(19) 대한민국특허청(KR)
(12) 등록특허공보(B1)

(51) Int. Cl.[6]
G06F 17/30

(45) 공고일자 2000년01월15일
(11) 등록번호 10-0240243
(24) 등록일자 1999년10월26일

(21) 출원번호 10-1997-0004166
(22) 출원일자 1997년02월12일
(65) 공개번호 특1997-0066948
(43) 공개일자 1997년10월13일
(30) 우선권주장 96-54588 1996년03월12일 일본(JP)

(73) 특허권자 마쯔시다덴기산교 가부시키가이샤 모리시타 요이찌
일본국 오사카후 가도마시 오아자 가도마 1006반지
(72) 발명자 기노시타 데쓰야
일본국 사이타마켄 우라와시 기시마찌 7-2-9-401
오야마 다카마사
일본국 가나가와켄 가마쿠라시 다이 3-6-15-612
기쿠치 츄이치
일본국 지바켄 이치가와시 히라타 4-11-8
(74) 대리인 김기종, 권동용, 최재철

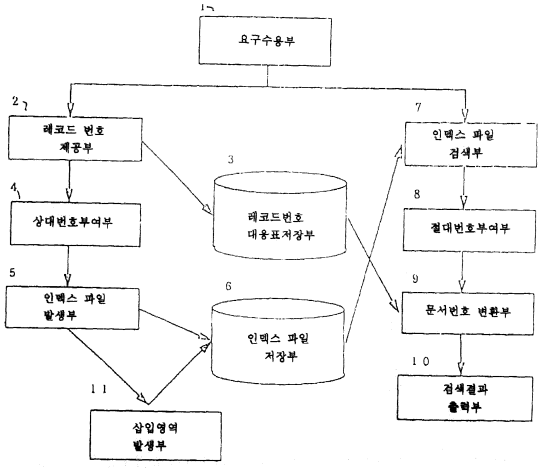
심사관 : 엄인권

(54) 데이터 검색장치

요약

검색 요구와 등록요구를 수용하는 요구수용 수단과; 상기 등록요구에 응답하여 상기 레코드 번호 데이터를 유일하게 규정하는 법칙에 의해 상기 데이터에 포함된 색션 데이터에 따라 등록될 데이터 단위인 각각의 레코드에 레코드 번호를 제공하는 레코드 번호 제공수단과; 상기 레코드 번호와 하나 이상의 스페어 레코드가 그룹중 하나의 그룹을 형성하도록 상기 레코드를 그룹화하고 그룹화한 번호로부터의 그룹번호 데이터가 대응 상대 번호 데이터를 발생시키는 상대 번호 부여수단과; 상기 상대 레코드 번호 데이터와 상기 그룹 번호 데이터를 이용하여 상기 데이터로부터 인덱스 파일을 발생시키는 인덱스 파일 발생수단과; 상기 인덱스 파일을 저장하는 인덱스 파일 저장수단과; 상기 요구 수용수단으로부터의 상기 검색요구에 응답하여 상기 인덱스 파일 저장수단의 상기 인덱스 파일에 따라 상기 검색 요구에 의해 표시된 레코드를 검색하는 인덱스 파일 저장수단과; 상대 번호 데이터와 하나의 검색된 레코드의 그룹 번호 데어터를 상기 레코드 번호 데이터로 반환하는 절대 번호 부여수단과; 상기 절대 번호 부여수단에 따라 검색결과를 출력시키는 검색결과 출력수단을 구비한 데이터 검색장치를 제공하는 것이다.

대표도



명세서

[발명의 명칭]

데이터 검색장치

[도면의 간단한 설명]

제1도는 본 발명의 제3실시예에 이용된 본 발명의 제1실시예의 데이터 검색장치의 블록도.

제2도는 제2및 제3실시예에 언급된 제1실시예의 흐름도.

제3도는 기록된 문서의 상태를 도시한 제1실시예의 예시도.

제4도는 대응표의 상태를 도시한 제1실시예의 예시도.

제5도는 절대 레코드 번호와 상태번호 사이의 관계를 도시한 제1실시예의 예시도.

제6도는 인덱스 파일을 도시한 제1실시예의 예시도.

제7a도는 문자열과 출현도수 사이의 관계를 도시한 예시도.

제7b도는 두 개의 소정의 문자를 포함하는 문자체인과 출현도 정보수 사이에 관계를 도시한 예시도.

제8도는 인덱스 파일을 도시한 실시예의 예시도.

제9도 및 제10도는 제1실시예의 검색처리의 흐름도.

제11도는 제2실시예의 데이터 검색장치의 블록도.

제12도는 문서등록작동을 도시한 제2실시예의 흐름도.

제13도는 문서의 레코드 번호를 도시한 제2실시예의 예시도.

제14도는 인덱스 파일 발생처리를 도시한 제3실시예의 흐름도.

제15도는 인덱스 파일을 도시한 제3실시예의 예시도.

제16도는 인덱스 파일을 발생하는 데이터를 도시한 제3실시예의 예시도.

제17도는 선행기술의 데이터 검색장치에 의해 이용된 인덱스 파일의 데이터 포맷의 예시도.

* 도면의 주요부분에 대한 부호의 설명

1 : 요구수용부 2 : 레코드번호 제공부

3 : 레코드번호 대응표 저장부 4 : 상대 번호 부여부

5 : 인덱스 파일 발생부 6 : 인덱스 파일 저장부

7 : 인덱스 파일 검색부 8 : 절대번호 부여부

9 : 문서번호 변환부 10 : 검색결과 출력부

[발명의 상세한 설명]

[발명의 목적]

[발명이 속하는 기술분야및 그분야의 종래기술]

본 발명은 요구에 응답하여 데이터를 검색하기 위해 데이터를 제공하는 데이터 검색장치에 관한 것이다.

등록문서를 관리하기 위해 데이터 베이스에 기록될 문서를 레코드로 분할하고 이 레코드에 번호를 부여하여 레코드 번호 데이터를 저장하는 데이터 검색장치가 공지되어 있고 각각의 레코드는 등록검색을 위해 데이터 베이스에 저장된 데이터 단위이다.

이러한 종래의 데이터 검색장치는 소정의 메모리 영역에 레코드번호와 문자출현 데이터를 저장한다.

제17도는 선행기술의 데이터 검색장치에 의해 이용되는 인덱스 파일의 데이터 포맷의 예시도이다.

제17도에서, 인데스 파일로 32비트의 소정의 메모리 영역은 레코드 번호 데이터의 12 비트와 문자출현데이터의 12 비트를 저장한다. 따라서, 인덱스 파일의 이 포맷은 1048576(2^{20})레코드만의 관리를 제공한다는 문제점이 있다. 또한, 선행기술에서, 인덱스 파일의 크기(비트)는 레코드의 번호가 증가하는 경우 증가해야 한다는 문제점이 있다.

