

	(19) 대한민국특허청(KR) (12) 공개특허공보(A)	(11) 공개번호 10-2014-0095024 (43) 공개일자 2014년07월31일
(51) 국제특허분류(Int. Cl.) H04N 1/387 (2006.01) H04N 5/913 (2006.01)	(71) 출원인 고려대학교 산학협력단 서울특별시 성북구 안암로 145, 고려대학교 (안암동5가)	
(21) 출원번호 10-2014-0006761	(72) 발명자 김형중 서울특별시 서초구 신반포로 270, 117동 2601호(반포동, 반포자이아파트)	
(22) 출원일자 2014년01월20일 심사청구일자 2014년01월20일	(74) 대리인 특허법인엠에이피에스	
(30) 우선권주장 1020130007602 2013년01월23일 대한민국(KR)		

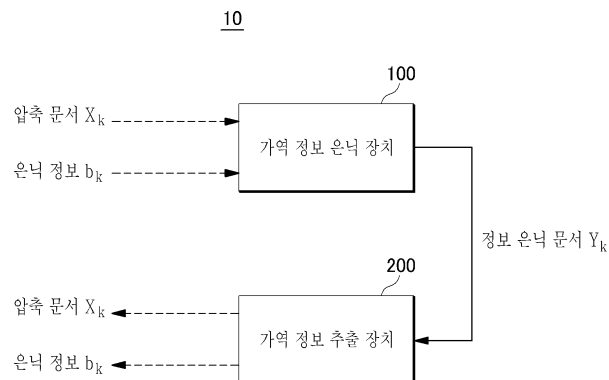
전체 청구항 수 : 총 20 항

(54) 발명의 명칭 가역 정보 은닉 시스템 및 방법

(57) 요약

이미 압축된 문서 파일에 은닉 정보 데이터가 인코딩된 정보 은닉 문서 파일로부터 가역 정보 추출 시, 기설정된 복원 포맷을 적용하여 정보 은닉 문서 파일을 기설정된 길이의 비트 블록 단위로 디코딩하고, 디코딩된 비트 블록에 포함된 복수의 문자를 기설정된 정보 분석 기준에 기초하여 분석하고, 분석한 결과에 따라 유의미 지수, 무의미 지수 및 문서 지수 중 적어도 하나를 포함하는 정보 지수를 산출하고, 산출된 정보 지수와 기설정된 정보 지수 임계값을 비교한 결과에 기초하여 비트 블록에 은닉된 은닉 정보 비트를 추출하고, 비트 블록 별로 디코딩 이전의 비트 값을 은닉 정보 비트의 값에 기초하여 변환시키고, 비트 값이 변환된 비트 블록을 복원 포맷으로 디코딩하여 원본 문서 파일로 복원한다.

대표도 - 도1



이 발명을 지원한 국가연구개발사업

과제고유번호	10045515
부처명	산업통상자원부
연구사업명	국가표준기술력향상사업
연구과제명	영상 재압축 기술 표준화 연구개발
기 여 율	1/2
주관기관	고려대학교 산학협력단
연구기간	2013.05.01 ~ 2015.04.30이 발명을 지원한 국가연구개발사업
과제고유번호	2012R1A2A2A01015587
부처명	교육과학기술부
연구사업명	중견연구자지원사업
연구과제명	H.264 기반 정보은닉 기술 개발
기 여 율	1/2
주관기관	고려대학교 산학협력단
연구기간	2013.05.01 ~ 2015.04.30

특허청구의 범위

청구항 1

가역 정보 은닉 장치에 있어서,

원본 문서 파일을 기설정된 압축 포맷으로 인코딩하여 압축 문서 파일을 출력하는 문서 압축부;

상기 압축 문서 파일의 비트스트림을 각각 기설정된 제 1 길이를 갖는 복수의 비트 블록으로 분할하고, 기설정된 은닉 정보 데이터를 각각 기설정된 제 2 길이를 갖는 복수의 은닉 정보 비트로 분할하여 상기 복수의 비트 블록에 각각 할당하고, 할당된 상기 은닉 정보 비트의 값에 기초하여 상기 비트 블록의 비트 값을 변환시키는 정보 은닉 처리부; 및

상기 비트 값을 변환시킨 비트 블록들을 포함하는 정보 은닉 문서 파일을 제공하는 정보 은닉 문서 제공부를 포함하는 가역 정보 은닉 장치.

청구항 2

제 1 항에 있어서,

상기 정보 은닉 처리부는,

상기 할당된 은닉 정보 비트의 값에 따라 상기 비트 블록의 비트 값을 1 비트 단위로 유지 또는 반전시키는 가역 정보 은닉 장치.

청구항 3

제 1 항에 있어서,

상기 정보 은닉 처리부는,

상기 비트 블록의 길이와 대응된 길이를 갖는 복수의 난수 비트스트림을 생성하고,

상기 할당된 은닉 정보 비트에 매칭된 난수 비트스트림 및 상기 비트 블록을 1 비트 단위로 배타적 논리합(XOR) 또는 배타적 부정 논리합(XNOR) 연산하는 가역 정보 은닉 장치.

