(19) 대한민국특허청(KR)
(12) 등록특허공보(B1)

(45) 공고일자 2009년02월10일
(11) 등록번호 10-0882864
(24) 등록일자 2009년02월03일

(51) Int. Cl.
G06F 17/30 (2006.01) *G06F 17/00* (2006.01)
G06F 9/00 (2006.01)
(21) 출원번호 10-2007-0120759
(22) 출원일자 2007년11월26일
심사청구일자 2007년11월26일
(56) 선행기술조사문헌
KR1019980071452 A
JP12227921 A

(73) 특허권자
한국전자통신연구원
대전 유성구 가정동 161번지
(72) 발명자
지형근
대전 유성구 관평동 운암네오미아 아파트 603동 1001호
홍도원
대전 유성구 전민동 엑스포아파트 212-1704
(74) 대리인
한양특허법인

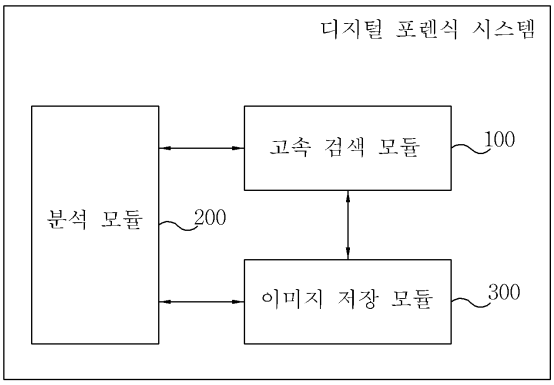
전체 청구항 수 : 총 10 항 심사관 : 김상철

(54) 디지털 포렌식 시스템을 위한 대용량 데이터 고속 검색시스템 및 방법

(57) 요 약

디지털 포렌식 시스템을 위한 대용량 데이터 고속 검색 시스템 및 방법이 개시된다. 디지털 포렌식 시스템을 위한 대용량 데이터 고속 검색 방법은, 이미지 저장 모듈이 검색하고자 하는 디스크 이미지를 입력받는 단계; 분석 모듈에 의해 이미지 저장 모듈로부터 입력된 디스크 이미지를 분석하여 디스크 이미지상에 존재하는 파일 목록을 구성하는 단계; 고속 검색 모듈에 의해 이미지 저장 모듈로부터 입력된 디스크 이미지에 대해 파일별로 클러스터를 재배열하는 단계; 고속 검색 모듈에 의해 텍스트 정보를 가지고 있는 파일로부터 텍스트 정보를 추출하여 저장하는 단계; 및 고속 검색 모듈에 의해 비트단위 검색 기법에 의해 키워드 및 통상적인 표현을 검색하는 단계를 포함한다.

대 표 도 - 도2



이 발명을 지원한 국가연구개발사업

과제고유번호	2007-S-019-01
부처명	정보통신부 및 정보통신연구진흥원
연구사업명	IT성장동력기술개발
연구과제명	정보투명성 보장형 디지털 포렌식 시스템 개발
주관기관	한국전자통신연구원
연구기간	2007-03-01 ~ 2008-02-28

특허청구의 범위

청구항 1

디지털 포렌식 시스템을 위한 대용량 데이터 고속 검색 시스템에 있어서,

조사하고자 하는 디스크의 디스크 이미지를 저장하는 이미지 저장 모듈;

상기 이미지 저장 모듈로부터 입력된 디스크 이미지를 분석하여 디스크 내 파일들이 저장된 클러스터를 분석하는 분석 모듈; 및

상기 이미지 저장 모듈로부터 디스크 이미지를 입력받아 키워드를 검색하여 그 검색 결과를 제공하는 고속 검색 모듈을 포함하고,

상기 고속 검색 모듈은 입력된 디스크 이미지에 대해 파일별로 클러스터를 재배열하고, 텍스트 정보를 가지고 있는 파일로부터 텍스트 정보를 추출해서 저장하여, 비트단위 검색 기법에 의해 검색하는 것을 특징으로 하는 디지털 포렌식 시스템을 위한 대용량 데이터 고속 검색 시스템.

청구항 2

제 1 항에 있어서,

상기 고속 검색 모듈은,

패턴 매칭 보드를 사용하여 원하는 다중 키워드를 동시에 검색하는 것을 특징으로 하는 디지털 포렌식 시스템을 위한 대용량 데이터 고속 검색 시스템.

