

9 データベース化に付随する機能

最近のU社システムの世界では、全文検索が容易に出来るソリューション・パッケージ・ソフトが作成され、市販されている。これはいろいろのトータル・システムを構築する上でのツールとして利用が可能であり、本システムの構築についても経済的、時間的観点からその利用について大いに検討すべきである。全文データベース・システムには、3つの形態が考えられる。

① 二次情報データベースに全文データベースを付加して構成されるデータベース

このデータベースでは、二次情報データベースの各レコードに対応する文書が結び付けられており、検索は、二次情報データベース上で行う。検索結果の表示に一次情報データベースの内容を利用する。この方式は、従来のデータベースに拡張機能を追加することによりシステムを比較的容易に構築できる利点がある。

② 一次情報のみからなるデータベース

このタイプのデータベースは、2種類に分類され一つは、文書を文字列として扱うものと、章/節といった文書構造を考慮するものに分けられる。前者のシステムで中心的な役割を果たすのが文字列照合機能である。後者のシステムでは、文字列照合機能も重要な役割を果たすが、さらに文書構造を記述するために、SGML (Standard Generalized Markup Language) などが利用される。

③ ①と②のシステムの融合型で、文書から検索によく使用される部分を二次情報として取り出し、一次情報と合わせて検索の効率を上げるシステム

データベースは、①と同様一次情報データベースと二次情報データベースで構成されるが、一次情報でも検索できる点が異なる。

本システムのように全文のデータを検索する手法としては、文書構造を考慮しない方法と、文書構造について考慮する方法の2とおりを考慮する必要がある。このうち、文書構造を考慮しない全文型検索手法としては、

① 文字列照合アルゴリズムによる方法

② 索引語による全文検索の方法

等が考えられる。この索引語は、人間が作成する場合と、システムが自動的に作成する場合が考えられる。もちろん自動索引を可能とするためには、それに対応するソフトの作成が必要でこのような機能が作成できるかどうかは、投入出来る予算を勘案する必要がある。しかし、データの基となる自作教材を、大勢の指導員が作成するとなれば、検索語の付け方について各指導員が個々に対応したとき、統一された検索語の付け方は、極めて困難とならざるを得ない。したがって、本システムのような場合は自動インデックス機能が、システムの円滑な運用を考慮した場合必要と考えられる。このほかに

③ 類似検索法

が考えられる。また、文書構造を考慮する検索方法としては

① 文書構造の扱いを利用者に委ねる方法

② システムが文書構造を解析のうえ変換を行い処理する方法

があると考えられる。この選択もソフトウェアの開発費を考慮した上で判断する必要がある。

(1) 検索キーワードの設定 (マルチ・インデックス化)

教材開発支援システムのデータベースを、指導員が利用し検索する場合、検索する範囲やその分野を限定することにより、目的とする情報（データベースの内容）を的確に短時間で入手することが可能となる。このための手段として必要とされるインデキシングを容易に行えるようにするためには、キーワードを制御する辞書機能を、当システムに内蔵する必要がある。キーワードをつけるための方法として次の2とおりがある。