청구항 4

제 1 항에 있어서,

상기 원본 문서 파일의 비트스트림을 복수의 비트 블록으로 구분하고, 상기 비트 블록 별로 포함된 복수의 문자를 기설정된 정보 분석 기준에 따라 분석하여 유의미 지수, 무의미 지수, 및 문서 지수 중 적어도 하나를 산출하고, 상기 산출된 유의미 지수, 무의미 지수, 및 문서 지수 중 적어도 하나를 정보 지수 임계값으로 설정하는 정보 지수 임계값 설정부를 더 포함하되,

상기 정보 은닉 문서 제공부는,

상기 정보 지수 임계값을 가역 정보 추출 장치로 제공하는 가역 정보 은닉 장치.

청구항 5

제 4 항에 있어서,

상기 정보 분석 기준은,

기설정된 사전(dictionary) 및 컴퓨터 표준 코드 중 적어도 하나로 설정된 가역 정보 은닉 장치.

청구항 6

가역 정보 추출 장치에 있어서,

압축 문서 파일에 은닉 정보 데이터가 인코딩된 정보 은닉 문서 파일을 입력받는 압축 문서 입력부;

기설정된 복원 포맷을 적용하여 상기 정보 은닉 문서 파일을 기설정된 길이의 비트 블록 단위로 디코딩하는 복원 전처리부;

상기 디코딩된 비트 블록에 포함된 복수의 문자를 기설정된 정보 분석 기준에 기초하여 분석하고, 상기 분석한 결과에 따라 유의미 지수, 무의미 지수 및 문서 지수 중 적어도 하나를 포함하는 정보 지수를 산출하는 정보 분석부;

산출된 상기 정보 지수와 기설정된 정보 지수 임계값을 비교한 결과에 기초하여 상기 비트 블록에 은닉된 은닉 정보 비트를 추출하고, 상기 비트 블록 별로 상기 디코딩 이전의 비트 값을 상기 은닉 정보 비트의 값에 기초하여 변환시키는 은닉 정보 추출부; 및

상기 은닉 정보 추출부를 통해 상기 비트 값이 변환된 상기 비트 블록을 상기 복원 포맷으로 디코딩하여 원본 문서 파일로 복원하는 문서 복원부를 포함하는 가역 정보 추출 장치.

청구항 7

제 6 항에 있어서,

상기 은닉 정보 추출부는,

상기 비트 블록 당 한 자릿수 비트 단위의 상기 은닉 정보 비트를 추출하며,

상기 산출한 정보 지수와 상기 정보 지수 임계값을 비교한 결과에 따라 제 1 비트 값을 갖는 제 1 은닉 정보 비트 또는 상기 제 1 비트 값이 반전된 제 2 비트 값을 갖는 제 2 은닉 정보 비트를 추출하되,

추출된 상기 은닉 정보 비트의 값에 따라 상기 비트 블록 별 비트 값을 1 비트 단위로 유지 또는 반전시키는 가역 정보 추출 장치.

청구항 8

제 6 항에 있어서,

상기 압축 문서 입력부는,

가역 정보 은닉 장치로부터 상기 원본 문서 파일의 비트 블록 별로 산출된 상기 정보 지수 임계값을 더 입력받으며,

상기 정보 지수 임계값은,

상기 유의미 지수, 무의미 지수 및 문서 지수 중 적어도 하나의 임계값을 포함하는 가역 정보 추출 장치.

청구항 9

제 6 항에 있어서,

상기 정보 분석부는,

기설정된 사전(dictionary)을 상기 정보 분석 기준으로 사용하여, 상기 비트 블록에 포함된 복수의 문자에 의한 단어 중 상기 사전에 포함된 유의미 단어를 추출하고,

상기 유의미 단어의 개수에 기초하여 상기 유의미 지수를 산출하는 가역 정보 추출 장치.

청구항 10

제 6 항에 있어서,

상기 정보 분석부는,

기설정된 글자 코드, 숫자 코드, 부호 코드 및 제어 코드를 포함하는 컴퓨터 표준 코드를 상기 정보 분석 기준으로 사용하여, 상기 비트 블록에 포함된 복수의 문자 별로 대응하는 상기 컴퓨터 표준 코드를 추출하고, 상기 추출된 컴퓨터 표준 코드 중 상기 부호 코드에 대응된 문자의 개수에 기초하여 상기 무의미 지수를 산출하는 가역 정보 추출 장치.

청구항 11

제 6 항에 있어서,

상기 정보 분석부는,

기설정된 글자 코드, 숫자 코드, 부호 코드 및 제어 코드를 포함하는 컴퓨터 표준 코드를 상기 정보 분석 기준으로 사용하여, 상기 비트 블록에 포함된 복수의 문자 별로 대응하는 상기 컴퓨터 표준 코드를 추출하고, 상기 추출된 컴퓨터 표준 코드 중 상기 글자 코드 및 숫자 코드 중 적어도 하나에 대응된 문자의 제 1 개수와 상기 부호 코드 및 제어 코드 중 적어도 하나에 대응된 문자의 제 2 개수에 기초하여 상기 문서 지수를 산출하는 가역 정보 추출 장치.

청구항 12

가역 정보 은닉 장치가 기압축된 문서 파일에 정보를 은닉하는 방법에 있어서,

(a) 원본 문서 파일을 기설정된 압축 포맷으로 인코딩하여 압축 문서 파일을 출력하는 단계;

(b) 상기 압축 문서 파일의 비트스트림을 각각 기설정된 제 1 길이를 갖는 복수의 비트 블록으로 분할하는 단계;

(c) 기설정된 은닉 정보 데이터를 기설정된 제 2 길이 단위로 분할한 복수의 은닉 정보 비트를 상기 복수의 비트 블록에 할당하는 단계;

(d) 할당된 상기 은닉 정보 비트의 값에 기초하여 상기 비트 블록의 비트 값을 변환시키는 단계; 및

(e) 상기 비트 값을 변환시킨 비트 블록들을 포함하는 정보 은닉 문서 파일을 제공하는 단계를 포함하는 가역 정보 은닉 방법.