또한, 선행기술의 데이터 검색장치에서, 문서번호, 페이지번호 및 레코드번호 사이의 대응을 나타내는 대응표를 제공해야 한다. 따라서, 새로운 대응표가 필요하고 문서갱신으로 인한 페이지 수가 증가하는 경우 대응표를 분류해야 한다는 문제점이 있다.

본 발명의 목적은 개량된 데이터 검색장치를 제공하는 것이다.

본 발명에 따라, 제1데이터 검색장치는 검색 요구 데이터를 포함하는 검색요구와 등록 요구를 수용하는 입력부; 이 데이터에 포함된 색션 데이터에 따라 등록할 데이터를 레코드의 단위에 색션시키기 위해 입력부로 부터의 등록 요구에 응답하는 색션부; 모든 소정의 번호가 하나이 그룹으로 글룹하기 위해 레코드를 그룹화하여 번호를 부여하고 그룹번호 데이터와 번호가 부여되어 그룹화된 레코드로 인한 상대 레코드 번호 데이터를 발생시키는 번호 부여부; 이에따라 그룹 번호 데이터와 상대적인 레코드 번호 데이터를 저장하는 저장부; 입력부로 부터의 검색요구 데이터 및 저장부의 그룹 데이터 및 상대 레코드 번호 데이터에 따라 레코드중하나를 검색하는데 이용 될 검색 데이터를 발생시켜 출력시키는 검색 데이터 발생부를 구비하고 있다.

본 발명에 따라, 제2데이터 검색장치는 검색 데이터를 포함하는 검색 요구와 등록 요구를 수용하는 요구

수용부; 등록 용구에 응답하여 레코드 번호 데이터를 등록될 데이터의 단위인 각 레코드에 제공하는 레코드 번호 제공부; 레코드 번호데이터를 그룹으로 그룹시키고 레코드 번호내의 그룹 번호 데이터와 상대 레코드번호 데이터로 변환하는 상대 번호 부여부; 이 데이터 및 상대 레코드 번호 데이터로부터 인덱스 파일 발생하는 인덱스 파일 발생부; 이 인덱스 파일을 저장하는 인덱스 파일 저장부; 요구수용부로 부터의 검색 요구에 응답하여 인덱스 파일 저장부의 인덱스 파일에 따라 검색 데이터에 의해 표시된 레코드중 하나를 검색하는 인덱스 파일 검색부; 레코드중 검색된 레코드의 상대 번호 데이터 및 그룹 번호 데이터를 레코드 번호 데이터로 변환하는 절대 번호 부여부; 절대번호 부여부로 부터의 레코드 번호 데이터에 따라 검색결과를 출력시키는 검색결과 출력부를 구비하고 있다.

제2데이터 검색장치는 데이터 파일 저장부가 인덱스 파일을 저장할 때 인데스 파일을 장치 인덱스 파일 저장부에 저장하기 위해 저장공간을 인덱스 파일 저장부에 제공하는 삽입 영역 발생부를 구비하고 있다.

제2데이터 검색장치에서, 레코드 번호 제공부는 레코드 번호 데이터와 이 데이터에 포함된 색션 데이터 사이의 대응을 나타내는 레코드 번호 대응표를 더 포함하고 검색결과 출력부는 레코드 번호 대응표에 따라 절대 번호부여부로 부터의 레코드 번호 데이터로부터의 검색결과를 출력시킨다.

제2데이터 검색장치에서, 인덱스 파일 발생부는 각각의 그룹 번호 데이터를 저장하는 헤더와 상대 레코드 번호 데이터 및 대응 인덱스 데이터를 저장하기 위해 헤더를 추종하는 인덱스 데이터 영역을 포함하는 인덱스 파일을 발생시킨다.

본 발명에 따라, 제3데이터 검색장치는 검색요구와 등록요구를수용하는 요구수용부와; 등록요구에 응답하여 레코드 번호 데이터를 유일하게 규정하는 법칙을 통해 데이터에 속한 색션데이터에 따라 등록된 데이터의 단위인 각각의 레코드에 레코드 번호를 제공하는 레코드 번호 제공부; 레코드 번호와 하나 이상의 스페어 레코드가 그룹중 하나를 형성하도록 레코드를 그룹화하고, 그룹번호 데이터와 그룹화한 레코드 번호로부터의 대응상태 번호 데이터를 발생하는 상대 번호부여부와; 요구수용부로부터의 검색요구에 응답하여 인덱스 파일 저장부의 인덱스 파일에 따라 검색요구에 의해 표시된 레코드 중 하나를 검색하는 인덱스 파일 검색부와; 레코드중 검색된 레코드의 상대 번호 데이터와 그룹 번호 데이터를 기록번호 데이터로 변환하는 절대번호 부여부와; 이 절대번호 부여부로부터의 레코드 번호에 따라, 검색결과를 출력시키는 검색결과 출력부를 구비한다. 제3검색장치는 인덱스 파일 저장부가 인덱스 파일을 저장할 때 인덱스 데이터를 장치, 인덱스 파일저장부에 저장하기 위해 저장공간을 인덱스 저장부에 제공하는 삽입영역 발생부를 더 포함한다.

본 발명의 제1실시예를 아래에서 설명할 것이다.

제1도는 제3실시예에서 언급한 본 발명의 제1실시예의 데이터 검색 장치의 블록도이다.

데이터 검색장치는 각 처리의 요구를 수용하는 요구수용부(1), 등록해야할 문서에 레코드 번호(데이터)를 할당하고 문서번호, 페이지번호 및 레코드번호 사이의 대응을 나타내는 기록번호 대응표를 발생시키는 레코드 번호 제공부(2), 레코드번호 제공부(2)에 의해 발생된 레코드 번호 대응표를 저장하는 레코드 번호 대응표(3), 레코드 번호 제공부(2)에 의해 제공된 레코드 번호(데이터)를 그룹화하고 레코드 번호를 그룹내의 상대번호를 변환시키는 상대번호 부여부(4)와; 상대번호 부여부(4)에 의해 제공된 상대번호와 등록된 데이터로부터 데이터 파일을 발생시키는 인덱스 파일 발생부(5)와; 이 인덱스 파일 발생부(5)에 의해 발생된 인덱스 파일을 저장하는 인덱스 파일 저장부(6)와; 인덱스 파일 저장부(6)의 인덱스 파일을 사용하여 검색을 실행하는 인덱스 파일 검색부(7)와; 인덱스 파일 검색부(7)에 의해 검색된 상대 레코드 번호를 절대 레코드 번호로 변환하는 절대 번호 부여부(8)과; 레코드 번호 대응표 저장부(3)의 레코드 번호 대응표를 이용하여 문서번호와 페이지번호로 절대 번호부여부에 의해 검색된 레코드 번호로 변환하는 문서번호 변환부(9)와; 검색된 문서번호를 디스플레이하는 검색결과 출력부(10)와; 인덱스 파일 발생부(5)가 인덱스 파일을 인덱스 파일 저장부(6)에 저장하는 동안 앞서서 부가적인 데이터를 저장하기 위해 데이터 영역을 제거하는 삽입 영역 발생부(11)를 포함하고 있다. 제1실시예에서, 삽입영역 발생부(11)는 이용되지 않는다.