청구항 3

제 1 항에 있어서,

상기 고속 검색 모듈은,

패턴 매칭 보드를 이용하여 디스크 이미지 상의 모든 섹터와 텍스트로 변환된 파일들에서 키워드 및 통상적인 표현을 검색하는 것을 특징으로 하는 디지털 포렌식 시스템을 위한 대용량 데이터 고속 검색 시스템.

청구항 4

제 1 항에 있어서,

상기 이미지 저장 모듈은,

상기 고속 검색 모듈에서 변환된 텍스트 파일을 생성한 후, 변환된 텍스트 파일을 해당 디스크 이미지와 함께 저장하는 것을 특징으로 하는 디지털 포렌식 시스템을 위한 대용량 데이터 고속 검색 시스템.

청구항 5

제 1 항에 있어서,

상기 고속 검색 모듈은,

각 파일의 클러스터가 순서대로 인접하여 배치되도록 클러스터를 재배열하는 것을 특징으로 하는 디지털 포렌식 시스템을 위한 대용량 데이터 고속 검색 시스템.

청구항 6

디지털 포렌식 시스템을 위한 대용량 데이터 고속 검색 방법에 있어서,

(a) 이미지 저장 모듈이 검색하고자 하는 디스크 이미지를 입력받는 단계;

(b) 분석 모듈에 의해 상기 이미지 저장 모듈로부터 입력된 디스크 이미지를 분석하여 디스크 이미지상에 존재하는 파일 목록을 구성하는 단계;

(c) 고속 검색 모듈에 의해 상기 이미지 저장 모듈로부터 입력된 디스크 이미지에 대해 파일별로 클러스터를 재

배열하는 단계;

(d) 상기 고속 검색 모듈에 의해 텍스트 정보를 가지고 있는 파일로부터 텍스트 정보를 추출하여 저장하는 단계; 및

(e) 상기 고속 검색 모듈에 의해 비트단위 검색 기법에 의해 키워드를 검색하는 단계;

를 포함하는 디지털 포렌식 시스템을 위한 대용량 데이터 고속 검색 방법.

청구항 7

제 6 항에 있어서,

상기 단계 (b)는,

(b1) 상기 입력된 디스크 이미지를 분석하여 사용하는 파일 시스템을 파악하는 단계; 및

(b2) 디스크 이미지상에 존재하는 파일 목록을 구성하는 단계;

를 포함하는 디지털 포렌식 시스템을 위한 대용량 데이터 고속 검색 방법.

청구항 8

제 6 항에 있어서,

상기 단계 (c)는,

각 파일의 클러스터가 순서대로 인접하여 배치되도록 클러스터를 재배열하는 것을 특징으로 하는 디지털 포렌식 시스템을 위한 대용량 데이터 고속 검색 방법.

청구항 9

제 6 항에 있어서,

상기 단계 (d)는,

(d1) 텍스트 정보를 가지고 있는 파일로부터 각 문서의 포맷에 해당하는 파서를 이용해서 텍스트 정보를 추출하는 단계;

(d2) 추출된 텍스트 정보를 상기 이미지 저장 모듈에 의해 해당 디스크 이미지와 함께 저장하는 단계;

를 포함하는 디지털 포렌식 시스템을 위한 대용량 데이터 고속 검색 방법.

청구항 10

제 6 항에 있어서,

상기 단계 (e)는,

비트단위(bitwise) 검색 기법에 의해 패턴 매칭 보드를 사용하여 원하는 다중 키워드를 동시에 검색하는 것을 특징으로 하는 디지털 포렌식 시스템을 위한 대용량 데이터 고속 검색 방법.

명 세 서

발명의 상세한 설명

기 술 분 야

<1> 본 발명은 대용량 데이터 고속 검색 시스템 및 방법에 관한 것으로, 특히 디지털 증거(Digital Evidance)를 분석하기 위한 디지털 포렌식 시스템에서 대용량의 디스크 이미지로부터 파일 시스템을 구성하여 파일별로 클러스터를 재배열하고, 디스크 이미지 내의 텍스트 정보를 가지고 있는 파일(포맷이 있는 파일)들을 텍스트 파일로 변환한 후, 패턴 매칭 보드를 이용하여 비트단위(bitwise) 검색에 의해 특정 키워드나 통상적인 표현(regular expression)을 빠르고 정확하게 검색하는, 디지털 포렌식 시스템을 위한 대용량 데이터 고속 검색 시스템 및 방법에 관한 것이다.

