# 1H-1　情報検索システムにおける情報選択提供（ＳＤＩ）の実現手法

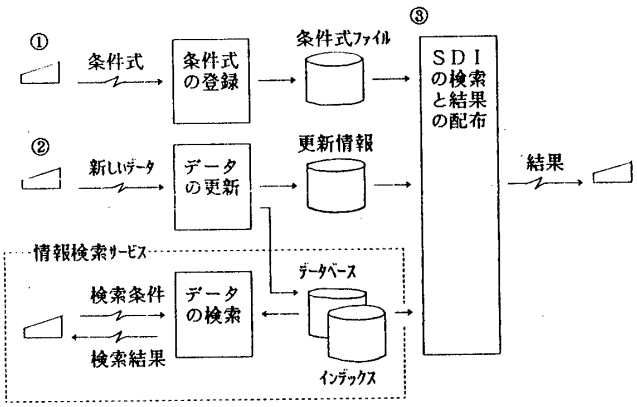
川下　満　　坪井　哲夫　　加藤　嘉明
(NTT電気通信研究所)

## 1．はじめに

ＳＤＩ( Selective Dissemination of Information )とは，システムに新たに追加，又は，変更されたデータのうち，事前に登録された条件に合致するものを配布するサービスである。

データのオンライン更新（オンラインサービス中の追加，変更，削除）を許す情報検索システムでのＳＤＩ処理の流れを図１に示す。この様なシステムでＳＤＩを提供するには，どのデータが更新されたかを知るための情報（更新情報と呼ぶ）の作成と参照を高速に行い，情報検索サービスの性能に影響を及ぼさない事が重要である。

本稿では，この要求を満たす更新情報の管理方式を提案する。




①検索条件の登録（条件式，実行周期，配布先，等）
②検索データのオンライン更新と更新情報の作成
③登録された条件式に従う検索と結果の配布

図１．ＳＤＩ処理の流れ

## 2．提案方式の概要

新たに更新したデータの中から条件に合致するものを高速に抽出するには，

①データ自体に更新された旨の印を付けるよりデータへのポインタを更新情報として持つ方が良い。

②条件に従う論理演算を行う前に検索対象を絞れることが必要。

（図２に条件式が，Ａ and Ｂ で，時刻ｔ以降に更新されたデータを対象に検索する例を示す。）

従って，更新情報の管理は，インデックスと同じ構造で行うことが適当である。この時，更新情報の持ち方として次の２通りが考えられる（図３）。

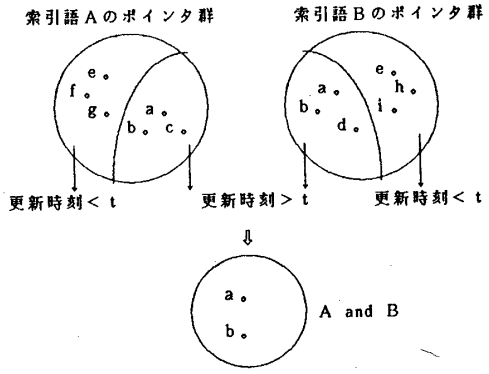


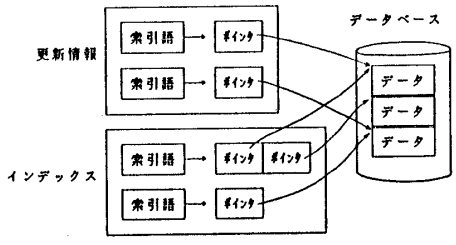

図２　ＳＤＩの検索方法

・方式ａ：更新情報をインデックスとは別個に持つ。

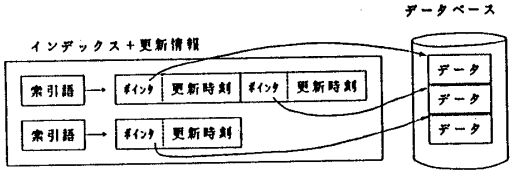
ＳＤＩの検索は，更新情報（構造は，インデックスと同様）で行う。情報検索サービスとは，参照するのがインデックスではなく更新情報である点が異る。

・方式ｂ：更新情報とインデックスを統合する。

更新情報がインデックスと同じ構造となることに着目し，データへのポインタ毎に更新時刻を持つことで両者を統合したもの。ＳＤＩの検索時には，検索対象の索引語配下のポインタ群から，この更新時刻をもとにＳＤＩ対象のポインタを選択した後，索引語毎のポインタ群間の論理演算を行えば良い。




方式ａ




方式ｂ

図３　更新情報の持ち方

---

Index of Database for Selective Disemination of Information
Mitsuru KAWASHIMO , Tetsuo TSUBOI , Yosiaki KATO
NTT Electrical Communications Laboratories

方式bは，更新時刻をポインタ毎に持つためポインタのサイズが数倍となり，インデックスの1ブロック内に納まるポインタ数が数分の1となるため，インデックスの参照時のI／O回数が増す。

方式aは，データの更新に伴いインデックスの更新とは別に更新情報の追加処理が必要であり，その分方式bと比べてI／O回数等が多い。しかし，情報検索サービスでは更新情報を参照しないので影響が無い。

