

オンライン情報選択提供 (SDI) における結果保持方式について

7H-2

石黒正典 田中一敏
NTT情報通信処理研究所

1. はじめに

SDI (Selective Dissemination of Information) は、利用者があらかじめ登録した検索条件式をセンタ側で周期的に実行し、その結果を利用者に配布するものであり、情報検索サービスにおいて有用な機能である。我々は、鮮度の高い情報の提供を目的に、データのオンライン更新を実現し¹⁾、SDI機能の実現に有効な、時間単位に基づく更新情報の管理方式を提案した²⁾。

本稿ではさらに、利用者毎に異なる配送契機を考慮した効率良い更新情報の保存方法を提案する。

2. SDIの基本メカニズム

オンライン更新データを管理する「更新情報管理用IVL: InVerted List」を用意する。本IVLは時間を索引語とし、対応する時間に更新されたデータベースレコードへのポインタ (RID) 群を管理する。登録された検索条件式は「SDI条件式管理テーブル」で、利用者ID、配送契機等とともに管理される。図1にSDIの基本メカニズムを示す。

- ① 配送契機にSDI条件式を実行し、「SDI条件式実行結果RID群」を得る。
- ② ①と並行して「更新情報管理用IVL」の必要な索引語を参照し、「配送候補RID群」を得る。
- ③ 「SDI条件式実行結果RID群」と「配送候補RID群」との論理積を行うことにより、「SDI配送RID群」を得、それに対応するデータベース内のレコード実体を利用者に配送する。

3. 実現方式

検討上前提としたサービス条件を以下に示す。

- (1) SDI条件式毎に、(a)時間間隔指定(例.3時間おき)、(b)日間隔時刻指定(例.10日おきの12:00)、(c)特定日の時刻指定(例.毎月15日の10:00)の3種類の配送契機を指定可能とした。1日のサービス時間帯は決まっているものとし、以降(a)を「当日配送のSDI条件式」、(b)及び(c)を「日またがりのSDI条件式」と呼ぶ。
- (2) 配送はセンタ起呼/端末起呼いずれも可能とする。
- (3) 配送失敗時、及び「日またがりのSDI条件式」は、配送完了まで「配送RID群」を保存する。以上より「配送RID群」保持方式案を比較する。

案1: 保存ファイル不要化方式(図2)

「更新情報管理用IVL」が「配送候補RID群」の管理/保存を兼ねることにより、保存用の別ファイルを不要化する。「当日配送のSDI条件式」/「日またがりのSDI条件式」の双方に適用可能なように、1索引語を1時間に対応させ1年分程度の索引語(24×365=8760)を用意する。【索引語数の削減のため、時間管理用、日管理用(1索引語は1日に対応)の2種類の「更新情報管理用IVL」を用意し、サービス終了時に時間管理用の全索引語のRIDをマージし、日管理用の索引語配下につなぎかえる改善策が考えられる。しかし、その日の配送済みRIDが日管理用の「更新情報管理用IVL」で重複管理され、次回の配送契機に配送済みの情報が配送対象となってしまう方式として不十分である。】