<2> 본 발명은 정보통신부 및 정보통신연구진흥원의 IT 성장동력기술개발 사업의 일환으로 수행한 연구로부터 도출된 것이다[과제관리번호: 2007-S-019-01, 과제명:정보투명성 보장형 디지털 포렌식 시스템 개발(Development of Digital Forensic System for Information Transparency)].

<3>

배 경 기 술

<4> 컴퓨터 포렌식(computer forensic)은 컴퓨터 시스템에서 자료를 수집하고 분석하여 분석된 자료에 대한 보고서를 작성하는 일련의 과정을 말한다. 컴퓨터 포렌식은 범죄 수사에 있어서 여러 가지 증거 자료들이 범죄자의 컴퓨터 시스템 또는 이와 관련된 다양한 저장 장치로부터 발견됨에 따라 관심이 집중되고 있는 분야이다.

<5> 컴퓨터 포렌식은 원하는 데이터를 찾기 위한 검색의 반복이라고 해도 과언이 아닐 만큼 검색이 많이 이루어지지만 저장매체의 용량이 급격이 증가함에 따라 관련 증거를 검색하는데 수일 이상이 걸리는 경우가 많아 조사에 어려움을 가지고 있다. 일반적으로, 컴퓨터 포렌식을 위한 검색 방법은 인덱스기반(index-based) 검색 방법과 비트단위(bitwise) 검색 방법이 있다.

<6> 인덱스기반(index-based) 검색 방법은 파일 기반 검색 방법으로써 디스크 상의 모든 파일들에 포함되어 있는 모든 단어에 대하여 사전에 인덱스를 생성하고 검색하는 방법이다. 이 방법의 장점은 초기 인덱싱 후에는 실시간 검색이 가능하고, HWP, PDF 등의 다양한 파일 포맷에 대하여 검색이 가능하다는 점이다. 그러나, 인덱스기반 검색 방법은 초기 인덱싱 과정에 많은 시간이 걸리며, 논리적 파일 단위로 검색하기 때문에 슬랙(slack) 영역이나 비할당 영역에 있는 데이터를 검색할 수가 없어 디지털 포렌식 시스템에 사용하는데 큰 약점을 가지고 있다.

<7> 도 1은 종래의 인덱스를 이용한 정보 검색 방법을 설명한 흐름도이다.

<8> 인덱스를 이용한 정보 검색 방법은 디스크 등에 저장된 대량의 문서를 고속으로 검색하기 위해 인덱스를 생성하여(S10) 데이타베이스로 갱신하고(S11), 인덱스 파일(index file)을 생성하고(S12), 검색할 문자열(search character string)을 검색 엔진(Search Engine)으로 입력하여(S13) 검색엔진에 의해 인덱스 파일을 이용하여 지정된 문자열과 문자의 나열이 유사한 문자열을 포함하는 문서를 고속으로 검색하여(S14) 그 검색 결과를 디스플레이한다(S15).

<9> 검색 시스템의 색인파일은 문자연쇄파일, 위치정보파일, 확장문자연쇄파일, 확장위치정보파일로 구성된다. 문자연쇄파일에는 가변길이 연쇄, 고정길이 연쇄, 단락패턴과 그것에 대응하는 문서번호, 문서내 위치번호가 위치정보파일의 어디에 위치하는가가 저장되며, 위치정보파일에는 문서번호, 문서내 위치번호가 저장된다. 확장문자연쇄파일에는 확장문자연쇄와 그것에 대응하는 가변길이 연쇄번호, 가변길이 연쇄내 위치번호가 확장위치 정보파일의 어디에 위치하는가가 저장되며, 확장위치정보파일에는 가변길이 연쇄번호, 가변길이 연쇄내 위치번호가 저장된다. 이 색인 파일들을 사용하여, 지정된 문자열과 문자의 나열이 유사한 문자열을 포함하는 문서를 고속으로 검색한다.

<10> 비트단위(bitwise) 검색 방법은 디스크의 처음부터 끝까지 모든 비트를 검색하는 방법이다. 이 방법의 장점은 슬랙 영역이나 비할당 영역에 존재하는 데이터를 검색할 수 있고, 키워드뿐만 아니라 복잡한 통상적인 표현(regular expression)을 이용한 검색도 가능하며, 파일 헤더와 같이 테스트가 아닌 이진 데이터(binary data)도 검색 가능하다.