제1실시예에의 작동을 설명할 것이다. 제2도는 제2및 제3실시예에서 언급한 제1실시예의 흐름도를 도시한다. 제3도는 레코드된 문서의 상태를 도시한 실시예를 나타낸다. 제4도는 대응표의 상태를 도시한 실시예를 도시한다. 데이터 검색장치에 포함된 cpu(데이터 검색장치에 포함되었지만 도시되지 않은 중아처리장치)는 제2도의 흐름도에 도시되어 있듯이 프로그램을 실행한다.

사용자가 문서를 등록할 때, 문서등록 요구를 수용하고 이에 응답하여 레코드 번호 제공부(2)는 단계(s1)에서 등록된 데이터의 단위인 각각의 레코드에 레코드 번호를 제공한다. 즉, 등록될 데이터인 문서는 페이지를 포함하는 모든 페이지는 레코드 번호가 제공된 레코드로 처리된다.

예를 들어, 문서가 제3도에 도시되어 있듯이 등록되면, 문서(D1)의 1 페이지에는 레코드번호 1에 제공되고, 문서(D1)의 2 페이지에는 레코드 번호 2가 제공되고, 문서(D2)의 1페이지에는 레코드 번호 3에 제공된다.

cpu는 레코드번호, 문서번호 및 페이지 번호 사이의 관계를 나타내는 레코드 번호 대응표에 대해 테이블 데이터를 발생시킨다. cpu는 테이블 데이터를 사용하여 레코드 번호 대응표를 갱신하고 단계(s2)에서 대응표 저장부(3)에서 갱신된 레코드 번호 대응표를 저장한다.

제3도에 도시된 문서가 등록될 때, 레코드 번호 대응표의 데이터 상태가 제4도에 도시된 것처럼 된다. cpu는 단계(s3)에서 레코드를 그룹화한다. 65536레코드가 하나의 그룹으로 그룹화된다고 하면 그룹번호는 그룹 번호 = 크기 로드번호/65536 으로 주어진다. 단, 은 몫 즉, 그룹 번호가 레코드번호/65536의 분수를 제거함으로써 제공된 정수이다.

다음 단계(s4)에서 그룹의 상대 레코드 번호가 상대 레코드 번호 = 레코드번호 %/65536 이고 단 오퍼레이터 %는 나머지를 나타낸다.

절대 레코드 번호와 상대 번호 사이의 관계를 나타내는 실시예를 도시한 제5도에 도시되어 있듯이, 절대

레코드 번호가 상대 번호로 변환된다. 예를들면, 절대 레코드 번호 65537 그룹 1에서 상대 레코드 번호로 변환 되고 그룹 2에서 레코드 번호 13072 는 상대 레코드 번호 0으로 변환된다.

다음, cpu는 단계(s5)에서 인덱스 파일을 발생시킨다. 제6도는 인덱스 파일을 도시한 실시예를 예시한다. 인덱스 파일은 두 개의 문자를 포함하는 체인에서의 두 개의 문자의 출현도 정보수의 번호를 포함한다.

제7a도는 문자의 트레인과 출련시간의 번호사이의 관계를 도시한다.

제7b도는 두 개의 소정의 문자를 포함하는 문자열과 출현도 정보수의 번호 사이의 관계를 도시한다. 등록된 데이터가 "abaabcac"이면, 각각의 문자의 출현 시간의 번호가 제7a도에 도시되어 있다. 다음, 문자 체인이 중첩을 지닌 등록된 데이터의 상부로부터 연속적으로 제공되도록 등록된 데이터가 분할된다. 즉,문자열이 두 개의 문자마다 색션되고 이색션점이 하나의 문자마다 시프트된다.

다음, 문자열이(a, b)이고 이의 출현도 정보수의 번호가(1, 1)이다. 다음 문자열이(b, a)이고 출현도 정보수의 번호가(1, 2)이다.

이와 유사하게, cpu는 모든 문자열의 출현도 정보수의 번호를 얻는다. 제8도는 인덱스 파일을 도시한 실시예의 예시이다.. 그룹 1에서 레코드 번호에 등록된 데이터가 "ababcbcbcababc"이면 그룹 1에서의 레코드번호(2)에 등록된 "bbaabbcabc" 이고 그룹 2의 레코드 번호 1에서 등록된 데이터가 "bcababcbc" 이고 그룹(2)의 레코드 번호(2)에 등록된 데이터가 "acbcabbaabc 이다. cpu는 제8도에 도시되어 있듯이, 문자열(a, b) 및 (b, c)에 대한 출현시간의 번호를 산출한다. 최종 레코드가 인덱스 파일을 제공하도록 처리될때까지 cpu는 인덱스 파일을 발생하기 위해 단계(s1-s7)을 반복 실행한다. 즉 cpu는 최종 레코드가 단계(s6)에서 처리되었는지를 결정한다. 만일 레코드가 최종이 아니면, 처리기 단계(s1)에 복귀한다.

검색 처리를 설명할 것이다.

제9도와 제10도는 본 실시예의 검색처리의 흐름도를 도시한다.

사용자가 검색될 문자열(검색문자열)의 데이터을 포함하는 검색요구를 입력할 때, 요구수용부(1)는 검색요구를 수용하고 검색문자의 데이터를 인덱스 파일부(7)에 공급한다. 인덱스 파일 검색부(7)는 문자열을 문자체인으로 분할한다.

검색 문자열이 abc 이면 인덱스 파일 검색부(7)는 단계 (s11)에서 이를 문자체인(a, b) 및 (b, c)로 분할한다. 인덱스 파일 검색부(7)는 단계(s12)에서 인덱스 파일 저장부(6)의 각각의 파일에 있어서의 문자체인(a, b) 및 문자체인 (b, c)의 출현도 정보수 파일로부터 문자체인(a, b) 및 (b, c)의 출현도 정보수를 얻는다.