청구항 13

제 12 항에 있어서,

상기 (d) 단계는,

상기 할당된 은닉 정보 비트의 값에 따라 상기 비트 블록의 비트 값을 1비트 단위로 유지 또는 반전시키는 가역 정보 은닉 방법.

청구항 14

제 12 항에 있어서,

상기 (d) 단계는,

상기 비트 블록의 길이와 대응된 길이를 갖는 복수의 난수 비트스트림을 생성하는 단계;

상기 할당된 은닉 정보 비트의 값 별로 서로 상이한 상기 난수 비트스트림을 매칭시키는 단계; 및

상기 할당된 은닉 정보 비트에 매칭된 난수 비트스트림 및 상기 비트 블록을 1 비트 단위로 배타적 논리합(XOR) 또는 배타적 부정 논리합(XNOR) 연산하는 단계를 포함하는 가역 정보 은닉 방법.

청구항 15

제 12 항에 있어서,

상기 (e) 단계 이전에,

상기 원본 문서 파일의 비트스트림을 복수의 비트 블록으로 구분하는 단계;

상기 원본 문서 파일의 비트 블록 별로 포함된 복수의 문자를 기설정된 정보 분석 기준에 따라 분석하여 유의미 지수, 무의미 지수, 및 문서 지수 중 적어도 하나를 산출하는 단계; 및

상기 산출된 유의미 지수, 무의미 지수, 및 문서 지수 중 적어도 하나를 정보 지수 임계값으로 설정하는 단계를 더 포함하되,

상기 (e) 단계는,

상기 정보 지수 임계값을 가역 정보 추출 장치로 더 제공하는 가역 정보 은닉 방법.

청구항 16

가역 정보 추출 장치가 정보 은닉 문서 파일에 은닉된 은닉 정보 데이터를 추출하는 방법에 있어서,

- (a) 압축 문서 파일에 은닉 정보 데이터가 인코딩된 상기 정보 은닉 문서 파일을 입력받는 단계;
- (b) 기설정된 복원 포맷을 적용하여 상기 정보 은닉 문서 파일을 기설정된 길이의 비트 블록 단위로 디코딩하는 단계;
- (c) 상기 디코딩된 비트 블록에 포함된 복수의 문자를 기설정된 정보 분석 기준에 기초하여 분석하여, 상기 분석한 결과에 따라 유의미 지수, 무의미 지수 및 문서 지수 중 적어도 하나를 포함하는 정보 지수를 산출하는 단계;
- (d) 산출된 상기 정보 지수와 기설정된 정보 지수 임계값을 비교한 결과에 기초하여 상기 비트 블록에 은닉된 은닉 정보 비트를 추출하는 단계;
- (e) 상기 비트 블록 별로 상기 디코딩 이전의 비트 값을 상기 은닉 정보 비트의 값에 기초하여 변환시키는 단계; 및
- (f) 상기 비트 값이 변환된 비트 블록을 상기 복원 포맷으로 디코딩하여 원본 문서 파일로 복원하는 단계를 포함하는 가역 정보 추출 방법.

청구항 17

제 16 항에 있어서,

상기 (d) 단계는,

상기 비트 블록 당 한 자릿수 비트 단위의 상기 은닉 정보 비트를 추출하되,

상기 산출한 정보 지수와 상기 정보 지수 임계값을 비교한 결과에 따라 제 1 비트 값을 갖는 제 1 은닉 정보 비트 또는 상기 제 1 비트 값이 반전된 제 2 비트 값을 갖는 제 2 은닉 정보 비트를 추출하고,

상기 (e) 단계는,

상기 추출된 은닉 정보 비트의 값에 따라 상기 비트 블록 별 비트 값을 1 비트 단위로 유지 또는 반전시키는 가역 정보 추출 방법.

청구항 18

제 16 항에 있어서,

상기 (c) 단계는,

기설정된 사전(dictionary)을 상기 정보 분석 기준으로 사용하여, 상기 비트 블록에 포함된 복수의 문자에 의한 단어 중 상기 사전에 포함된 유의미 단어를 추출하는 단계; 및

상기 유의미 단어의 개수에 기초하여 상기 유의미 지수를 산출하는 단계를 포함하는 가역 정보 추출 방법.

청구항 19

제 16 항에 있어서,

상기 (c) 단계는,

기설정된 글자 코드, 숫자 코드, 부호 코드 및 제어 코드를 포함하는 컴퓨터 표준 코드를 상기 정보 분석 기준으로 사용하여, 상기 비트 블록에 포함된 복수의 문자 별로 대응하는 상기 컴퓨터 표준 코드를 추출하는 단계; 및

상기 추출된 컴퓨터 표준 코드 중 상기 부호 코드에 대응된 문자의 개수에 기초하여 상기 무의미 지수를 산출하

는 단계를 포함하는 가역 정보 추출 방법.