<11> 그러나, 비트단위 검색 방법은 MS Office, 한글워드(HWP), PDF 파일 등과 같이 아스키(ASCII) 포맷으로 저장되어 있지 않는 파일에 대하여는 기본적으로 검색이 불가능하며, 디스크의 모든 비트를 검색하므로 검색 시간이 많이 걸리는 단점이 있다. 또한, 비트단위 검색 방법은 하나의 파일을 여러 클러스터(cluster)에 나누어 저장되어 있고, 이 클러스터들이 서로 인접해 있지 않은 경우, 및 찾고자 하는 키워드가 두 클러스터 경계에 걸쳐 있는 경우에는 검색이 안되는 단점이 있었다.

발명의 내용

해결 하고자하는 과제

<12> 본 발명은 상기한 문제점을 해결하기 위해 제안된 것으로, 디지털 증거(Digital Evidence)를 분석하기 위한 디지털 포렌식 시스템에서 대용량의 디스크 이미지에서 파일별로 클러스터를 재배열하고, 디스크 이미지 내의 텍스트 정보를 가지고 있는 파일(포맷이 있는 파일)들을 텍스트 파일로 변환한 후, 패턴 매칭 보드를 이용하여 비

트단위(bitwise) 검색에 의해 대용량의 저장매체에서 특정 키워드나 통상적인 표현(regular expression)을 빠르고 정확하게 검색하는, 디지털 포렌식 시스템을 위한 대용량 데이터 고속 검색 시스템 및 방법을 제공하는데 그 목적이 있다.

과제 해결수단

<13> 본 발명의 목적을 달성하기 위하여, 본 발명은 디지털 포렌식 시스템을 위한 대용량 데이터 고속 검색 시스템으로서, 조사하고자 하는 디스크의 디스크 이미지를 저장하는 이미지 저장 모듈; 상기 이미지 저장 모듈로부터 입력된 디스크 이미지를 분석하여 디스크 내 파일들이 저장된 클러스터를 분석하는 분석 모듈; 및 상기 이미지 저장 모듈로부터 디스크 이미지를 입력받아 키워드를 검색하여 그 검색 결과를 제공하는 고속 검색 모듈을 포함하고, 상기 고속 검색 모듈은 입력된 디스크 이미지에 대해 파일별로 클러스터를 재배열하고, 텍스트 정보를 가지고 있는 파일로부터 텍스트 정보를 추출해서 저장하여, 비트단위 검색 기법에 의해 검색하는 것을 특징으로 한다.

<14> 상기 고속 검색 모듈은, 패턴 매칭 보드를 사용하여 원하는 다중 키워드를 동시에 검색하는 것을 특징으로 한다.

<15> 상기 고속 검색 모듈은, 패턴 매칭 보드를 이용하여 디스크 이미지 상의 모든 섹터와 텍스트로 변환된 파일들에서 키워드 및 통상적인 표현을 검색하는 것을 특징으로 한다.

<16> 상기 이미지 저장 모듈은, 상기 고속 검색 모듈에서 변환된 텍스트 파일을 생성한 후, 변환된 텍스트 파일을 해당 디스크 이미지와 함께 저장하는 것을 특징으로 한다.

<17> 상기 고속 검색 모듈은, 각 파일의 클러스터가 순서대로 인접하여 배치되도록 클러스터를 재배열하는 것을 특징으로 한다.

<18> 본 발명의 다른 목적을 달성하기 위하여, 본 발명은 디지털 포렌식 시스템을 위한 대용량 데이터 고속 검색 방법으로서, (a) 이미지 저장 모듈이 검색하고자 하는 디스크 이미지를 입력받는 단계; (b) 분석 모듈에 의해 상기 이미지 저장 모듈로부터 입력된 디스크 이미지를 분석하여 디스크 이미지상에 존재하는 파일 목록을 구성하는 단계; (c) 고속 검색 모듈에 의해 상기 이미지 저장 모듈로부터 입력된 디스크 이미지에 대해 파일별로 클러스터를 재배열하는 단계; (d) 상기 고속 검색 모듈에 의해 텍스트 정보를 가지고 있는 파일로부터 텍스트 정보를 추출하여 저장하는 단계; 및 (e) 상기 고속 검색 모듈에 의해 비트단위 검색 기법에 의해 키워드를 검색하는 단계를 포함한다.

<19> 상기 단계 (b)는, (b1) 상기 입력된 디스크 이미지를 분석하여 사용하는 파일 시스템을 파악하는 단계; 및 (b2) 디스크 이미지상에 존재하는 파일 목록을 구성하는 단계를 포함한다.