제8도의 경우, 그룹 1의 레코드 번호 1에서 문자체인(a, b)의 출현도 정보수의 번호는 (1, 1)이고, 문자체인(b, c)의 출현도 정보수의 번호는(1, 1)이고, 문자체인(b, c)의 출현도 정보수의 번호는(2, 1)이다. 따라서, 인덱스 파일 검색부(7)는 인데스 파일 저장부(6)로부터 문자체인(a, b)의 출현도 정보수(1, 1)의 번호와 문자체인(b, c)의 출현도 정보수(2, 1)의 번호를 얻는다.

다음, 인덱스 파일 검색부(7)는 단계(s13)에서 인덱스 파일 저장부(6)의 문자체인(a, b) 및 (b, c)의 출현도 정보수의 얻어진 번호의 그룹번호들을 비교한다. 출현도 정보수의 얻어진 번호가 같지 않으면, 인덱스 파일 검색부(7)는 그룹번호가 작은 그룹에 존재하는 문자체인(a, b) 및 (b, c)의 인덱스 파일의 포인터중 하나를 다음 그룹의 출현도 정보수의 번호의 최고치까지 증가시키는 단계(s13) 및 (s14)를 되풀히 한다. 문자체인(a, b) 및 (b, c)의 출현시간, 데이터의 얻어진 번호의 그룹번호가 같으면, 인덱스 파일 검색부(7)는 단계(15)에서 출현도 정보수의 얻어진 번호의 레코드 번호를 비교한다.

제8도의 경우에, 모든 그룹 번호가 1이기 때문에, 인덱스 파일 검색부(7)는 레코드 번호를 비교한다.

상대 레코드 번호가 동일한 그룹에서 서로 다르면, 인덱스 파일 검색부(7)는 출현도 정보수의 얻어진 번호가 단계(s16)에서 그룹의 최종 데이터 인지를 결정한다. 만일 출현도 정보수의 얻어진 번호가 그룹에서 최종 데이터가 아니면, 인덱스 파일 검색부(7)는 단계(s17)에서 레코드 번호가 적은 출현도 정보수의 후행번호를 얻는다. 인덱스 파일 검색부(7)는 단계(s15)~ 단계(s17)의 처리를 되풀이 한다. 출현도 정보수의 얻어진 번호가 그룹의 최종 데이터이면, 인덱스 파일 검색부(7)는 단계(s18)에서 후행그룹의 상부에서 출현도 정보수의 번호를 얻고, 단계(s13)~단계 (s18)의 처리를 되풀이 한다. 그룹의 레코드 번호가 같으면, 인덱스 파일 검색부(7)는 단계(s19)에서 출현시간의 번호들을 비교한다.

제8도의 경우에, 그룹의 양 레코드 번호가 1 이기 때문에, 인덱스 파일 검색부(7)는 출현도 정보수의 번호들을 비교한다.

인덱스 파일 검색부(7)는 다음과 같이 출현도 정보수의 번호를 비교한다.

문자체인(a, b)에서의 문자"b" 의 출현 시간의 비교는 문자체인(b,c)에서의 문자 "b"의 출현도 정보수의 번호와 비교된다. 출현도 정보수의 양번호가 서로 일치하면, 문자체인(a, b)의 문자 "b"가 문자체인(b, c)의문자 "b"와 같은 문자이다. 이것은 두 개의 문자체인 연속 문자열, 즉, 문자열(abc)라는 것을 의미한다.

출현시간의 양번호가 서로 일치하지 않으면, 인덱스 파일 검색부(7)는 단계(s10)에서 출현도 정보수의 번호가 작은 문자체인에 대해 추종하는 문자체인이 출현도 정보수의 번호를 얻고, 단계(s5)로부터 처리를 되풀이 한다.

제8도의 경우에, (a, b)의 출현도 정보수의 번호가 (1, 1)이고 (b, c)이 출현도 정보수가 (2, 1)이어서 문자체인(a, b)의 제2문자(b)의 출현시간의 번호가 문자체인(b, c)(COM1)의 제1문자(b)의 출현시간 번호와 일치하지 않기 때문에, 인덱스 파일 검색부(7)는 동일한 그룹(2, 2)의 레코드 번호(1)의 출현도 정보수의 번호를 얻고 값이 작은 출현시간의 번호를 갖는 문자체인(a, b)에 대해 추종한다.

다음, 출현도 정보수의 양번호가 동일한 레코드 번호 및 동일한 그룹에 있기때문에 인덱스 파일 검색부

(7)는 양출현 주파수(COM2)를 비교한다. 출현도 정보수의 번호가 2 이기 때문에 문자(abc)가 검색되어진다.

출현도 정보수가 서로 일치하면, 즉, 문자열에 검색되면, 레코드 번호가 단계(s11)에서 다음식에 의해 그룹번호 및 레코드 번호로부터 산출된다.

레코드 번호 = 65536 * 그룹번호 + 그룹의 레코드 번호.

다음, 인덱스 파일 검색부(7)는 단계(s12)에서 레코드 번호 대응표 저장부(3)에 저장된 제4도에 도시된 대응표를 이용하여 레코드 번호로부터 문서번호와 페이지 번호를 얻는다.

다음 단계(s13)에서 인덱스 파일 검색부(7)는 검색될 문서번호와 페이지 번호로 단계(s12)에서 얻어진 문서번호와 페이지번호를 출력시킨다. 최종 데이터가 처리될때까지 이 처리가 되풀이 된다.

인덱스 파일 발생부(5)는 각각의 그룹번호 데이터를 저장하는 헤더(60)와 상대 레코드 번호 데이터를 저장하여 인덱스 파일에 대응시키기 위해 헤더(60)에 대해 추종하는 인덱스 데이터 영역(62)를 포함하는 인덱스 파일을 발생시킨다. 헤더(60)과 각각의 인덱스데이터 영역(62)는 크기가 동일하다. 언급했듯이 본 실시예의 데이터 검색장치에서, 문서가 등록될 때, 레코드 번호가 모든 페이지에 할당되고, 이 레코드 번호는 그룹화되어 제어될 상대 레코드 번호로 변환되어 동일한 범위의 레코드 번호가 모든 그룹에 되풀이하여 이용될 수 있다.

본 실시예에서, 인덱스 파일이 두 개의 문자를 포함하는 문자체인의 출현도 정보수에 의해 제공된다. 그러나, 문제차인의 절대 위치 데이터와 같은 문자체인의 출현을 나타내는 기타 데이터가 인덱스 파일에 대해 이용될 수 있다. 또한, 본 실시예에서, 등록의 하나의 데이터가 모든 페이지에 제공된다. 그러나, 하나의 레코드를 기간 등과 같은 데이터를 색션함으로써 색션된 모든 문서, 페이지 또는 모든 항(패러그래프)와 같은 모든 소정의 색션에 제공할 수 있다.