청구항 20

제 16 항에 있어서,

상기 (c) 단계는,

기설정된 글자 코드, 숫자 코드, 부호 코드 및 제어 코드를 포함하는 컴퓨터 표준 코드를 상기 정보 분석 기준으로 사용하여, 상기 비트 블록에 포함된 복수의 문자 별로 대응하는 상기 컴퓨터 표준 코드를 추출하는 단계; 및

상기 추출된 컴퓨터 표준 코드 중 상기 글자 코드 및 숫자 코드 중 적어도 하나에 대응된 문자의 제 1 개수와 상기 부호 코드 및 제어 코드 중 적어도 하나에 대응된 문자의 제 2 개수에 기초하여 상기 문서 지수를 산출하는 단계를 포함하는 가역 정보 추출 방법.

명세서

기술분야

[0001] 본 발명은 압축된 문서 파일에 정보 데이터를 가역적으로 은닉하고, 압축 문서 상에 은닉된 정보 데이터를 추출하는 시스템 및 방법에 관한 것이다.

배경기술

[0002] 많은 정보들이 디지털 형태로 저장됨에 따라, 쉽게 복사되거나 변경될 수 있는 디지털 매체를 보호하기 위하여 여러 가지 정보 보호 기법들이 개발되고 있다.

[0003] 그 중 정보 은닉(information hiding) 기법은 암호화와 함께 기밀성을 보장하기 위해 널리 사용되고 있다. 이러한, 정보 은닉 기법은 디지털 매체를 변경함으로써 디지털 매체에 데이터를 삽입할 수 있다.

[0004] 이와 관련하여, 대한민국공개특허 제 10-2007-0076130호(발명의 명칭: 문서 인증을 위한 정보 삽입 및 추출 방법)는, 정보의 삽입 및 추출 과정에서 원정보(Original Data)의 손실이 발생하지 않는 무손실 정보 은닉(Lossless Data Embedding) 방법에 관한 것으로, 이미지를 다수개의 블록으로 분할하고, 분할된 각 블록을 서브블록으로 분할하고, 분할된 각 서브블록 중 하나의 픽셀에 인증 정보를 삽입하되, 서브블록 중 최소 픽셀값을 갖는 픽셀에 인증 정보(이진정보)를 삽입하는 기술이 개시되어 있다.

[0005] 그러나 문서를 압축할 경우 난수 수준의 출력이 생성됨에 따라, 원본 문서가 아닌 압축된 문서에 가역적으로 정보를 은닉하는 것은 용이하지 않다는 한계가 있었다.

발명의 내용

해결하려는 과제

[0006] 본 발명의 실시예는 전술한 종래 기술의 문제점을 해결하기 위한 것으로서, 이미 압축된 문서 파일에 추가로 은닉 정보 데이터를 인코딩하는 정보 은닉 장치 및 방법과, 정보가 은닉된 압축 문서 파일로부터 은닉 정보 데이터를 추출하는 정보 추출 장치 및 방법을 제공하고자 한다.

[0007] 다만, 본 실시예가 이루고자 하는 기술적 과제는 상기된 바와 같은 기술적 과제로 한정되지 않으며, 또 다른 기술적 과제들이 존재할 수 있다.

과제의 해결 수단

[0008] 상술한 기술적 과제를 달성하기 위한 기술적 수단으로서, 본 발명의 일 측면에 따른 가역 정보 은닉 장치는, 원본 문서 파일을 기설정된 압축 포맷으로 인코딩하여 압축 문서 파일을 출력하는 문서 압축부; 상기 압축 문서 파일의 비트스트림을 각각 기설정된 제 1 길이를 갖는 복수의 비트 블록으로 분할하고, 기설정된 은닉 정보 데이터를 각각 기설정된 제 2 길이를 갖는 복수의 은닉 정보 비트로 분할하여 상기 복수의 비트 블록에 각각 할당하고, 할당된 상기 은닉 정보 비트의 값에 기초하여 상기 비트 블록의 비트 값을 변환시키는 정보 은닉 처리부; 및 상기 비트 값을 변환시킨 비트 블록들을 포함하는 정보 은닉 문서 파일을 제공하는 정보 은닉 문서 제공부를

포함한다.

[0009] 본 발명의 다른 측면에 따른 가역 정보 추출 장치는, 압축 문서 파일에 은닉 정보 데이터가 인코딩된 정보 은닉 문서 파일을 입력받는 압축 문서 입력부; 기설정된 복원 포맷을 적용하여 상기 정보 은닉 문서 파일을 기설정된 길이의 비트 블록 단위로 디코딩하는 복원 전처리부; 상기 디코딩된 비트 블록에 포함된 복수의 문자를 기설정된 정보 분석 기준에 기초하여 분석하고, 상기 분석한 결과에 따라 유의미 지수, 무의미 지수 및 문서 지수 중 적어도 하나를 포함하는 정보 지수를 산출하는 정보 분석부; 산출된 상기 정보 지수와 기설정된 정보 지수 임계값을 비교한 결과에 기초하여 상기 비트 블록에 은닉된 은닉 정보 비트를 추출하고, 상기 비트 블록 별로 상기 디코딩 이전의 비트 값을 상기 은닉 정보 비트의 값에 기초하여 변환시키는 은닉 정보 추출부; 및 상기 은닉 정보 추출부를 통해 상기 비트 값이 변환된 상기 비트 블록을 상기 복원 포맷으로 디코딩하여 원본 문서 파일로 복원하는 문서 복원부를 포함한다.