<20> 상기 단계 (c)는, 각 파일의 클러스터가 순서대로 인접하여 배치되도록 클러스터를 재배열하는 것을 특징으로 한다.

<21> 상기 단계 (d)는, (d1) 텍스트 정보를 가지고 있는 파일로부터 각 문서의 포맷에 해당하는 파서를 이용해서 텍스트 정보를 추출하는 단계; (d2) 추출된 텍스트 정보를 상기 이미지 저장 모듈에 의해 해당 디스크 이미지와 함께 저장하는 단계를 포함한다.

<22> 상기 단계 (e)는, 비트단위(bitwise) 검색 기법에 의해 패턴 매칭 보드를 사용하여 원하는 다중 키워드를 동시에 검색하는 것을 특징으로 한다.

효 과

<23> 이상에서 설명한 바와 같이, 본 발명은 디지털 포렌식 시스템에서 대용량의 디스크 이미지에서 파일 시스템을 구성하여 파일별로 클러스터를 재배열하고, 포맷이 있는 파일들을 텍스트 파일로 변환한 후, 패턴 매칭 보드를 이용하여 비트단위(bitwise) 검색에 의해 원하는 키워드 및 통상적인 표현(regular expression)을 빠르고 정확하게 검색하고, 디지털 포렌식 시스템에서 검색의 신뢰성 및 속도를 향상시킬 수 있다.

발명의 실시를 위한 구체적인 내용

<24> 이하, 본 발명의 바람직한 실시예를 첨부된 도면을 참조하여 상세하게 설명한다.

<25> 도 2는 본 발명에 의한 고속 검색 모듈을 포함한 전체 디지털 포렌식 시스템의 구성도이다.

<26> 본 발명에 따른 디지털 포렌식 시스템은 고속 검색 모듈(100), 분석 모듈(200), 및 이미지 저장 모듈(300)로 구성된다.

<27> 이미지 저장 모듈(300)은 조사하고자 하는 디스크 이미지를 제공하고, 상기 고속 검색 모듈(100)에서 변환된 텍스트 파일을 생성한 후, 변환된 텍스트 파일을 해당 디스크 이미지와 함께 저장한다.

<28> 분석 모듈(200)은 입력된 디스크 이미지가 어느 파일 시스템을 사용하는지를 분석하고, 디스크 내 파일들이 파일 시스템의 어느 클러스터에 저장되었는지를 분석한다.

<29> 고속 검색 모듈(100)은 분석 모듈(200)로부터 검색 요청이 들어오면 이미지 저장 모듈(300)로부터 디스크 이미지를 입력받아, 입력된 디스크 이미지로부터 파일 시스템을 구성하여 파일별로 클러스터들을 재배열하고, 텍스트 정보를 가지고 있는 파일(이하, 포맷이 있는 파일로 칭함)들을 텍스트로 변환하여 저장하며, 패턴 매칭 보드를 이용하여 이미지 상의 모든 섹터와 텍스트로 변환된 파일들에서 원하는 키워드 및 통상적인 표현(regular expression)을 검색하여 그 검색 결과를 다시 분석 모듈(200)에 전송한다.

<30> 텍스트 정보를 가지고 있는 파일(포맷이 있는 파일)은 디스크 이미지 내 MS Office 파일, 한글워드(HWP), PDF 등과 같이 아스키(ASCII) 포맷으로 저장되어 있지 않는 파일을 의미한다.

<31> 패턴 매칭 보드는 일반적으로 네트워크에서 침입탐지시스템(IDS:Intrusion Detection System)에서 사용하는 보드이다. 패턴 매칭 보드는 네트워크상에서 패킷이 들어올 때, 일정 키워드나 통상적인 표현(regular expression)을 검사하여 침입을 탐지한다. 본 발명에서 사용되는 패턴 매칭 보드는 컴퓨터상에 키워드 또는 통상적인 표현(regular expression)을 검색하는데 사용하였다.

<32> 상기 고속 검색 모듈(100)은 비트단위(bitwise) 검색 기법에 의해 패턴 매칭 보드를 사용하여 원하는 다중 키워드를 동시에 검색한다.

<33> 분석 모듈(200)은 고속 검색 모듈(100)로 검색을 요청하고, 고속 검색 모듈(100)로부터 검색 결과를 수신받아 검색된 키워드에 대한 분석을 수행한다.

<34> 도 3은 본 발명에 따른 디지털 포렌식 시스템을 위한 대용량 데이터 고속 검색 방법에 대한 플로우챠트이다.