제2실시예를 설명할 것이다.

제11도는 제2실시예에 데이터 검색장치의 블록도이다.

제2실시예의 데이터 검색장치는 여러처리를 위해 요구를 수용하는 요구수용부(12), 문서번호 및 페이지 번호로부터 레코드 번호를 유일하게 결정하는 법칙에 따라 레코드 번호를 제공하는 레코드 번호 제공부(13), 이 레코드 번호 제공부(13)에 의해 제공된 레코드 번호를 그룹화 하는 상대 번호 부여부(14), 등록될 상대 번호 부여부(14) 데이터에 의해 변환된 상대 레코드 번호를 포함하는 인덱스 파일을 발생시키는 인덱스 파일 발생부(15), 인덱스 파일 발생부(15)에 의해 발생된 인덱스 파일을 저장하는 인덱스 파일 저장부(16), 인덱스 파일 저장부(16)에 저장된 인덱스 파일을 이용하여 데이터를 검색하는 인덱스 파일 검색부(17), 인덱스 파일 검색부(17)에 의해 검색된 레코드의 상대 레코드 번호를 절대번호로 변환하는 절대번호부여부(18), 레코드 번호 제공부(13)에 의해 실행된 변환이 역변환을 통해 레코드 번호로부터 문서번호와 페이지 번호를 제공하는 문서번호 변환부(19) 및 검색결과로 문서 번호를 출력시키는 검색결과 출력부(20)를 포함한다.

먼저, 문서를 등록하는 작동을 설명할 것이다. 제12도는 문서등록 작동을 도시한 제2실시예의 흐름도를 도시한다. 제13도는 문서의 레코드 번호를 도시한 제2실시예의 예시이다.

사용자가 등록요구를 요구수용부(12)에 입력할 때, 용구수용부(12)는 이동록요구를 수용한다. 다음, 레코드 번호 부여부(13)는 문서번호 및 페이지 번호로부터의 레코드 번호를 유일하게 결정하는 법칙에 의해 레코드 번호를 제공한다.

레코드 번호는 다음식으로부터 산출된다. 레코드 번호 = (문서번호 -1) * 하나의 문서의 페이지의 최대번호 + 페이지번호, 단, 레코드 번호는 제13도에 도시되어 있듯이, 하나의 문서의 페이지 번호를 소정의 번호에 설정함으로써 유일하게 결정될 수 있다.

이처리후 문서를 등록하는 처리는 제2도에 도시된 단계(s3 ~ s8)의 처리와 유사하다.

검색작동은 제19도 및 제10도에 도시된 처리에 의해 수용된다. 그러나,문서 번호와 페이지 번호는 다음식에 의해 단계(s12)에서 레코드 번호로부터 산출된다.

문서번호 P 레코드 번호/하나의 문서의 페이지의 최대수 + 1.

페이지 번호 = 레코드 번호 % 하나의 문서의 최대 페이지 번호.

언급했듯이, 제2실시예의 데이터 검색 장치에서 많은량의 레코드 번호가 레코드 번호를 그룹화하고, 상대 레코드 번호로 제어함으로써 제어될 수 있어서 하나의 문서의 최대 페이지 번호로부터의 레코드 번호 및 페이지 번호를 유일하게 결정하는 법칙에 의해 페이지 번호를 할당할 수 있다.

따라서, 제1실시예에서 필요한 대응표를 발생할 필요가 없다. 또한, 문서를 갱신함으로써 페이지 번호가 증가하면, 레코드 번호의 번호 부여가 소정의 최소 페이지내에 규칙적으로 제공될 수 있다.

또한, 레코드 번호의 기타 변환이 레코드 번호 = 문서의 최대번호 *(페이지수 -1) + 문서번호와 같은 문서번호 및 페이지 번호로부터의 레코드 번호를 유일하게 결정하면 본 실시예에 응용가능하다.

제3실시예를 설명할 것이다.

제3실시예의 구조와 작동은 제1실시예와 유사하다. 제1실시예와 제3실시예의 차이전은 삽입 영역 발생부(11)가 더 제공되었다는 것이다. 삽입영역 발생부(11)는 삽입영역은 발생시키고 출현도 정보수를 이 삽입영역에 저장하는 반면, 인덱스 파일 발생부(5)는 인덱스 파일을 인덱스 파일 저장부(6)에 저장한다.

문서등록작동은 제2도에 도시된 단계 1~4의 처리와 유사하게 실행된다.

제14도는 인덱스파일 발생 처리를 도시한 제3실시예의 흐름도를 도시한다.

단계(s41)에서, 인덱스 파일 발생부는 제1실시예와 유사한 것으로 두 개의 문자를 포함하는 문자 체인의 출현도 정보수를 발생시킨다. 인덱스 파일 발생부(5)는 단계(s42)에서 출현도 데이터수를 저장하기 위해 삽입 영역이 없는지를 결정한다. 출현도 정보의 수를 저장하는 삽입 영역에 없는 경우 삽입영역 발생부(11)는 단계(s43)에서 저장영역이 소정의 크기를 발생한다. 단계(s42)에서 출현도 정보수를 저장하기 위한 삽입영역이 있는 경우, 인덱스 파일 발생부(5)는 단계(s44)에서 출현도 정보수를 삽입영역에 저장시킨다.

제15도는 인덱스 파일을 도시한 제3실시예의 예시이다. 제16도는 인덱스 파일을 발생하기 위해 데이터를 도시한 제3실시예의 예시이다. 그룹1에서 레코드 번호 1을 지닌 문장을 "ababcbcababc"이고, 그룹 1에서 레코드 번호(2)를 갖는 문장이 "bbaabbcabc" 이고, 그룹2에서 레코드 번호 2를 갖는 문장이 "bcababcbc" 라고 했다. 다음, 인덱스 파일 발생부(5)는 제1실시예와 유사하게 문자체인(a, b)의 출현도 정보의수를 나타내는 인덱스 파일을 발생시키고, 소정의 크기의 또 다른 인덱스 영역이 제16도에 도시되어 있듯이 모든 그룹에 끼워진다.

출현도 정보수를 인덱스 파일에 더하는 경우는 문서가 더 등록되고 등록된 문서가 갱신되어 출현도 정보수가 갱신되어야 한다.

이처리후 데이터 등록 처리는 데이터 등록 처리를 완료하기 위해 제2도에 도시되어 있듯이 제1실시예에 단계(s6~s7)의 처리와 유사하게 실행된다.

검색 작동은 제9도 및 제10도에 도시된 처리와 유사하게 실행된다.

