組織化し、希望する内容のニュースの検索を可能とするシステムを開発した。その他、[11, 12] などの研究がある。

こうした自動分類アプローチを用いて、番組を体系づけ、検索効率化を図ることが求められる。

3.7 情報検索の QoS 制御

データの検索を行う際に、数 1000 万規模の利用者がアクセスするデータベースに対して効率のよい情報検索を行うためには、最短ルートで情報の検索を行うことが必要である。しかし、既存技術では、情報放送システムのようにデータベースの規模が巨大で、かつ、伝送量が非対称的で、動的に構成が変化するシステムに対する検索技術は開発されていない。そこで、複製の配置、伝送速度などを考慮した最小アクセスパス計算アルゴリズムを確立し、利用者の要求にあった検索の Quality of Service (QoS) 制御が行えることが必要である。QoS のパラメータとしては、(1) 所要時間、(2) 料金、(3) 検索範囲、(4) 検索結果の返し方 (逐次的か、一括か) などが考えられる。

4 まとめ

本稿では、情報放送の機能モデル及びその利用イメージを説明し、次に情報放送に必要となる超大規模分散データベースシステムの要件について述べた。情報放送ではデータ複製が重要であり、その複製データを超多数地点に最適に配置し、利用者にはディレクトリサービスを提供することが重要となる。また利用者やデータベースサイトは情報放送により配信されたデータを情報フィルタリングして自分のデータベースに蓄積する。そのフィルタリング情報の獲得を半自動化することが必要である。またデータ量が膨大であることから、データを自動的に分類する機構も必要となる。利用者は各種パラメータを使い、自分の要望レベルに合致し

たアクセスパスを発見できるように、情報検索の QoS 制御機能も提供される必要がある。情報放送上のデータベースシステムの研究はまだ始まったばかりであり、今後は上記の各機能についてどのような方式が適しているか検討していきたい。

参考文献

- [1] Imielinski, T., et al.: "Energy Efficient Indexing on Air," *Proc. of ACM SIGMOD94*, pp. 25-36, Minneapolis Minnesota, May 1994.
- [2] NHK 放送技術研究所 (編): 「マルチメディア時代のデジタル放送技術辞典」、丸善、1994。
- [3] 「NHK が BS 向けのデジタル伝送方式、降雨時などに備え 3 種類の変調方式組み込む」、日経ニューメディア、97.6.2, pp. 2-3。
- [4] 中田靖、渡辺博則: 「インターネットでは届かない」、日経マルチメディア、97.4, pp. 48-51。
- [5] 中田靖: 「デジタル放送が主役になる日」、日経マルチメディア、97.4, pp. 52-59。
- [6] 白田由香利、飯沢篤志: 「情報放送システム上での並行処理制御方式の考察」、情報処理学会研究会報告、97-DBS-113-44、1997 年 7 月。
- [7] R. Orfaily, D. Harkey, and J. Edwards: *Instant CORBA*, John Wiley & Sons, Inc., 1997.
- [8] 伊藤耕一郎、河野浩之、長谷川利治: 「異種データベースからの相関ルールによる知識発見- WWW 検索式の生成支援システムへの適用 -」、電子情報通信学会 第 8 回 データ工学ワークショップ (DEWS'97) 論文集、pp. 91-96、1997 年 3 月。
- [9] 田中克己: 「ネットワーク社会とマルチメディアデータベース」、情報処理、Vol. 38, No. 1, pp. 24-29、1997。
- [10] 有木康雄、杉山善明: 「データベースの内容検索と自己組織化 - 画像・音声認識によるニュース映像の検索付けと分類・検索 -」、高度データベース 平成 8 年度研究成果報告会 講演論文集、Vol. 3、1997-01、重点領域研究「高度データベース」総括班、pp. 13-22、1997。
- [11] D. R. Cutting, D. R. Karger, and J. O. Pedersen: "Constant inter-action-time scatter/gather browsing of very large document collections," *Proc. of the SIGIR'93*, March 1993.
- [12] 田中克己、上原邦昭、有木康雄: 「動画データベース: 内容記述とコンテンツによる構造化」、Computer Today、No. 79, pp. 34-39、サイエンス社、1997 年 5 月。