<35> 분석 모듈(200)은 이미지 저장 모듈(300)로부터 조사하고자 하는 디스크 이미지를 입력받게 되면(S110), 해당 이미지의 파일 시스템을 분석한다(S120).

<36> 파일 시스템은 저장장치 내에서 데이터를 읽고 쓰기 위해 미리 정해진다. 따라서, 분석 모듈(200)은 입력된 디스크 이미지가 어떤 파일 시스템을 사용하는지를 알아내고, 파일 시스템(file system)을 분석함으로써 디스크 내에 어떤 파일들이 어떤 클러스터(cluster)에 어떤 방식으로 저장되어 있는지를 알 수 있다.

<37> 하나의 파일이 여러 클러스터에 나뉘어 저장되는 경우, 연속된 클러스터에 순차적으로 저장되지 않는 상황이 많이 발생한다. 또한, 찾고자 하는 키워드가 이웃하지 않은 두 클러스터 경계에 걸쳐 있는 경우에는 검색되지 않는다. 따라서, 디지털 포렌식 시스템은 검색하기 전에 파일별로 클러스터들이 순차적으로 위치하도록 클러스터들을 재배열하는 과정이 반드시 필요하다.

<38> 분석 모듈(200)에 의해 파일 시스템을 분석하여 디스크 이미지 내에 어떤 파일들이 어느 클러스터들에 저장되어 있는지를 알아낸 후, 고속 검색 모듈(100)이 파일별로 클러스터들이 순차적으로 위치하도록 재배치한다(S130).

<39> 고속 검색 모듈(100)은 도 4에 도시된 바와 같이, 파일별로 클러스터들을 재배열한 후, 디스크 이미지 내에서 텍스트 정보를 가지고 있는 파일(포맷이 있는 파일)들을 찾아 이 파일들을 텍스트 파일로 변환하고, 변환된 텍스트 파일을 이미지 저장 모듈(300)로 저장한다.

<40> 디스크 이미지 내 MS Office 파일, 한글워드(HWP), PDF 등과 같이 아스키(ASCII) 포맷으로 저장되어 있지 않는 파일은 기본적으로 검색이 불가능하기 때문이다.

<41> 고속 검색 모듈(100)은 디스크 이미지 내에서 텍스트 정보를 가지고 있는 파일(포맷이 있는 파일)의 존재 여부를 판단한다(S140).

<42> 고속 검색 모듈(100)은 디스크 이미지 내 포맷이 있는 파일이 존재하는 경우, 이러한 포맷이 있는 파일들을 각 파일 포맷에 맞는 파서(parser)를 이용하여 파일의 원시 데이터에서 텍스트 데이터만을 추출하여 텍스트 파일을 생성한 후에, 이미지 저장 모듈(300)로 변환된 텍스트 파일을 해당 디스크 이미지와 함께 저장한다(S150).

<43> 그 다음 고속 검색 모듈(100)은 디스크 이미지 전체와 변환된 텍스트 파일에 대하여 패턴 매칭 보드를 이용하여 비트단위(bitwise)로 검색을 수행한다(S160).

<44> 비트단위 검색의 단점은 검색 시간이 오래 걸린다는 것이다. 비트단위 검색은 여러 개의 키워드를 동시에 검색하는 경우가 많이 발생하는데, 이러한 경우에는 더욱 많은 검색 시간이 필요하다. 그러나, 비트 단위 검색은 패턴 매칭 보드를 이용하게 되면 일정한 시간에 다중 키워드 검색이 가능하다. 따라서, 본 발명에 따른 디지털 포렌식 시스템의 고속 검색 모듈(100)은 패턴 매칭 보드를 이용하여 해당 디스크 이미지 전체를 검색하고, 검색되지 않는 포맷이 있는 파일(예: MS Office, HWP, PDF 문서)들을 검색하기 위해 텍스트로 변환된 파일들을 순차적으로 검색한다.

<45> 본 발명에 따른 디지털 포렌식 시스템을 위한 고속 검색 방법은 슬랙 영역이나 비할당 영역에 존재하는 데이터를 검색할 수 있고, 키워드 뿐만 아니라 복잡한 통상적인 표현(regular expression)을 이용한 검색도 가능하며, 파일 헤더와 같이 텍스트가 아닌 이진 데이터도 검색이 가능하다.

<46> 도 5는 고속 검색 과정에 있어서, 파일 슬랙(file slack) 스페이스를 나타낸 도면이다.