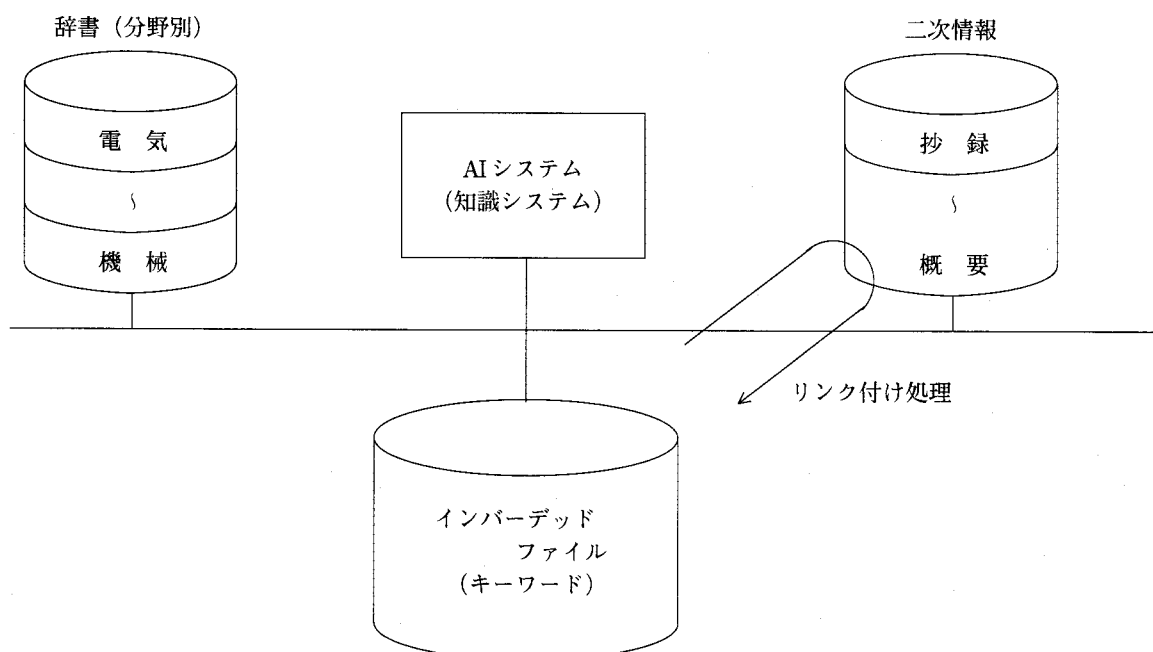
① 自作教材の作成者が自らキーワードを登録する方法

これは、新規の教材作成時に、入力段階で作成者が自分で、一定のルールに従いキーワードを登録する方法である。登録する対象者が多ければ多いほどキーワードの均一性に問題が生じやすく、後日の利用に不都合が生じやすい。

② 既存の自作教材やモデル教材にシステムが自動的にキーワードを作成し登録する方法

これは、二次情報（例えば、抄録や概要等）からシステム内既存の辞書機能と、AIシステム（統計機能と学習機能を合わせ持つシステム）の応用により、機械的にキーワードを作成する方法であり、本システムの利用形態を考慮した場合、この方法が適当であると考えられる。

この概念図は下記のとおりである。



(2) 検索機能 (部分、全文)

a 利用者の検索ニーズ

訓練用教材を開発又は加工・編集し活用する利用者にとって、期待される検索機能を推察すると、求める情報は以下の4種類に大別できる。

- ① 類似する訓練コースのモデル教材本体
- ② 類似する訓練コースの指導案

③想定する訓練用教材の一部となる図面等の教材部品

④想定する訓練用教材の参考となる資料情報

は検索機能を実現する上で、検討すべき課題は少なくないが、以下の検討事項については、システムの基本設計に与える影響も大きく、将来の拡張性を制限することにもつながるので、十分な期間をかけ詳細な検討を行う必要があると考えられる。

①キーワードの設定に関する事項

②レスポンスビリティに関する事項

③検索の結果として提示される情報に関する事項

b 教材本体の検索

事業団本部で作成しているカリキュラムモデルに対応するモデル教材本体の検索については、カリキュラムモデル・コードからその教材の概要がほぼ把握出来ると考えられるので、カリキュラムモデル・コードによる検索が考えられる。

この場合、その機能にはコードの一部にマスクをかけることが出来るようにする必要がある。また、カリキュラムモデル以外の教材を含む本体の検索で検索キー指定によるものでは、各系別分類名、表題、要旨、訓練目標、教科の細目、教科の内容、本文そのものが考えられる。各系別分類名、表題、本文についてはキーワードの設定を行わず、全体を検索の対象にし、要旨、訓練目標、教科の細目、教科の内容については、データ入力段階で、キーワードを設定し検索キーによる検索と、全体を対象とする検索の2種類の機能を提供するべきと考えられる。