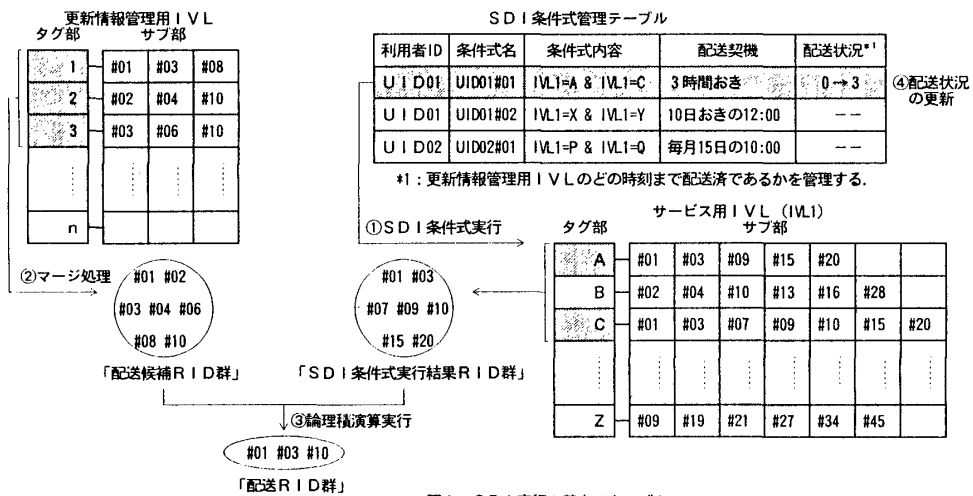


図1. SDI実行の基本メカニズム

案2：利用者ファイル保存方式（図3）

「更新情報管理用 I V L」は1時間単位に1日分のみとする。利用者毎に「配送 R I D 群」保存用ファイル（「配送 R I D 群保存ファイル」）を1つ用意し、サービス終了時点で未配送の「配送 R I D 群」を該当する利用者の保存用ファイルへ書き出す。

案3：I V Lファイル保存方式（図4）

（案2）と同様に「更新情報管理用 I V L」は時間単位に1日分のみとし、「配送 R I D 群保存ファイル」に I V L を用いる（1索引語は1 S D I 条件式に対応）。サービス終了時点で未配送の「配送 R I D 群」を該当 S D I 条件式の索引語配下に連結する。

4. 評価

ファイル容量、S D I 実行性能等の観点から各方式案の評価を行った。①1日当たりの更新データ数、②S D I 利用者数、③S D I 条件式の特徴（当日配送が多い/日またがりが多い）等が考慮すべきポイントとなる。各方式案の比較結果を表1に示す。

- ・案1は、日またがりの条件式が多い場合、参照する索引語数の範囲が広がり処理時間が問題になる。
- ・案2/案3は、サービス終了時の処理が多いが、利用者に不都合を与えるものではない。
- ・案2ではファイル容量面で負荷が大きすぎる。
- ・案3がファイル容量、S D I 実行処理時間において良好である。

我々が適用するサービス例では、①=3000件、②=2000人、③=日またがりの S D I 条件式が多い、であり、これに対しては案3が適切といえる。当日配送の S D I 条件式が多いサービスであれば、案1が適切である。

5. 今後の課題

適用対象のサービスは、1日のサービス時間帯が決まっていることを前提としたが、24時間運転のサービスへの適用も考慮する必要がある。案3のサービス終了時に行う処理を深夜のトラフィックが低い時間帯にバックグラウンドで実行する等の対策が必要である。また、株式情報のようなリアルタイム性が必要なサービスに適切な方式も検討課題である。

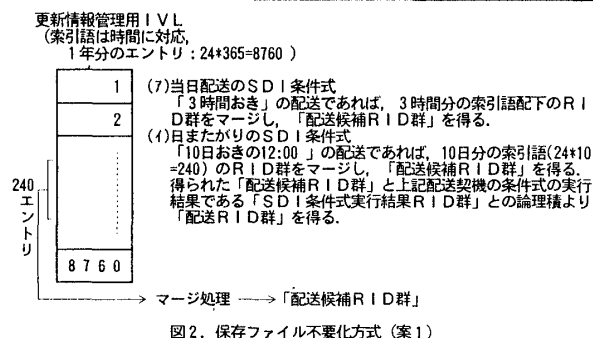
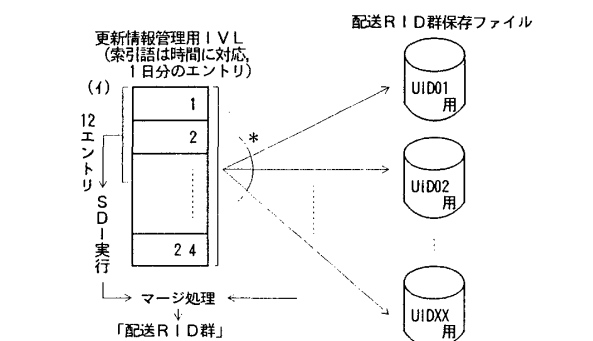


