

氏 名 小 畑 征 二 郎

授 与 学 位 工 学 博 士

学位授与年月日 昭和 62 年 1 月 14 日

学位授与の根拠法規 学位規則第 5 条第 2 項

最 終 学 歴 昭和 38 年 3 月

東北大学教育学部学校教育学科卒業

学 位 論 文 題 目 オンライン情報蓄積・検索システムに関する研究

論 文 審 査 委 員 東北大学教授 野口 正一 東北大学教授 城戸 健一

東北大学教授 重井 芳治 東北大学助教授 宮崎 正俊

## 論 文 内 容 要 旨

計算機システムを使って会話的に情報を検索するシステム、いわゆるオンライン情報検索システムは、1966年に初めて米国のロッキード社において開発された。その後、同じようなシステムが各所で開発され、検索サービスが行われるようになった。しかし、これらの多くは専門家がデータベースを構築し、利用者には検索機能だけを開放するシステムである。ところが、計算機システムの利用が多様化するに伴い、個人や研究室などが単位で独自にデータを収集し、データベース化してデータの整理や検索に利用したいという要望が高まってきた。

本論文の目的は、これらの要求に応えるシステムのもつべき機能を明確にして、その基本構成およびその実現方式などについて論じることである。

本論文は、7章からなり各章の概要は次の通りである。

第1章は、序論である。

第2章では、先ずオンライン情報蓄積・検索システムのもつべき機能を明確にしている。具体的には、(1)データベースの枠組の作成、(2)データの蓄積と修正、(3)データの検索と結果の表示、(4)データベースの管理・保全、(5)利用者ビューによる外部ファイルの作成などの機能を最低限もつべきで、これらはすべてオンラインで利用可能であるべきとしている。次に、システムで扱うデータをモデル化し、システムの基本設計を行っている。その設計によって得られるシステムは、処理部とデータベース部からなる構造モデル、および機能を表現するエンドユーザ言語（以下、EULとい

う)と論理コマンドからなる機能モデルの2層モデルからなるとしている。また、処理部は利用者とのインターフェースをとる利用支援部と処理の中核をなす機能部からなる。利用支援部は、EULを用いて利用者とのインターフェースを取る部分であり、機能部に対しては論理コマンドを発行する。

データベース部は、データの集合であるデータ・ファイルとインデックス・ファイルからなる。インデックス・ファイルは、インデックス・タームとそのインデックス・タームをキーワードとするデータへのポインタの集合からなるインデックス・レコードの集まりである。このインデックス・レコードのアドレッシングには、インデックス・タームをキーとしたハッシュ法を用いている。このハッシュ法を使うことによって、オンライン・リアルタイムでのデータの更新を高速かつ円滑に行うことが可能となっている。

また、処理部とデータベース部との間には、データベースの構造およびデータの構成要素やその属性を記述したデータベース定義ファイルを置いて、処理部とデータベース部の独立性を保証している。