언급했듯이 제3실시예의 데이터 검색 장치에서 소정의 크기의 삽입 영역을 발생시킨다. 따라서, 등록된 문서가 갱신되거나 새로운 문서가 더 갱신되면, 삽입영역에 부가된 출현도 정보수를 부가할 수 있어서 데이터 삽입으로 인한 등록 속도를 감소시키는 것을 방지한다.

본 실시예에서, 레코드 번호가 상대적으로 번호를 부여함으로써 제공된다. 그러나, 상대적으로 번호가 부여되지 않은 일반 레코드 번호를 지닌 레코드 번호를 제어할 수 있다.

(57) 청구의 범위

청구항 1

검색 요구 데이터를 포함하는 검색요구와 등록요구를 수용하는 입력수단과; 상기입력수단으로 부터의 상기 등록요구에 응답하여 상기 데이터에 포함된 색션 데이터에 따라 레코드의 단위로 등록될 데이터를 색션하는 색션수단과; 상기 레코드의 모든 소정의 번호가 하나의 그룹으로 그룹화 되도록 상기 레코드를 그룹화하여 번호를 부여하고 번호가 부여되어 그룹화한 레코드로 인한 그룹번호 데이터와 상대 레코드 번호 데이터를 발생시키는 번호 부여수단과; 상기 그룹번호 데이터와 이로 인한 상기 상대 레코드 번호 데이터를 저장하는 저장수단과; 상기 입력수단으로 부터의 상기 검색요구 데이터와 상기 저장수단의 상대 레코드 번호 데이터에 따라 하나의 상기 레코드를 검색하기 위해 사용되는 검색 데이터를 발생하여 출력시키는 검색 데이터 발생수단을 구비한 데이터 검색장치.

청구항 2

검색데이터를 포함하는 검색요구 및 등록 요구를 수용하는 요구 수용수단과; 상기 등록 요구에 응답하여 등록될 데이터의 단위인 각각의 레코드에 레코드번호 데이터를 제공하는 레코드 번호 제공수단과; 상기 레코드 번호 데이터를 그룹으로 그룹화하고 상기 레코드 번호 데이터를 각각의 상기 그룹내의 그룹 번호 데이터 및 상대 레코드 번호 데이터는 변환하는 상대 번호 부여수단과; 상기 데이터와 상기 상대 레코드 번호 데이터로부터 인덱스 파일을 반생시키는 인덱스 파일 발생수단과; 상기 인덱스 파일을 저장하는 인덱스 파일 저장수단과; 상기 요구 수용수단으로 부터의 상기 검색요구에 응답하여 상기 인덱스 파일저장수단의 상기 인덱스 파일에 따라 상기 검색 데이터에 의해 표시된 하나의 레코드를 검색하는 인덱스 파일 검색수단과; 상기 번호 데이터와 상기 검색된 하나의 레코드의 그룹번호 데이터를 상기 레코드 번호 데이터를 변환하는 절대번호 부여수단과; 상기 절대번호 부여수단으로부터의 레코드 번호 데이터에 따라 검색결과를 출력시키는 검색결과 출력수단을 구비한 데이터 검색장치.

청구항 3

제2항에 있어서, 상기 레코드 번호 제공부는 상기 레코드 번호 데이터와 상기 데이터에 포함된 색션 데이터 사이의 대응을 나타내는 대응표에 레코드 번호를 발생시키고, 상기 검색결과 출력수단을 상기 레코드 번호 대응표에 따라 절대번호부여수단으로부터의 상기 레코드 번호 데이터로부터 상기 검색결과를 출력시키는 데이터 검색장치.

청구항 4

제2항에 있어서, 상기 인덱스 파일 발생수단은 각각의 상기 그룹번호 데이터를 저장하기 위한 헤더와 상대 레코드 번호 데이터를 저장하기 위해 상기 헤더에 대해 추종하고 상기 인덱스 데이터에 대응하는 상기 인덱스 데이터 영역을 포함하는 인덱스 파일을 발생시키고, 상기 헤더와 각각의 상기 인덱스 데이터 영역은 크기가 같은 데이터 검색장치.

청구항 5

검색요구와 등록요구를 수용하는 요구수용 수단과; 상기 등록요구에 응답하여 상기 레코드 번호 데이터를 유일하게 규정하는 법칙에 의해 상기 데이터에 포함된 색션 데이터에 따라 등록될 데이터 단위인 각각의 레코드에 레코드 번호를 제공하는 레코드 번호 제공수단과; 상기 레코드 번호와 하나 이상의 스페어 레코드가 그룹중 하나의 그룹을 형성하도록 상기 레코드를 그룹화하고 그룹화한 번호로부터의 그룹번호 데이터가 대응 상대 번호 데이터를 발생시키는 상대 번호 부여수단과; 상기 상대 레코드 번호 데이터와 상기 그룹 번호 데이터를 이용하여 상기 데이터로부터 인덱스 파일을 발생시키는 인덱스 파일 발생수단과; 상

기 인덱스 파일을 저장하는 인덱스 파일 저장수단과; 상기 요구수용수단으로부터의 상기 검색요구에 응답하여 상기 인덱스 파일 저장수단의 상기 인덱스 파일에 따라 상기 검색 요구에 의해 표시된 레코드를 검색하는 인덱스 파일 검색수단과; 상대 번호 데이터와 하나의 검색된 레코드의 그룹 번호 데이터를 상기 레코드 번호 데이터로 반환하는 절대 번호 부여수단과; 상기 절대 번호 부여수단에 따라 검색결과를 출력시키는 검색결과 출력수단을구비한 데이터 검색장치.

청구항 6

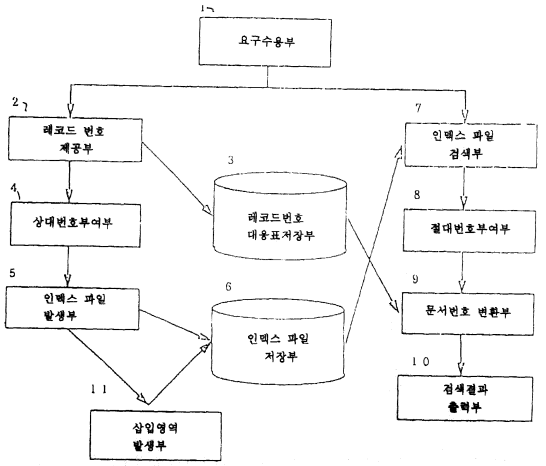
제2항에 있어서, 상기 인데스 파일 저장수단이 상기 인덱스 파일을 저장할때 장치 상기 인덱스 데이터를 상기 인덱스 파일 저장수단에 저장하기 위해 상기 저장 공간을 상기 인덱스 저장수단에 제공하는 삽입영역 발생수단을 구비한 데이터 검색장치.