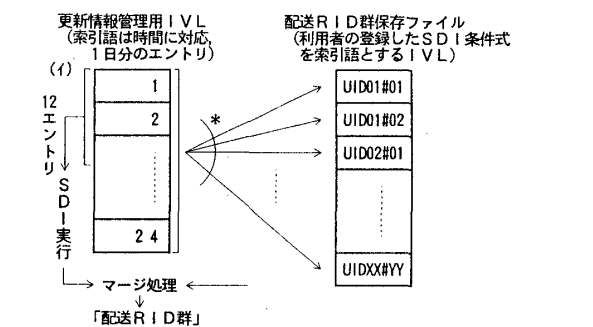
図2. 保存ファイル不要化方式（案1）



*サービス終了時において、当日未配送の S D I 条件式を実行し、該条件式対応に得られる R I D 群を各利用者の配送 R I D 群保存用ファイルに格納する。

- (7)当日配送の S D I 条件式（案1）に同じ。
- (4)日またがりの S D I 条件式「10日おきの12:00」の配送であれば、更新情報管理用 I V L の12:00 までの索引語配下の R I D 群に対する S D I 実行結果と、配送 R I D 群保存ファイルに保持されている R I D 群とをマージし、「配送 R I D 群」を得る。

図3. 利用者ファイル保存方式（案2）



*サービス終了時において、当日未配送の S D I 条件式を実行し、該条件式対応に得られる R I D 群を配送 R I D 群保存ファイルの該当する索引語配下に保存する。

- (7)当日配送の S D I 条件式（案1）に同じ。
- (4)日またがりの S D I 条件式「10日おきの12:00」の配送であれば、更新情報管理用 I V L の12:00 までの索引語配下の R I D 群に対する S D I 実行結果と、該 S D I 条件式に対応する配送 R I D 群保存ファイル（I V L）の索引語配下の R I D 群とをマージし「配送 R I D 群」を得る。

図4. I V Lファイル保存方式（案3）

表1. 各方式案の比較

| 方式案 | ファイル容量 | | 「日またがりの S D I 条件式」の 実行時間 | サービス終了時の処理 |
|--------------------|-----------------------------------|---|--|---|
| | 更新情報管理用 I V L | 配送 R I D 群保存ファイル | | |
| 保存ファイル不要化方式（案1） | 時間を索引語とする1年分（24*365=8760）のエントリが必要 | 不要。更新情報管理用 I V L が配送 R I D 群保存ファイルの代替となる。 | 長期の日またがりにおいては、マージ対象となる索引語数が急増する。 | 特別な処理は不要。 |
| 利用者ファイル保存方式（案2） | 時間を索引語とする1日分（24）のエントリが必要 | 利用者毎に1つの配送 R I D 群保存ファイルが必要。利用者が増える度に運用側でファイルを追加する必要あり。 | 配送 R I D 群保存ファイルより得る R I D 群（前日分）と更新情報管理用 I V L より得る R I D 群（当日分）とのマージを行う。 | 1日分の更新情報管理用 I V L の R I D 群を配送候補 R I D 群として日またがりの S D I 条件式を実行し、得られた配送 R I D 群を配送 R I D 群保存ファイルに格納する。 |
| I V L ファイル保存方式（案3） | 案2と同じ。 | ユーザ数 m で、1人平均 n 個の S D I 条件式を登録すると、m * n エントリ分の索引語が必要。条件式が追加登録されても、動的に追加可能。 | 案2と同じ。 | 案2と同じ。 |

【参考文献】

- 1) 坪井, 川下: 情報検索システムに於けるインバーテッドファイルのオンライン更新実現手法, 情報33回全国大会 (1H-2)
- 2) 川下, 坪井: 情報検索システムにおける情報選択提供 (S D I) の実現手法, 情報33回全国大会 (1H-1)