

## 대용량 전경화상자료기지구축을 위한 화일색인화

리금수, 한남철, 강춘성

경애하는 최고령도자 김정은동지께서는 다음과 같이 말씀하시였다.

《새 세기 산업혁명의 불길높이 우리 나라를 지식경제강국으로 일떠세워야 합니다.》

도로를 따라가면서 촬영한 전경화상자료는 위성화상자료에 비해볼 때 훨씬 더 큰 대용량자료이다. Oracle과 같은 자료기지관리체계는 전경화상자료와 같은 대용량자료를 저장하는데 시간소비가 많고 비용이 많이 들며 공간검색이 불가능하다.

논문에서는 화일색인화와 그에 기초한 검색방법을 리용하여 대용량 전경화상자료의 저장과 관리를 진행하는 방법에 대하여 서술하였다.

### 1. 전경화상자료관리를 위한 화일색인화

화일색인화는 전경화상자료편성을 위한 방법중의 하나이다. 화일색인화는 색인구조에 기초하여 등록부관리에서 불필요한 자료쓰기동작뿐아니라 실행시 기억기사용을 줄인다.

논문에서는 전경화상자료기지를 하쉬등록부블록과 ID블록, 자료블록으로 설계하였다.(그림 1) 하쉬등록부블록은 등록부화일들의 하쉬등록부표를 포함하고 ID블록들은 화일의 ID와 등록부를 위한 등록부항목을 포함하며 자료블록들은 전경화상자료들로 구성된 자료화일로 이루어진다. 세가지 류형들가운데서 하쉬등록부와 ID블록들은 자료화일을 위한 메타자료를 포함한다.

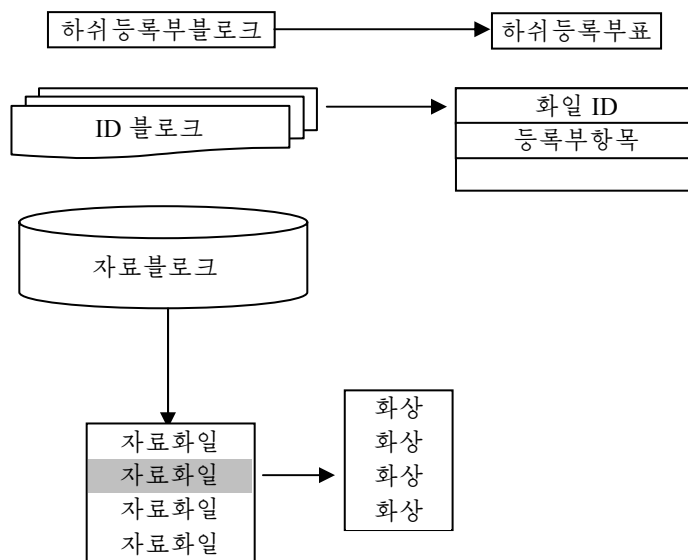


그림 1. 색인화일을 위한 블록의 분류와 배치

구체적인 색인화일구성과정은 다음과 같다.

① 등록부관리를 위한 코드화

자료등록부의 코드는 전경화상자료의 참조자리표를 제공한다. 매 전경화상의 지리자리표들은 코드와 지리자리표사이 변환에 의하여 얻어진다.

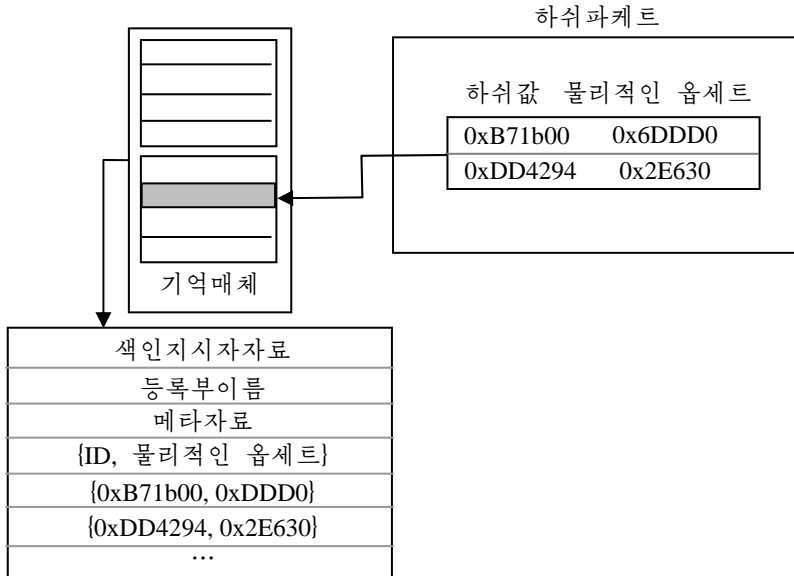


그림 2. 하쉬에 기초한 등록부구조

등록부는 연산속도를 높이기 위하여 하쉬함수와 하쉬자료구조로 관리된다. 하쉬등록부분록들은 1개의 쌍을 이루는 ID(하쉬값)와 그것의 물리적위치를 포함한다.(그림 2)

② 색인화일구성

전경화상자료등록부의 코드화는 등록부의 색인으로 볼수 있다. 전경화상자료는 많은 자료등록부들로 이루어지며 매 등록부에서 전경화상자료화일은 순서대로 하나씩 저장된다. 전경화상자료화일은 색인정보를 가진 색인화일들에 대응된다.