区 分	対象項目指定あり	対象項目指定なし
設定キーワード 対 象	要旨、訓練目標、教科の細目、教科の内容のうち指定された対象（複数指定可）に設定されたキーワードを対象に検索する。	要旨、訓練目標、教科の細目、教科の内容の全てに対して設定されたキーワードを対象に検索する。
全 文 対 象	分類名、表題、要旨、訓練目標、教科の細目、教科の内容、本文のうち指定された対象（複数指定可）の全文を対象に検索する	分類名、表題、要旨、訓練目標、教科の細目、教科の内容、本文の全てを対象とし、その全文を検索する。

検索の結果、検索キーの条件に合致する対象が存在した場合、利用者に提供される情報は、今後検討される必要があるが、次の3項目を含めることが適当と考えられる。

①カリキュラムモデルコード

②分類名、表題

③要旨

このことにより、利用者は提示された情報を基に、希望に合致した教材全体を必要に応じて入手出来るであろう。

c 訓練指導案の検索

訓練指導案の検索についても、教材本体と同様にカリキュラムモデルコードによる検索機能は提供さ

れるべきであろう。

そのほかに、分類名、表題、訓練レベル、訓練対象者、訓練目標、使用する機械器具、訓練時間を対象に、検索できることが必要とされるであろう。これら検索対象項目のうち、訓練対象者、訓練目標についてはキーワードの設定を検討する必要があるが、分類名、表題、訓練レベル、使用機械器具、訓練時間については、全文検索のみとして問題はないであろう。

区 分	対象項目指定あり	対象項目指定なし
設定キーワード 対 象	訓練対象者、訓練目標のいずれかに設定されたキーワードを対象に検索する。	訓練対象者、訓練目標の両方に設定されたキーワードを対象に検索する。
全 文 対 象	分類名、表題、訓練レベル、訓練対象者、訓練目標、使用機械器具、訓練時間のうち指定された対象（複数指定可）の全文を対象に検索する。	分類名、表題、訓練レベル、訓練対象者、訓練目標、使用機械器具、訓練時間の全てを対象とし、その全文を検索する。

ただし、分類名、表題、訓練レベル、使用機械器具、訓練時間については、検索キー設定時に、項目を特定すべきとも考えられ、実施段階で検討が必要であろう。

検索の結果、提示される情報は次に示すもので十分であると考ええる。

- ①カリキュラムモデルコード
- ②分類名、表題
- ③要旨
- ④教科の細目、教科の内容

d 教材部品の検索

教材部品の検索においても、カリキュラムモデル・コードによる検索機能が、前記の各検索と同様に提供されるべきかどうかは検討の余地がある。

教材部品においては、教科の細目、教科の内容が検索対象の中心になると予測され、対象の多量さ、検索キー条件の合致判定の難しさ（論理和・積の設定基準等）が容易に想像され、アップ・ポスティングを含めたシソーラスの必要性が痛感される。このような状況で恒常的な運用を含めて考えた時に、その実現可能性に知識システムの活用も検討されるべきである。

運用段階に移行した後も、均一な品質が保てるような基準（ルール）の設定が可能であるか否かがそのポイントであろうが、とりあえず、基準（ルール）の設定を模索してみるべきであると考ええる。

教材部品の検索では、検索の対象、情報提示の対象について、その単位をどのように決定するのか検討する必要がある。また、対象の単位が決定されたとしても、検索キーの条件に合致する対象が決定できた場合において、提示すべき対象が、該当する対象のみに止めるのか、何等かの条件を満たす付随情報も付加して提示すべきであるのかも、一層の分析、検討が望まれる。

e 参考資料の検索

参考資料の検索は、前記3種類の検索と明確に異なり、検索の対象が教材以外であり、その範囲は、運用が進むにつれて拡大していくと考えるべきものである。

この分野については、既存のデータベースも少なからずあり、その利用も十分検討すべきであるが、他

システムとのインターフェースは別途検討するものとして、それとは別に、自システム内にも独自のデータベースを内包すべきである。

参考資料データベースは、収容することが望ましいデータもさることながら、現実的に収集可能なデータから逐次入力することが戦略的情報システム構築のセオリーからも妥当と考える。