第3章では、先ず第2章で設計した論理システムを、特定しない計算機システム上で実現する方式について述べている。

実現するシステム(以下、実現システムという)の構造の概要を示すと、図1のようになる。

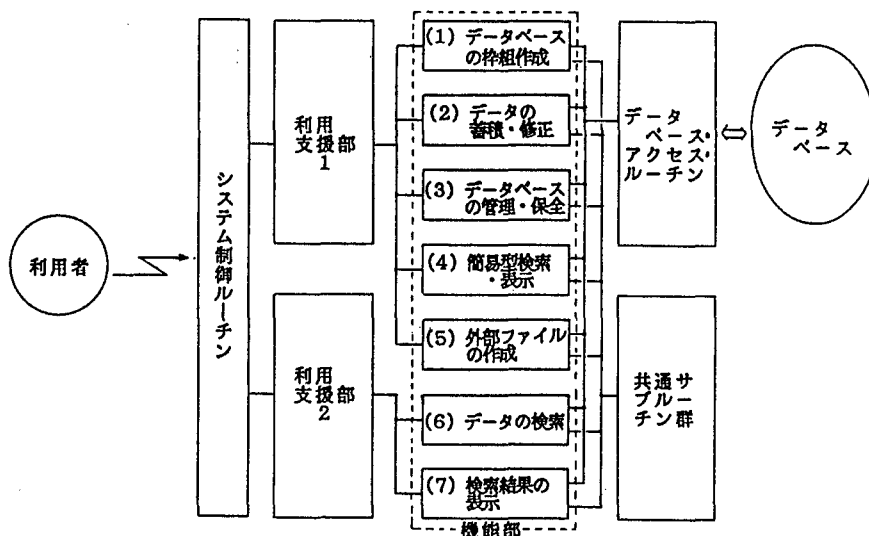


図1 システムの構造

利用支援部1はシステム誘導型のEULを実現する部分で、利用支援部2はシステム受動型のEULを実現する部分である。システム誘導型とシステム受動型の二つのEULをもつことによって幅広い利用者層の利用に対応することができる。

機能部は、七つのモジュールに分けてある。また、各モジュールで共通に使われる部分は、共通サブルーチンおよびデータベース・アクセス・ルーチンとしてまとめてある。

一般に、インデックス・ファイルのインデックス・タームのもつデータポインタの数は不定である。実現システムでは、インデックス・レコードを固定長とし、その固定長のレコードに入り切らないデータポインタを、同じく固定長のオーバーフロー・レコードにリンクして格納する方式を取っている。また、データ・ファイルに関してもデータ・レコードを固定長とし、1レコードに入り切らない部分を同じく固定長のオーバーフロー・レコードにリンクして1データを構成するようにしている。

実現システムの特徴は、論理コマンドのコマンド名がモジュール名（サブルーチン名）に対応し、コマンドのパラメータがサブルーチン・コールにおける引き数に対応させて実現できること、また、インデックス・ファイルとデータ・ファイルの論理レコードを固定長の物理レコードの連結で実現していることである。

また、第3章では本システムを実現するための必要条件を述べ、実際に日本電気K.K.のACOS-FOTRANでオンライン情報蓄積・検索システムUNIQ-1(Universal Information Query-1)として実現した例を示している。このことによって、システムの基本設計およびその実現方式の妥当性が証明されている。

第4章では、システムの性能と最適化に関して考察している。まず、実験とその結果の考察によりシステム性能の決定要因を明確にしている。それによると、システム性能の決定要因は、蓄積するデータの特性に適合したインデックス・ファイルの構造である。次に、以上の結果をもとにインデックス・レコードのサイズを変数とするシステムの評価関数を定義して、その関数を使ったシステムの最適化の手法について述べている。評価関数の係数は、蓄積するデータのキーワードの度数分布によって決定する。

また、その最適化の手法を実際のデータベースの作成に適用し、その効果の程度も示している。

第5章では、システム（UNIQ-1）のディスク入出力に対して、ディスクキャッシュ・システム（以下、DCSという）を適用した場合の効果について考察している。

最初に、一般的なディスク入出力におけるDCSの性能評価を試みている。それによると、一般のディスク入出力における入力時間は、DCSを使用しない場合と比較して最大7.5倍速くなる。また、DCSの効果は、シーケンシャル・ファイルではディスク装置とディスクキャッシュ間のデータの転送単位であるブロックのサイズにのみ依存し、ランダム・ファイルではディスクキャッシュのサイズにのみ依存することが判明した。また、DCSを使わない場合のディスク装置に対する論理入出力レコードのサイズとその入出力時間の関係、およびDCSを使用したときの論理入力レコードのサイズとその入力時間の関係も明確になった。

次に、以上の性能評価の結果をもとに、DCSをUNIQ-1に適用した場合の効果について考察している。その結果、データの蓄積ではその入出力時間は最大約3.3倍速くなり、検索では最大約7.5倍速くなることが明らかになった。

第6章では、システムの利用分析を行っている。利用分析は、UNIQ-1を使って1980年7月より検索サービスを行っている金属学関係の文献データベースMETADEXの利用記録をもとに行っている。それによると