[0010] 본 발명의 또 다른 측면에 따른 가역 정보 은닉 장치가 기압축된 문서 파일에 정보를 은닉하는 방법은, (a) 원본 문서 파일을 기설정된 압축 포맷으로 인코딩하여 압축 문서 파일을 출력하는 단계; (b) 상기 압축 문서 파일의 비트스트림을 각각 기설정된 제 1 길이를 갖는 복수의 비트 블록으로 분할하는 단계; (c) 기설정된 은닉 정보 데이터를 기설정된 제 2 길이 단위로 분할한 복수의 은닉 정보 비트를 상기 복수의 비트 블록에 할당하는 단계; (d) 할당된 상기 은닉 정보 비트의 값에 기초하여 상기 비트 블록의 비트 값을 변환시키는 단계; 및 (e) 상기 비트 값을 변환시킨 비트 블록들을 포함하는 정보 은닉 문서 파일을 제공하는 단계를 포함한다.

[0011] 본 발명의 또 다른 측면에 따른 가역 정보 추출 장치가 정보 은닉 문서 파일에 은닉된 은닉 정보 데이터를 추출하는 방법은, (a) 압축 문서 파일에 은닉 정보 데이터가 인코딩된 상기 정보 은닉 문서 파일을 입력받는 단계; (b) 기설정된 복원 포맷을 적용하여 상기 정보 은닉 문서 파일을 기설정된 길이의 비트 블록 단위로 디코딩하는 단계; (c) 상기 디코딩된 비트 블록에 포함된 복수의 문자를 기설정된 정보 분석 기준에 기초하여 분석하여, 상기 분석한 결과에 따라 유의미 지수, 무의미 지수 및 문서 지수 중 적어도 하나를 포함하는 정보 지수를 산출하는 단계; (d) 산출된 상기 정보 지수와 기설정된 정보 지수 임계값을 비교한 결과에 기초하여 상기 비트 블록에 은닉된 은닉 정보 비트를 추출하는 단계; (e) 상기 비트 블록 별로 상기 디코딩 이전의 비트 값을 상기 은닉 정보 비트의 값에 기초하여 변환시키는 단계; 및 (f) 상기 비트 값이 변환된 비트 블록을 상기 복원 포맷으로 디코딩하여 원본 문서 파일로 복원하는 단계를 포함한다.

발명의 효과

[0012] 전술한 본 발명의 과제 해결 수단에 의하면, 원본 데이터를 수정하지 않고도 완전하게 데이터를 복원할 수 있는 가역 인코딩 기법을 통해, 기압축된 상태의 문서 파일의 데이터 손실없이도 은닉 정보 데이터를 추가로 인코딩할 수 있다.

[0013] 또한, 본 발명의 과제 해결 수단 중 어느 하나에 의하면, 압축 문서 파일 상에 은닉 정보 데이터가 추가 인코딩된 문서 파일을, 비트 블록 별로 포함된 문자의 속성에 따라 산출한 정보 지수에 기초하여 디코딩함으로써 정확한 디코딩 결과를 획득할 수 있다.

도면의 간단한 설명

[0014] 도 1은 본 발명의 일 실시예에 따른 가역 정보 은닉 시스템의 구성을 나타낸 블록도이다.

도 2는 본 발명의 일 실시예에 따른 가역 정보 은닉 장치의 구성을 나타낸 블록도이다.

도 3은 본 발명의 일 실시예에 따른 압축 문서 파일의 비트스트림 구조를 설명하기 위한 도면이다.

도 4는 본 발명의 일 실시예에 따른 문서 파일의 비트 블록에 포함된 문자를 설명하기 위한 일례이다.

도 5는 본 발명의 일 실시예에 따른 가역 정보 은닉 방법을 설명하기 위한 순서도이다.

도 6은 본 발명의 일 실시예에 따른 가역 정보 추출 장치의 구성을 나타낸 블록도이다.

도 7은 본 발명의 일 실시예에 따른 정보 지수를 산출하는 방식을 설명하기 위한 아스키 코드 테이블이다.

도 8은 본 발명의 일 실시예에 따른 가역 정보 추출 방법을 설명하기 위한 순서도이다.

발명을 실시하기 위한 구체적인 내용

[0015] 명세서 전체에서, 어떤 부분이 다른 부분과 "연결"되어 있다고 할 때, 이는 "직접적으로 연결"되어 있는 경우뿐

아니라, 그 중간에 다른 소자를 사이에 두고 "전기적으로 연결"되어 있는 경우도 포함한다. 또한 어떤 부분이 어떤 구성요소를 "포함"한다고 할 때, 이는 특별히 반대되는 기재가 없는 한 다른 구성요소를 제외하는 것이 아니라 다른 구성요소를 더 포함할 수 있는 것을 의미한다.