参考資料データベースについては、データの類別、シソーラスの生成をあきらめるのも、ひとつの考え方である。その代わりとしては、キーワード設定のシンプルな基準を設定し、そのキーワードに対する自動検索が、高速に行なえる仕掛けの構築に力を傾注すべきであろう。

なお、検索の結果提供するデータについても、カリキュラムモデル・コードに類するコード体系を設定する必要があると共に、表題、要旨等の取り扱いが検討されなければならない。何れにせよ、参考資料の検索については、他の3種類の検索と比べ、システムの構築にあたり、検討、決定すべき要素が多いことは否めず、決定に際しては、全ての基準を単純化すべきことを付言しておく。

(3) 加工・編集機能

これは、教材データベースの一次情報（全文データ）を検索し、提示されたデータを、施設の器材やカリキュラムに適応するように、指導員自らデータを再編集する機能である。このため、ワープロ機能を使い、文章の追加、変更、削除を行ったり、図形、写真、イメージの追加、変更、削除をグラフィック機能を使用し、また、全体のページレイアウトを自動レイアウト機能で校正し直すことが求められる。ワークステーション上では、これらの機能が統合されたDTP（デスクトップパブリッシング）機能として存在するし、パソコン上でも、WINDOWSを使ったOAツール（ワープロ、表計算、図形処理）やCADツールで、データベースを加工・編集することができる。加工された一次情報は、自動的に文書管理や履歴管理されることが好ましい。このため、訓練科目、施設の実体に見合った利用者インターフェイスGUI（グラフィクス・ユーザー・インターフェイス）を採用する必要がある。

(4) 出力機能（伸張）

教材データベースの情報を、各種デバイスやメディアに出力する機能である。

- ①FDD（5インチ・フロッピーディスク、3.5インチ・フロッピーディスク）
- ②CD-I（CD-ROM化するデバイス、ISO9660フォーマット）
- ③カートリッジ・テープ（8mmカートリッジ・テープ）
- ④磁気テープ（6250バイト／インチ）
- ⑤光磁気ディスク（書換え可能型MOディスク）
- ⑥LBP（レーザー・プリンター、400BPI以上の解像度、日本語Post Script対応）
- ⑦カラー・プリンター（24万色、400BPI以上の解像度）
- ⑧カラー・スライド（パソコン接続）
- ⑨プロッター（2次元、3次元図形）

等の出力デバイスが想定される。このため、各種デバイスに対応するドライバーソフトやデバイス・ハンドラー・ソフトが必須となる。

データの登録時に圧縮されたデータは、出力時に、伸張技術を利用し、元のデータに復元する必要がある。

る。また、マルチメディア対応のデバイス、ソフトも市販されているので、今後の教材作成ツールとして、訓練時の利用方法についても検討すべき課題である。

(5) 標準（オープン対応）型データベース

当教材開発支援システムを構成するデータベースは、従来のコンピュータ・メーカー固有のデータベースから、業界標準のリレーショナル・データベースを採用すべきである。

データベースの構造モデルは、ネットワーク型、リレーショナル型、オブジェクト型が存在する。ネットワーク型は、「レコード」指向により大規模データ管理を築く定型業務向き（例えば銀行の顧客データベース）で、手続き的データ操作言語（DML）を要する。

リレーショナル型は、「集合」「値」指向でデータの独立性を提供する非定型業務向きで、2次元の表（テーブル）で関係を表し、非手続き的データ操作言語（SQL）を要する。

オブジェクト型は、「オブジェクト」指向によりデータと手続きの一体化（カプセル化）が可能で、将来マルチメディア対応データベースやCADデータベースに適用される。教材データベースは辞書、一次情報、二次情報とそれぞれ目的別にデータベースが構築され、それぞれの情報の関係が2次元で操作できる非定型業務であることによりリレーショナル・データベースが適当と考えられる。