<47> 클러스터는 운영체제가 저장장치에서 데이터를 읽고 쓰는 논리적인 기본 단위이다. 파일 시스템은 클러스터 단위로 파일을 저장하게 된다. 만약, 클러스터의 크기가 4096Byte인 경우, 파일 시스템은 1000Byte 크기의 파일이 저장되더라도 4096Byte를 할당하며, 나머지 3096Byte의 공간을 사용되지 않고 버려지게 된다. 이러한 남은 공간을 슬랙(slack) 스페이스라고 한다. 이 슬랙 스페이스는 컴퓨터 포렌식에서 매우 중요한 의미를 갖는다. 대부분의 파일 시스템은 파일이 지워졌을 때 파일의 내용을 지우지 않고 파일에 대한 포인터만을 삭제하기 때문이다.

<48> 만약, 4000Byte 크기의 파일이 지워지고 그 공간에 1000Byte 크기의 파일이 덮어 씌어 지는 경우, 지워진 파일의 나머지 3000Byte의 데이터는 그대로 디스크에 남아있게 되지만 이 3000Byte의 데이터는 파일 단위의 검색을 통해 내용을 검색할 수가 없다. 그러나, 비트단위(bitwise) 검색 기법을 통해 디스크의 처음부터 끝까지 검색하게 되면, 고속 검색 모듈(100)은 이러한 지워진 데이터에 대해서도 내용 검색이 가능하게 된다.

<49> 본 발명에 따른 고속 검색 방법은 디스크 이미지에 대해서 비트단위(bitwise) 검색을 통해 디스크 내의 모든 문자열 및 패턴을 고속으로 검색하고, 슬랙(slack) 영역에 존재하는 데이터를 검색할 수 있고, 통상적인 표현(regular expression)을 이용한 검색이 가능하고, 파일 헤더와 같이 텍스트(text)가 아닌 이진(binary) 데이터 검색이 가능하다.

<50> 이상에서 설명한 바와 같이, 본 발명의 바람직한 실시예를 참조하여 설명하였지만, 본 발명이 속하는 기술 분야에서 통상적인 지식을 가진자가 하기의 특허청구범위에 기재된 본 발명의 기술적 사상 및 영역으로부터 벗어나지 않는 범위 내에서 본 발명을 다양하게 수정 또는 변형하여 실시할 수 있다.

도면의 간단한 설명

<51> 도 1은 종래의 인덱스를 이용한 정보 검색 방법을 설명한 흐름도.

<52> 도 2는 본 발명에 의한 고속 검색 모듈을 포함한 전체 디지털 포렌식 시스템의 구성도.

<53> 도 3은 본 발명에 의한 디지털 포렌식 시스템을 위한 대용량 데이터 고속 검색 방법을 도시한 플로우챠트.

<54> 도 4는 고속 검색 과정에 있어서, 클러스터 재배열을 나타낸 도면.

<55> 도 5는 고속 검색 과정에 있어서, 파일 슬랙(file slack) 스페이스를 나타낸 도면.