筆者らは，情報検索サービスの性能を重視し方式aが妥当と判断した。本稿では，以下，方式aについて詳細に述べる。

## 3．更新情報管理の問題点と対策

SDIは，即時性がセールスポイントの1つであるが，要求される即時性の程度は利用者毎に異る。そこで，SDIの検索を実行する周期を，登録する条件式毎に指定することを可能とした。

この時，次の問題がある（図4）。

時刻$t_x$に更新したデータは，条件式1では時刻$t_{12}$にSDIの検索の対象とするので時刻$t_{13}$では検索の対象から外す必要がある。このため，更新情報中のこのデータへのポインタを時刻$t_{12}$に削除する必要がある。一方，条件式2では時刻$t_{22}$で検索の対象とするため時刻$t_{22}$以前にこのデータへのポインタを削除することはできない。

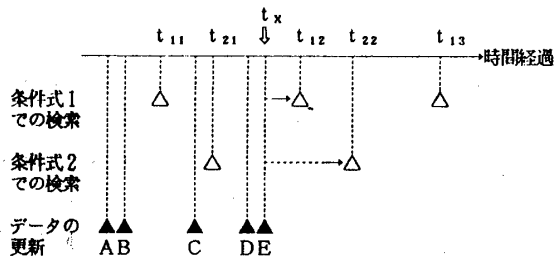



図4　データ更新と検索時刻

この問題は，更新したデータへのポインタを格納する更新情報のエントリをSDIの検索時に切替えることで対処できる（図5）。以下に本方式の特徴的な処理を説明する。

（1）更新情報の追加

更新したデータへのポインタを格納するエントリをSDIの検索時に切替えるため，データの更新処理側に使用するエントリ番号を通知する必要がある。これは，次の方式により矛盾なく行うことができる。

①データの更新処理は，「現エントリ」で示されるエントリに更新したデータへのポインタを格納する。

②SDIの検索処理は，実行直前に時刻テーブルの「現エントリ」で示されるエントリに現時刻を設定する。同時に，「現エントリ」の値を先頭の空きエントリに変更し，以後更新するデータへのポインタが新しいエントリに格納できる様にする。

エントリの切替えは，更新情報を格納するファイルの一部を排他制御することで矛盾なく行なえるので，制御は容易である。

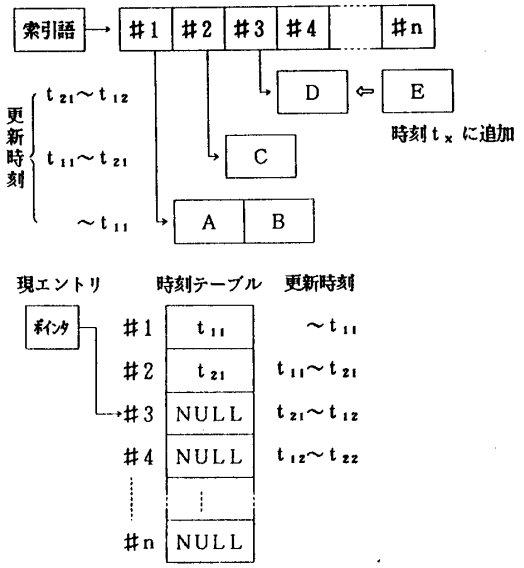



図5　時刻$t_x$での更新情報の作成

（2）更新情報の参照

図4の時刻$t_{12}$のSDIの検索時はエントリ番号2（時刻$t_{21}$）とエントリ番号3（時刻$t_{12}$，図5ではまだ時刻NULL）のポインタ群が対象である。時刻$t_{22}$ではエントリ番号3とエントリ番号4のポインタ群が対象となる。

どのエントリ番号のポインタ群を用いるかを知るには，条件式毎にSDIの前回の検索時刻を持ち，時刻テーブルから前回の検索時刻より大きい値を持つエントリを探せば良い。

なお，エントリ番号は時間経過に従って昇順に使用するので必要なエントリを探すことは容易である。

（3）更新情報の削除

更新情報の削除は，登録されている全ての条件式でのSDIの検索を終えたエントリから順に行うことが原則であるが，検索の実行周期が短期の条件式（分単位，時間単位）と長期の条件式（週単位，月単位，等）が混在すると，長期の条件式での検索が終了するまで更新情報が削除できずエントリ数が多くなり，必要なエントリを探す際の性能が悪くなる。

従って，1日のサービス終了時等の契機で検索の行われなかった条件式での検索を行い，結果を配布する時まで保存することにより，更新情報は削除することが望ましい。

## 4．おわりに

データのオンライン更新を許す情報検索システムでのSDIの更新情報の管理方式を提案した。提案した方式は，情報検索サービスの性能を重視し，このサービスへの影響を少なくすることを前提としている。

また，この方式は既存のインデックスの構造に影響を与えることなくSDIの実現を可能にするものである。

〔参考文献〕青木，他「マルチメディア情報検索システムについて」情処第33回全国大会（4H－1）