特に、各職業能力開発施設に設置する分散型データシステムには、オープンシステム対応型（ORACLE、SYBASE、INFORMIX等）を採用すべきである。

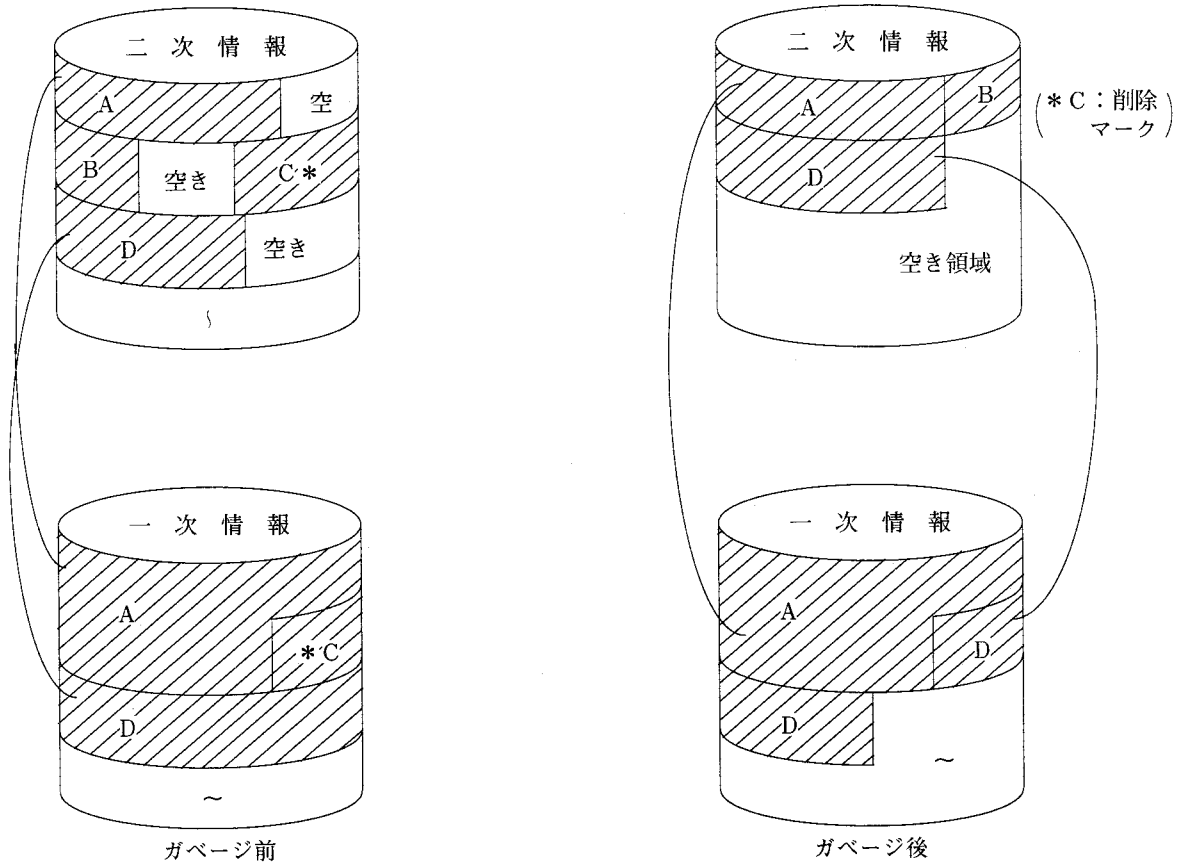
これは例えば、ORACLEリレーショナル・データベースシステムの特徴として、次のものがあげられる。

- ①分散検索、分散更新機能（2フェーズコミット）
- ②データベーストリガー（変更データに応じて、他のAPを自動的に実行）
- ③ストアード・プロシージャ（回線上の負荷軽減）
データベース内にSQLプロシージャを格納
- ④機密保護機能の強化
- ⑤データハンドリング強化（1フィールド2GB）
長大なテキストデータやイメージデータに対応
- ⑥マルチネットワーク対応（プロトコル変換機能）
- ⑦データベースの信頼性強化（ログファイルの二重化）