청구항 7

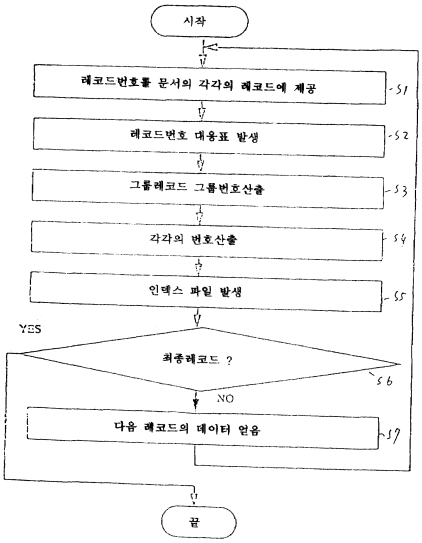
제5항에 있어서, 상기 인덱스 파일 저장수단의 상기 인덱스 파일을 저장할때 장치 상기 인덱스 파일 저장수단의 저장공간을 상기 인덱스 파일 저장수단의 상기 인덱스 데이터를 제공하는 삽입영역 발생부를 더 구비한 데이터 검색장치.

도면

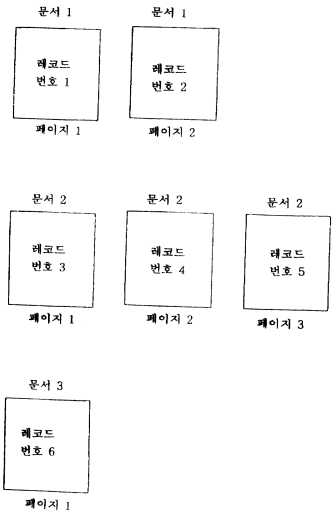
도면1



도면2



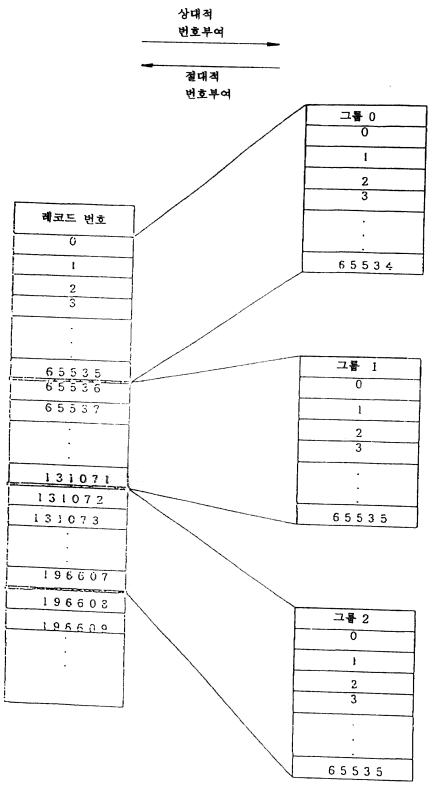
도면3



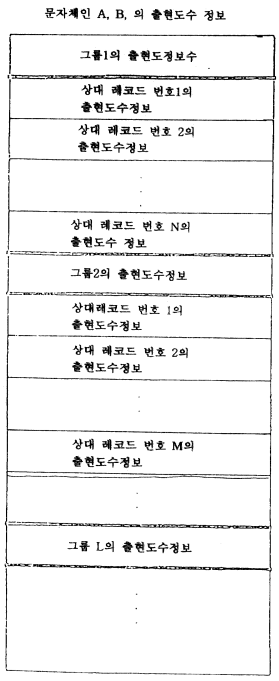
도면4

레코드 번호	문서 번호	페이지 번호
1	1	1
2	1	2
3	2	1
4	2	2
5	2	3
6	3	1
	. . .	

도면5


상대적
번호부여
절대적
번호부여
레코드 번호
0
1
2
3
65535
65536
65537
131071
131072
131073
196607
196608
196609
그룹 0
0
1
2
3
65534
그룹 1
0
1
2
3
65535
그룹 2
0
1
2
3
65535

도면6


문자제인 A, B, 의 출현도수 정보
그룹1의 출현도정보수
상대 레코드 번호1의
출현도수정보
상대 레코드 번호 2의
출현도수정보
상대 레코드 번호 N의
출현도수 정보
그룹2의 출현도수정보
상대레코드 번호 1의
출현도수정보
상대 레코드 번호 2의
출현도수정보
상대 레코드 번호 M의
출현도수정보
그룹 L의 출현도수정보

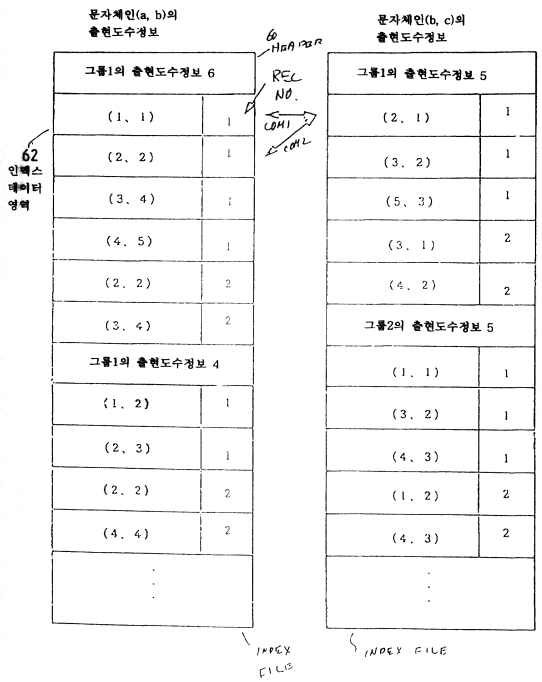
도면7a

문자열	a	b	a	a	b	c	a	c
출현도수	1	1	2	3	2	1	3	2

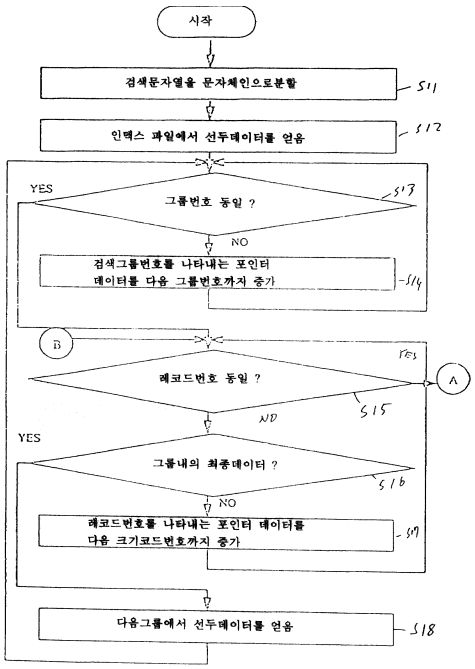
도면7b

문자열	출현도수정보
(a, a)	(2, 3)
(a, b)	(1, 1) (3, 2)
(a, c)	(3, 2)
	.
(b, a)	(1, 2)
(b, c)	(2, 1)
	.
(c, a)	(1, 3)
	.