- [0016] 도 1은 본 발명의 일 실시예에 따른 가역 정보 은닉 시스템의 구성을 나타낸 블록도이다.
- [0017] 도 1에서와 같이, 본 발명의 일 실시예에 따른 가역 정보 은닉 시스템(10)은, 가역 정보 은닉 장치(100) 및 가역 정보 추출 장치(200)를 포함한다.
- [0018] 구체적으로, 가역 정보 은닉 장치(100)는 설정된 압축 포맷으로 1차 인코딩된 압축 문서(X_k)에 은닉 정보 데이터(b_k)를 은닉시키는 2차 인코딩을 처리하고, 2차 인코딩 결과에 따른 정보 은닉 문서(Y_k)를 출력한다.
- [0019] 가역 정보 추출 장치(200)는 가역 정보 은닉 장치(100)로부터 제공된 정보 은닉 문서(Y_k)를 입력받아, 정보 은닉 문서(Y_k)로부터 은닉 정보 데이터(b_k)를 추출하여 압축 문서(X_k)를 획득한다. 그리고, 가역 정보 추출 장치(200)는 추출된 은닉 정보 데이터(b_k)에 기초하여 압축 문서(X_k)를 디코딩하여 원본 문서를 복원한다.
- [0020] 이하, 도 2 내지 도 5를 참조하여 본 발명의 일 실시예에 따른 가역 정보 은닉 장치의 구성 및 동작에 대해서 상세히 설명하도록 한다.
- [0021] 도 2는 본 발명의 일 실시예에 따른 가역 정보 은닉 장치의 구성을 나타낸 블록도이다.
- [0022] 도 2에서와 같이, 본 발명의 일 실시예에 따른 가역 정보 은닉 장치(100)는 문서 압축부(110), 정보 은닉 처리부(120), 정보 은닉 문서 제공부(130) 및 정보 지수 임계값 설정부(140)를 포함한다.
- [0023] 문서 압축부(110)는 원본 문서 파일을 기설정된 압축 포맷으로 인코딩(이하, '1차 인코딩'이라고 지칭함)하여 압축 문서 파일을 출력한다.
- [0024] 정보 은닉 처리부(120)는 문서 압축부(110)로부터 출력된 압축 문서 파일의 비트 스트림을 복수의 비트 블록으로 변환하고, 각 비트 블록에 은닉 정보 데이터를 할당 및 은닉 처리한다.
- [0025] 구체적으로, 정보 은닉 처리부(120)는 압축 문서 파일의 비트스트림을 각각 기설정된 길이(이하, 'L'이라고 지칭함)를 갖는 복수의 비트 블록으로 분할한다. 참고로, 하기에서 설명할 가역 정보 추출 장치(200)의 은닉 정보 추출 처리를 위하여, 정보 은닉 처리부(120)는 정보 은닉 문서 제공부(130)를 통해 비트 블록의 길이 정보(즉, 'L')를 더 제공할 수 있다.
- [0026] 그리고, 정보 은닉 처리부(120)는 사전에 설정된 은닉 정보 데이터의 비트스트림을 기설정된 길이(즉, 비트 자릿수)의 비트 단위로 분할하여 복수의 은닉 정보 비트를 생성한다. 이때, 정보 은닉 처리부(120)는 은닉 정보 데이터를 한 자릿수의 비트 단위 또는 두 자릿수 이상의 비트 단위로 분할할 수 있다.
- [0027] 예를 들어, 도 3은 본 발명의 일 실시예에 따른 압축 문서 파일의 비트스트림 구조를 설명하기 위한 도면이다.
- [0028] 도 3에 도시한 바와 같이, 압축 문서 파일의 비트스트림은 헤더 및 페이로드로 구성되며, 정보 은닉 처리부(120)는 압축 문서 파일의 비트스트림 중 헤더를 제외한 페이로드의 비트스트림을 기설정된 길이(즉, L) 단위로 분할하여 복수개의 비트 블록(X_k)을 구분한다. 이때, 정보 은닉 처리부(120)는 은닉할 은닉 정보 데이터의 크기에 기초하여 비트 블록의 길이를 설정할 수 있다. 즉, 정보 은닉 처리부(120)는 압축 문서 파일의 비트스트림을 은닉 정보 비트의 개수에 대응하는 개수의 비트 블록으로 분할할 수 있다.
- [0029] 정보 은닉 처리부(120)는 각각의 비트 블록 별로 은닉 정보 비트를 할당하고, 할당된 은닉 정보 비트의 값에 따라 해당 비트 블록의 비트 값을 변환시키는 인코딩(이하, '2차 인코딩'이라고 지칭함)을 수행한다.
- [0030] 구체적으로, 정보 은닉 처리부(120)는 각 비트 블록에 한 자릿수 비트 단위의 은닉 정보 비트가 할당된 경우, 해당 은닉 정보 비트의 값에 따라 비트 블록의 비트 값을 1비트 단위로 유지 또는 반전시킨다.
- [0031] 예를 들어, 정보 은닉 처리부(120)는 압축 문서 파일의 페이로드의 각 비트 블록들에 대해 하기 수학적 식 1에 따른 연산을 적용하여 정보를 은닉한다.
- [0032] <수학적 식 1>

$b_k=0$ 이면 $Y_k=X_k$

$b_k=1$ 이면 $Y_k=\bar{X}_k$

[0033]

[0034]

즉, 은닉 정보 비트의 값이 한 자릿수 비트인 '0' 인 경우, 비트 블록(X_k)의 비트를 그대로 유지시켜 출력한다. 반면, 은닉 정보 비트의 값이 한 자릿수 비트인 '1' 인 경우, 비트 블록(X_k)의 비트를 1 비트 단위로 반전시켜 출력한다. 참고로, 비트 블록의 비트 값을 유지시키거나 또는 반전시키는 은닉 정보 비트의 값은 상호 변경될 수 있다.

[0035]

또한, 정보 은닉 처리부(120)는 각 비트 블록에 두 자릿수 이상의 비트 단위의 은닉 정보 비트를 은닉시킬 수 있다.