(6) ガベージ、メンテナンス機能

データベース管理システムの機能として、ガベージ・コレクションがある。当教材開発支援システムが運用された場合、データベース上に新規データの登録、一部修正削除等のデータの更新が実行されることが想定される。その結果、ディスク上に空き領域が生じ、検索レスポンスの遅延やディスク領域の有効利用がはかれなくなる。

このため、休日等のシステムに負荷のかからないタイミングでガベージ・コレクションを実行しディスクの空き領域を整理する。また、メンテナンス機能としては、トレース、ダンプ、リンク付け、インストール、バックアップデータベース等のユーティリティ・ソフトが必要となる。



(7) セキュリティ機能 (ウイルスチェック、データベース更新、破壊防止)

FDD、CD-ROMの授受、LAN (ローカル・ネットワーク) の採用、パソコン通信の普及、ゲームソフト、パソコンの拡大により、現在発見されているウイルスは、1700種類を越えている。このウイルスにより、中央のデータベース、施設の分散データベース及びプログラム・ライブラリが破壊され、ウイルスの増殖にもつながる恐れがある。

クライアント/サーバ形式のオープン・システムでは、NETWARE サーバ上で動作するウイルス退治ソフト (ウイルス・プロテクション・アナライザ) が、各種市販されている。

集中データベース、分散データベースは、データベース・マネジャーのみしかデータベースを更新出来ないよう、セキュリティ機能をRDBMS (リレーショナル・データベース・マネジメント・システム) にもり込む必要がある。

データベース検索者 (指導員) も、システム運用時には、前もってシステムに登録しておき、DB利用時には、その都度利用者を確認する方法を取ることも考える必要がある。(利用者名、利用者コード、ユニークなID)

データベース更新は、リアルタイムで更新する方法と、ディレード (一定期間: 1週間) で更新する方法 (一括更新と分割更新) があるが本システムでは管理運営組織の関係からどのような方式を採用すべきか検討する必要がある。

中央のデータベースは、利用者側からは基本的にリード・オンリー (検索のみ) とし、レコードロック、ファイルロックといったプロテクト機能も合わせて持つ必要がある。

(8) データベース化するデバイス（磁気ディスク、光ディスク・ボックス）

当システムで、特に高速な検索や更新が必要とされる目的別データベース（例えば、辞書DBや検索キーワードDB等）は、高速磁気ディスクに搭載すべきである。また、大容量の目的別データベースが想定される自作教材本体等の一次情報DB群は、光ディスク（ディスク・ボックス型）上に構築することが、経済的側面から有効である。このようにセンター・システムを考え、教材データベース群は2階層の磁気ディスクと光ディスクから構成されるべきで、バックアップ・データベースは光ディスクに自動的に蓄えられることが必要であろう。

(9) データベース引用、検索、出力のカウント機能

中央の教材データベースは、各施設の共通資源でもあり有効に活用されるためには、常に最新の一次情報及び二次情報を維持する必要がある。このため、利用者側の要望や利用度合を的確に把握し、どの分野、対象の技術情報、図書情報、モデル教材、カリキュラムなど収集すれば良いか等のパラメータ情報（基礎情報）を収集するために、カウント機能も検討される必要がある。

例えば、一次情報は、①収集 ②分類・整理（キーワード、分類コード化）③蓄積（データベース化）④検索・利用 ⑤処分または廃棄 といったライフ・サイクルをたどる。カウント機能を設けることにより、これらのサイクル管理にデータベースの引用（類似教材から実際の訓練教材へ活用、流用したケース）、検索頻度、各種出力デバイスへの出力傾向をとらえることが可能であり、ライフ・サイクルの最終段階である情報の廃棄処分（データベース上から削除する）という処理へ進めることが出来る。これはすなわち一次情報（全文）のバージョン管理に連動するものである。