(1) METADEXのような文献データの検索では、使用回数およびデータの出力件数とも、デー

タの経年数と逆比例的に減少する。

(2) 検索の結果の出力では、出力するデータの内容および出力量によって、出力する形式や出力する装置を自由に選択できることが必要である。

(3) 定期的にデータが更新されるデータベースでは、自動的に検索できるようにする機能が必要である。

(4) 多くのデータのキーワードになっているインデックス・タームほど検索のキーワードとしても多く使われる。

などが明確になった。そして、(4)の事実は、今後のデータベースの設計に対して一つの指針を与えるものである。

また、本システムでは、使用したコマンドと質問式の利用記録が利用者の入力したままのイメージでとれるようにしてあり、これらを分析することによってシステムの利用教育に大いに役立てることができる。

第7章は、結言である。

以上、各章の概要で述べたように、オンライン情報蓄積・検索システムの基本設計ができ、実際にUNIQ-1として実現され、実用に供されていることは、本研究の最終目的までが達成されたことを示している。

そして、本論文第2章で示したように、オンライン情報蓄積・検索システムの基本構成を構造モデルと機能モデルの2層モデルからなるとして、概念設計、論理設計と組み立てる設計方式は、オンライン情報蓄積・検索システムの設計方式を確立するものである。また、ここで定義された論理コマンドは、高水準プログラミング言語のサブルーチン・コールにマッピングでき、実現方式の設計を容易にするものである。

第4章で述べたシステムの最適化の手法は、データの特성에応じてシステムの構成パラメータの最適値を求めるもので、この種の情報蓄積・検索システムにおいて有力な最適化の手法であり、第5章で述べたディスクキャッシュ・システムの利用は、この種の情報蓄積・検索システムの効率化を図るための有力な手段である。

更に、第6章におけるシステムの利用分析の結果は、情報蓄積・検索システムの設計や利用教育に対して有用な情報を提供するものである。

## 審査結果の要旨

今後の情報化社会において、一つの重要な情報資源はデータベースである。このため夫々の目的に従って自分の望む情報をデータベースとして構築し、この情報を希望する同じ分野の人々に自由に提供できるデータベースシステムの設計問題はきわめて重要な問題である。本論文の著者はこの立場に立ち、特に研究者を対象とした新しい研究者用データベースシステムの設計法を導いた。本設計法に従えば研究者は、自由に自分の目的のためのデータベースシステムを構築することができる。本論文はこれらの成果をまとめたもので全編7章よりなる。

第1章は序論である。

第2章では、情報蓄積・検索システムで扱うデータをモデル化し、システムのプロトタイプ設計法と、これにより導かれるシステムの論理設計法について論じている。

第3章では、第2章で導いた論理システムを任意の計算機システム上で実現する方法について研究し、特に本システムを東北大学大型計算機センターのACOSシステム1000で実現するための具体的設計法を与えている。本システムはオンライン情報蓄積・検索システムUNIQ-1として広く研究者により利用されており、有用な成果といえる。

第4章では、UNIQ-1で実現したシステムについて解析し、蓄積する情報のもつデータ構造とシステム性能との関係を明らかにしている。本解析によりシステムの性能を最適にする設計法が導かれ、この有効性は実システムによる実験により確かめられている。

第5章では、システムの各構成要素がどのようにシステムの性能に影響するかを調べ、それらの間の関係を明らかにしている。特にディスクキャッシュの大きさがシステムの性能に大きく影響することを示し、ディスクキャッシュの設計法を与えている。

第6章では、UNIQ-1で構築したデータベースMETADEXの利用記録を分析し、UNIQ-1のコマンド及び検索のための質問式の利用度を考察している。これにより今後効率的なコマンド設計の基礎データが与えられる。

第7章は結論である。

以上要するに本論文は、研究者を対象とした新しいオンライン情報蓄積・検索システム設計のための基礎理論、具体的構成論を明らかにしたもので、計算機工学、情報工学の発展に寄与するところが少なくない。

よって、本論文は工学博士の学位論文として合格と認める。