[0036]

구체적으로, 정보 은닉 처리부(120)는 비트 블록의 길이와 대응된 길이를 갖는 복수의 난수 비트스트림을 생성하고, 할당된 은닉 정보 비트의 값 별로 서로 상이한 난수 비트스트림을 매칭시킨다. 그리고, 정보 은닉 처리부(120)는 할당된 은닉 정보 비트에 매칭된 난수 비트스트림과 해당 비트 블록을 1 비트 단위로 배타적 논리합(XOR)(또는 배타적 부정 논리합(XNOR)) 연산시켜 2차 인코딩을 처리한다.

[0037]

예를 들어, 정보 은닉 처리부(120)는 은닉 정보 비트가 두 자릿수 비트인 경우, 하기 수학적 2를 적용하여, 각 비트 값 별로 매칭된 난수 비트스트림에 따라 각 비트 블록의 비트 값을 변형시킨다.

[0038]

<수학적 2>

$b_k=0$ 이면 (즉, 00_2) $Y_k=X_k \oplus R_0$

$b_k=1$ 이면 (즉, 01_2) $Y_k=X_k \oplus R_1$

$b_k=2$ 이면 (즉, 10_2) $Y_k=X_k \oplus R_2$

$b_k=3$ 이면 (즉, 11_2) $Y_k=X_k \oplus R_3$

[0039]

[0040]

즉, 은닉 정보 비트(b_k)의 값이 두 자릿수 비트인 '002' 인 경우 난수 비트스트림 ' R_0 ' 이 매칭되며, 비트 블록(X_k)과 난수 비트스트림(R_0)의 각 자릿수의 비트를 XOR(또는 'NXOR') 연산하여 정보가 은닉된 비트 블록을 출력한다. 이와 같은 방식으로, K개의 서로 다른 난수를 난수 비트스트림으로 사용할 경우, 각 비트 블록 당 ' $\log_2 K$ '의 은닉 정보 비트를 은닉시킬 수 있다.

[0041]

이처럼, 두 자릿수 이상의 은닉 정보 비트를 비트 블록에 은닉시킬 경우, 하기에서 설명할 가역 정보 추출 장치(200)의 은닉 정보 추출 처리를 위하여, 정보 은닉 처리부(120)는 정보 은닉 문서 제공부(130)를 통해 각 은닉 정보 비트의 값 별로 매칭된 난수 비트스트림을 더 제공할 수 있다.

[0042]

한편, 도 3에서와 같이, 압축 문서 파일의 비트스트림을 기설정된 길이 L로 분할할 경우, L보다 작은 길이를 갖는 비트 블록이 잔여할 수 있다. 즉, 도 3에서는 압축 문서 파일이 n개의 비트 블록(즉, X_1 내지 X_n)으로 분할되되, 마지막 비트 블록(즉, X_n)의 길이가 L보다 작은 M인 것을 나타내었다. 이때, 정보 은닉 처리부(120)는 비트 블록에 대해 기설정된 길이(즉, L) 미만의 잔여 비트를 제외한 비트 블록들에 은닉 정보 비트를 할당할 수 있다.

[0043]

정보 은닉 문서 제공부(130)는 2차 인코딩된(즉, 비트 값을 변환시킨) 비트 블록들을 포함하는 정보 은닉 문서 파일을 가역 정보 추출 장치(200)로 제공한다.

[0044]

한편, 하기에서 설명할 가역 정보 추출 장치(200)가 정보 은닉 문서 파일에 은닉된 은닉 정보 데이터의 추출 및 원본 문서 파일의 복원을 성공적으로 처리하기 위해서는, 원본 문서 파일의 비트 블록 별로 포함된 복수의 '문자'의 속성을 분석하여 산출된 정보 지수가 필요하다. 참고로, 본 발명의 일 실시예에서 '문자'는 글자, 숫자, 부호, 제어 문자 등 의사 소통을 위한 모든 기호 체계를 의미한다.

[0045]

이를 위해, 본 발명의 일 실시예에 따른 가역 정보 은닉 장치(100)는 정보 지수 임계값 설정부(140)를 포함한다.