(10) 課金、モニター機能

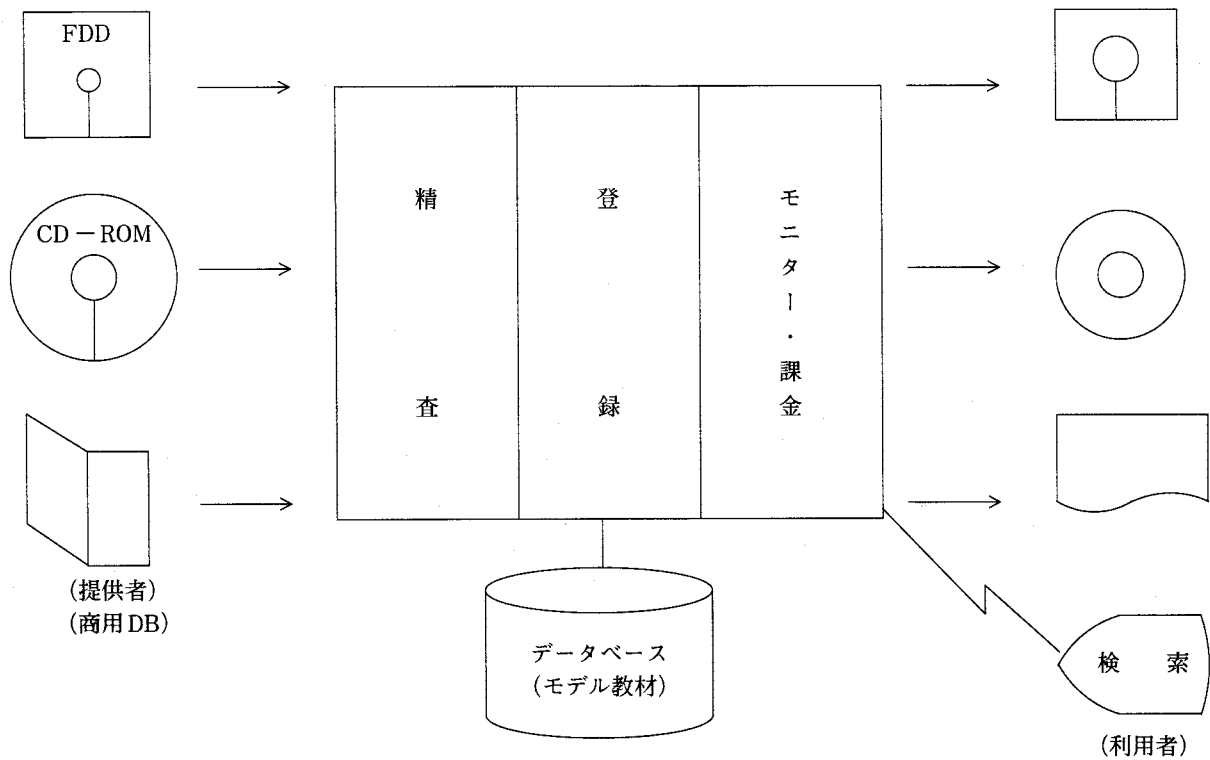
新規に登録されたモデル教材やカリキュラム等の目的別DBの状況や各施設の利用者から、どの位活用されているかを監視（モニター）する機能。

①二次情報の登録、更新、削除、検索回数

②一次情報の登録、更新、削除、検索回数

③外部ファイル（例えば、FDD、CD-ROM、印書など）へ出力した回数

データベースの利用度合いにより、今後の情報収集や教材提供者への見返りといった基礎情報を把握することが可能である。（課金機能）



一方、モニター機能により、各種目的別データベースを、如何に最適な配置にするか（利用度の高い目的データベースは、磁気ディスク上に配置するなど）がある。

他方、課金機能は、従来情報処理（コンピュータ使用料や回線料金も含む）にかかった費用を、部門毎に集計し、経費としてカウントされる機能である。この課金機能を利用して当システムの課金・計算式を作る。例えば

a 精査費用（許諾料含む）	(提供者)
b データベース入力費用	
c FDDで配布 × N回	(利用者)
d CD-ROMで配布 × N回	
e 紙ベースで配布 × N回	
f 回線経由で配布 × N回	

(利用者側の経費) - (提供者側の経費) = 回収費

回収費がプラスになれば、利用度合いが大きいことを表わす。