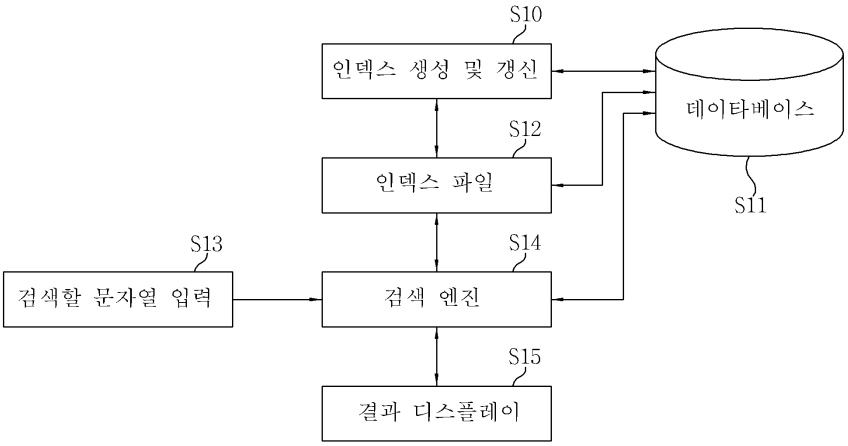
<56> <도면의 주요 부분에 대한 부호의 설명>

<57> 100: 고속 검색 모듈 200: 분석 모듈

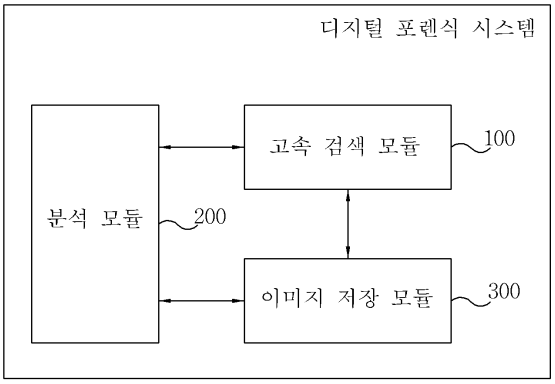
<58> 300: 이미지 저장 모듈

도면

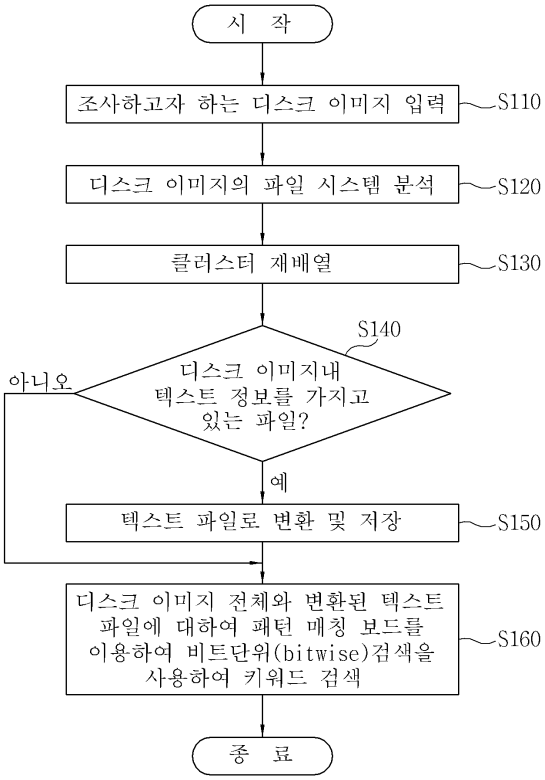
도면1


S10
인덱스 생성 및 갱신
데이타베이스
S12
인덱스 파일
S11
S13
S14
검색할 문자열 입력
검색 엔진
S15
결과 디스플레이

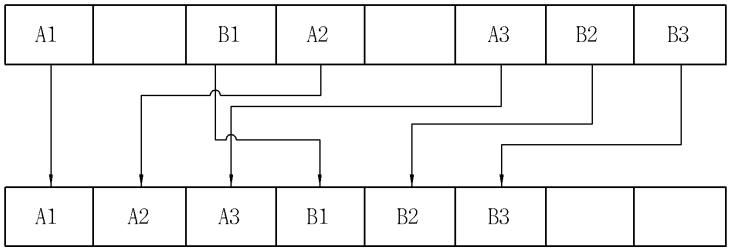
도면2


디지털 포렌식 시스템
분석 모듈
200
고속 검색 모듈
100
이미지 저장 모듈
300

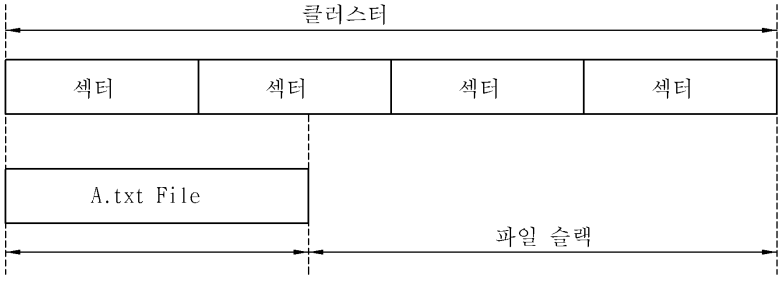
도면3


시 작
조사하고자 하는 디스크 이미지 입력
S110
디스크 이미지의 파일 시스템 분석
S120
클러스터 재배열
S130
S140
디스크 이미지내
텍스트 정보를 가지고
있는 파일?
아니오
예
텍스트 파일로 변환 및 저장
S150
디스크 이미지 전체와 변환된 텍스트
파일에 대하여 패턴 매칭 보드를
이용하여 비트단위(bitwise)검색을
사용하여 키워드 검색
S160
종 료

도면4


A1
B1
A2
A3
B2
B3
A1
A2
A3
B1
B2
B3

도면5


클러스터
섹터
섹터
섹터
섹터
A.txt File
파일 슬랙