색인화일은 순서로 된 매 전경화상자료화일에서의 저장된 자료위치만을 기록하며 추가적인 정보를 포함할수 있다. 매 자료화일들의 최대크기는 2GB로 하며 2 000개의 전경화상들은 하나의 화일로 작성된다. 화일들이 많아지면 화일검색속도가 떨어지기때문에 화일들은 여러개의 등록부들로 나누어 저장하며 1개의 등록부는 100개의 화일들을 포함한다. 매 화일들은 1MB정도의 용량을 가진다.

## 2. 색인화일에 기초한 전경화상자료의 검색

전경화상자료검색에는 색인화일에 의한 검색과 지리적위치에 따르는 검색이 있다.

색인화일에 의한 검색에서는 색인화일로부터 전경화상자료가 들어있는 레코드의 위치를 검색하여 자료화일을 얻는다. 이를 위하여 색인화일에서 해당한 레코드의 위치를 얻고 자료화일에서 그것의 레코드마당들을 읽어 자료를 얻는다.

색인화일로부터의 레코드검색과정은 다음과 같다.

① 코드로부터 색인화일의 등록부이름과 화일이름을 얻는다.

② 색인화일에서 색인정보를 알아낸다.

③ 색인정보에 있는 자료화일의 레코드위치를 리용하여 레코드값을 얻는다.

전경화상자료검색에서 색인화일에 의한 검색과 함께 지리적위치에 의한 검색도 매우 중요한 역할을 한다.

다분해능화상자료기지의 구성은 기초적인 원리들인 자료구조화와 공간색인알고리즘을 요구한다.[1] 화상이나 동화상과 같은 자료들의 공간검색은 측정공간과 수학적인 방법에 기초한다.[2]

전경화상자료의 공간검색은 분산된 주어진 어떤 위치에서의 전경화상의 검색이며 그 과정은 다음과 같다.

① 부분영역의 수를 얻는다.

부분영역에 드는 자료의 개수는 일정한 한계를 넘지 말아야 한다. 분할하는 살창수는 다음과 같이 계산된다.

$$N = \sqrt{(n/500)}$$

여기서  $N$ 은 분할살창의 행수(혹은 렬수),  $n$ 은 전체 전경화상자료의 개수이다.

전경화상자료에 대한 분할살창의 행수는 200개, 매 부분영역의 전경화상자료의 수는 500개이하로 정하였다.

② 부분영역의 전경정보들을 얻는다.

부분영역들의 전경화상자료들에는 지리자리표와 코드가 들어있다. 공간색인화일의 1개의 기록은 3개의 마당들을 가지며 그 크기는 20byte이다.(표 1)

표 1. 전경공간속성마당

속성이름	자료형	설명
ID	Integer	전경화상의 식별번호
photoX	Double	전경촬영기의 X자리표
photoY	Double	전경촬영기의 Y자리표
angle	Float	앞방향 촬영방위각
pitch	Float	앞방향 촬영경사각
date	DateTime	촬영날자 및 시간
addr	String	촬영지역의 주소
st_name	String	촬영한 도로이름
cnt	Integer	린접한 전경화상의 개수
Spot[cnt]	Object Array	린접한 전경화상정보배렬

photoX와 photoY마당들은 지리자리표 ( $X$ ,  $Y$ )를 가진 점에 의하여 표시되고 ID마당은 전경화상의 영역코드를 표시한다. 매 기록들은 photoX와 photoY의 증가순서로 공간색인화일에 저장된다.

③ 부분영역의 번호를 얻는다.

어떤 점을 포함하는 부분영역의 행렬번호는 다음과 같이 계산된다.

$$I = [(X - X_{\min})(X_{\max} - X_{\min}) \times N]$$

$$J = [(Y - Y_{\min})(Y_{\max} - Y_{\min}) \times N]$$

여기서  $I$ 와  $J$ 는 부분영역에서 점의 행과 렬번호,  $X_{\min}$ ,  $X_{\max}$ 는 전체 영역의 최소 및 최대  $X$ 자리표값들,  $Y_{\min}$ ,  $Y_{\max}$ 는 전체 영역의 최소 및 최대  $Y$ 자리표값들이다.

④ 주어진 부분영역의 공간색인화일에 포함된 전경화상자료들의 정보를 리용하여 전

경 화상자료를 검색한다.

공간색인화일에 포함된 자료들을 리용하여 검색하려는 점에서 제일 가까운 점의 코드 마당값을 얻는다. 주어진 코드마당값으로 색인화일에 의한 전경화상자료에 대한 정보를 검색하여 실지 전경화상자료를 얻었다.

Oracle자료기지를 가진 공개형체계와 공간색인화일을 리용하는 전경지리공간정보체계의 공간검색속도를 비교한 결과는 표 2와 같다.

표 2. 공간검색속도를 비교한 결과

검색자료수/개	Oracle을 리용한 공간검색/ms	공간색인을 리용한 공간검색/ms
100	21	11
200	43	22
500	89	60
2 000	387	245
5 000	1 302	520
10 000	2 130	1 320

## 맺 는 말

화일색인에 기초한 대용량 자료기지에서의 공간검색속도는 다른 자료기지관리체계의 검색속도보다 빠르다. 이 방법을 리용하면 대용량 전경화상자료들에 대한 효과적인 검색과 관리를 진행할수 있다.

## 참 고 문 헌

- [1] M. Wang et al.; Geo-Spatial Information Science, 3, 3, 52, 2000.
- [2] M. Welington et al.; Data Science Journal, 10, 23, 143, 2012.

주체109(2020)년 1월 5일 원고접수

## File Indexing for a Huge Amount of Panoramic Image Database Construction

*Ri Kum Su, Han Nam Chol and Kang Chun Song*

In this paper we presented a method for management and storage of a huge amount of panoramic imagery data using file indexing and its search.

Keywords: file index, panoramic image, retrieval method, hash directory