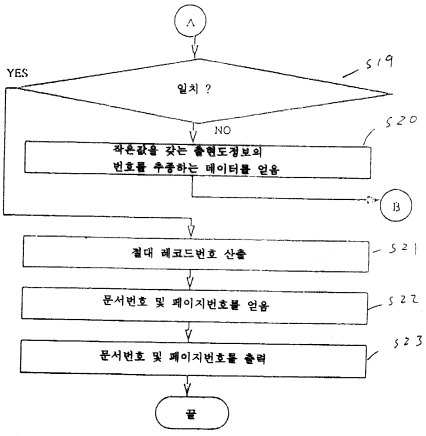
도면8



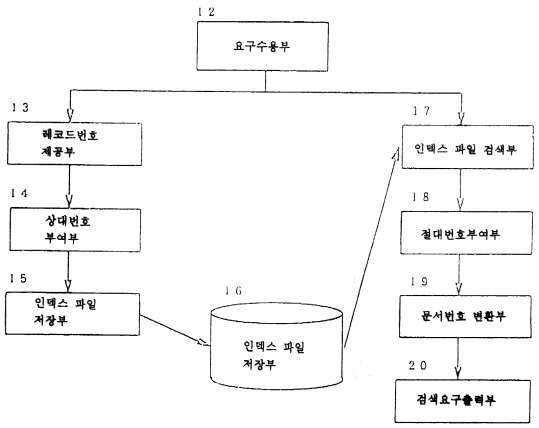
도면9



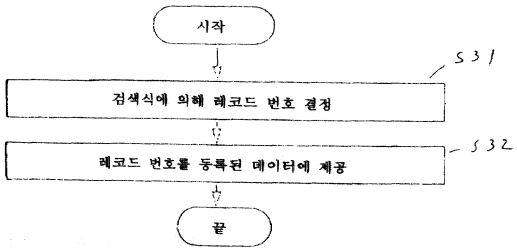
도면10



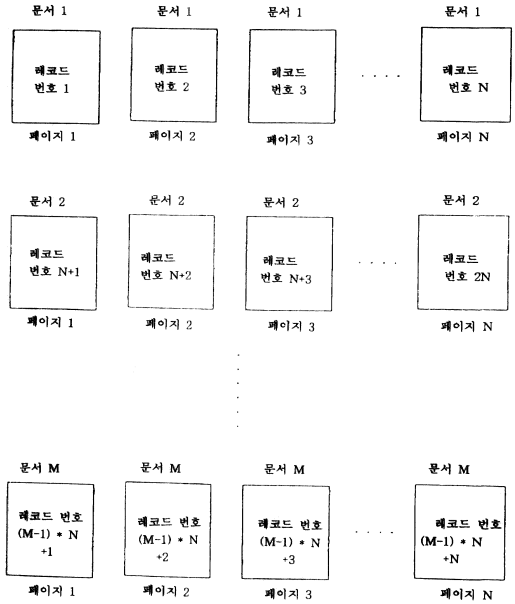
도면11



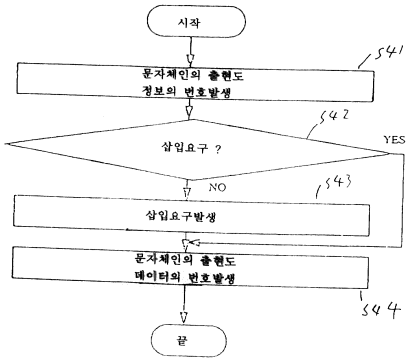
도면12



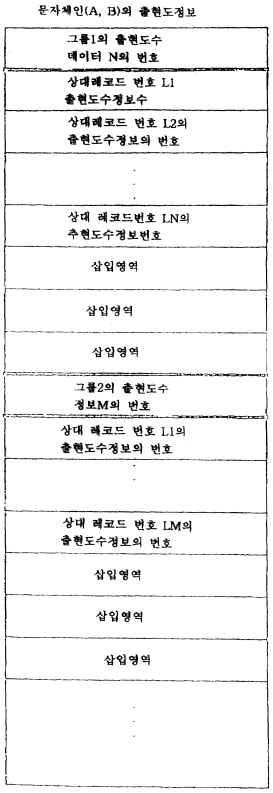
도면13



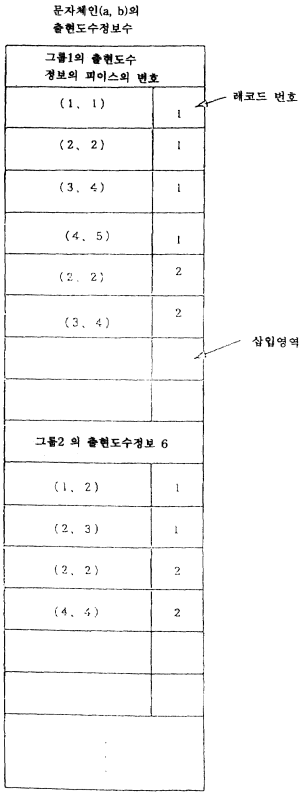
도면14



도면15


문자체인(A, B)의 출현도정보
그룹1의 출현도수
데이터 N의 번호
상대레코드 번호 L1
출현도수정보수
상대레코드 번호 L2의
출현도수정보의 번호
상대 레코드번호 LN의
추현도수정보번호
삽입영역
삽입영역
삽입영역
그룹2의 출현도수
정보M의 번호
상대 레코드 번호 L1의
출현도수정보의 번호
상대 레코드 번호 LM의
출현도수정보의 번호
삽입영역
삽입영역
삽입영역

도면16


문자체인(a, b)의
출현도수정보수
그룹1의 출현도수
정보의 피어스의 번호
(1, 1) 1
(2, 2) 1
(3, 4) 1
(4, 5) 1
(2, 2) 2
(3, 4) 2
레코드 번호
삽입영역
그룹2 의 출현도수정보 6
(1, 2) 1
(2, 3) 1
(2, 2) 2
(4, 4) 2

도면17

선행기술