- [0046] 정보 지수 임계값 설정부(140)는 원본 문서 파일의 비트스트림을 기설정된 크기를 갖는 복수의 비트 블록으로 분할하고, 각 비트 블록의 비트 값에 대응된 문자에 대해 기설정된 정보 지수를 산출한다. 본 발명의 일 실시예에서, 기설정된 정보 지수는 유의미 지수, 무의미 지수 및 문서 지수를 포함한다. 그리고, 정보 지수 임계값 설정부(140)는 산출된 정보 지수를 정보 지수 임계값으로서 설정한다.
- [0047] 정보 지수 임계값 설정부(140)는 정보 은닉 문서 제공부(130)를 통해 정보 지수 임계값을 가역 정보 은닉 추출 장치(200)로 제공한다.
- [0048] 구체적으로, 도 4는 본 발명의 일 실시예에 따른 문서 파일의 비트 블록에 포함된 문자를 설명하기 위한 일례이다.
- [0049] 도 4의 (a)에서는 문서 파일의 비트 블록에 유의미한 문자들이 포함된 경우를 나타내었다. 이때, 도 4의 (a)에 도시된 문자들이 원본 문서 파일의 비트 블록에 포함된 문자들이므로 가정하면, 정보 지수 임계값 설정부(140)는 해당 비트 블록의 문자들에 포함된 단어 및 기설정된 정보 분석 기준에 기초하여 유의미 지수 임계값을 산출한다. 이때, 기설정된 정보 분석 기준은 사전(dictionary) 데이터로서, 현존하는 단어가 저장된 데이터베이스로부터 획득할 수 있다. 참고로, 정보 지수 임계값 설정부(140)는 자체적으로 사전 데이터베이스를 포함하거나 외부 사전 데이터베이스를 이용하여 정보 분석 기준을 적용할 수 있다.
- [0050] 한편, 도 4의 (b)에서는 문서 파일의 비트 블록에 무의미한 문자들이 포함된 경우를 나타내었다. 이때, 도 4의 (b)에 도시된 문자들이 원본 문서 파일의 비트 블록에 포함된 문자들이므로 가정하면, 정보 지수 임계값 설정부(140)는 해당 비트 블록의 문자 별 속성과 기설정된 정보 분석 기준에 기초하여 무의미 지수 임계값 및 문서 지수 임계값을 설정할 수 있다. 이때, 기설정된 정보 분석 기준은 아스키 코드(ASCII) 및 유니코드(UNICODE) 등의 컴퓨터 표준 코드 데이터로 설정된다.
- [0051] 이처럼, 정보 지수 임계값 설정부(140)는 원본 문서 파일의 비트 블록에 포함된 문자의 속성을 기설정된 정보 분석 기준에 기초하여 분석하고, 분석의 결과에 따라 유의미 지수 임계값, 무의미 지수 임계값 및 문서 지수 임계값 중 적어도 하나를 설정한다.
- [0052] 한편, 정보 지수 임계값 설정부(140)가 원본 문서 파일에 기준하여 정보 지수 임계값을 산출하는 방식은, 하기에서 설명할 가역 정보 추출 장치(200)가 정보 은닉 문서 파일의 비트 블록 별로 처리하는 정보 지수 산출 방식과 대응된다. 따라서, 설명의 편의상 정보 지수 임계값 설정부(140)가 정보 지수 임계값을 산출하는 방식에 대한 상세한 설명은 생략하도록 한다.
- [0053] 이하, 도 5를 참조하여 본 발명의 일 실시예에 따른 가역 정보 은닉 방법에 대해서 상세히 설명하도록 한다.
- [0054] 도 5는 본 발명의 일 실시예에 따른 가역 정보 은닉 방법을 설명하기 위한 순서도이다.
- [0055] 먼저, 원본 문서 파일을 기설정된 압축 포맷으로 인코딩하여 압축 문서 파일을 출력한다(S510).
- [0056] 그리고, 압축 문서 파일의 비트스트림을 각각 기설정된 길이를 갖는 복수의 비트 블록으로 분할한다(S520).
- [0057] 그런 다음, 기설정된 은닉 정보 데이터를 기설정된 길이 단위로 분할한 복수의 은닉 정보 비트를 상기 복수의 비트 블록에 할당한다(S530).
- [0058] 그런 후, 할당된 은닉 정보 비트의 값에 기초하여 해당 비트 블록의 비트 값을 변환시킨다(S540).
- [0059] 구체적으로, 상기 단계 (S540)에서는 한 자릿수 비트인 은닉 정보 비트가 할당된 경우 은닉 정보 비트의 값에 따라 해당 비트 블록의 비트 값을 1비트 단위로 유지 또는 반전시킬 수 있다.
- [0060] 또한, 상기 단계(S540)에서는 적어도 두 자릿수 비트인 은닉 정보 비트가 할당된 경우, 각 은닉 정보 비트 값에 매칭된 난수 비트스트림을 적용하여 비트 블록의 비트 값을 변환시킬 수 있다. 이때, 비트 블록의 길이와 대응된 길이를 갖는 복수의 난수 비트스트림을 생성하고, 할당된 은닉 정보 비트의 값 별로 서로 상이한 난수 비트스트림을 매칭시킨 후, 할당된 은닉 정보 비트에 매칭된 난수 비트스트림 및 해당 비트 블록을 1 비트 단위로 배타적 논리합(XOR) 또는 배타적 부정 논리합(XNOR) 연산한다.
- [0061] 다음으로, 비트 값을 변환시킨 비트 블록들을 포함하는 정보 은닉 문서 파일을 가역 정보 추출 장치로 제공한다(S550).
- [0062] 이때, 상기 단계 (S550)에서는 은닉 정보 비트의 값 별로 매칭된 난수 비트스트림 및 비트 블록에 대해 기설정된 길이의 정보를 가역 정보 추출 장치(200)로 더 제공할 수 있다.

- [0063] 한편, 상기 (S520) 단계에서는 압축 문서 파일의 헤더를 제외한 페이로드의 비트스트림을 상기 복수의 비트 블록으로 분할하며, 이에 따라 상기 단계 (S530) 및 (S540)은 헤더를 제외한 페이로드의 비트 블록에 대해 처리될 수 있다.
- [0064] 또한, 상기 단계 (S550) 이전에, 원본 문서 파일의 비트스트림을 각각 기설정된 크기를 갖는 복수의 비트 블록으로 분할하고, 비트 블록 별로 포함된 문자에 대한 유의미 지수, 무의미 지수 및 문서 지수 중 적어도 하나에 대한 정보 지수 임계값을 산출하는 단계를 더 포함할 수 있다. 이때, 상기 단계 (S550)에서는, 가역 정보 추출 장치(200)로 정보 지수 임계값을 더 제공할 수 있다.
- [0065] 이하, 도 6 내지 도 8을 참조하여 본 발명의 일 실시예에 따른 가역 정보 추출 장치(200)의 구성 및 동작에 대해서 상세히 설명하도록 한다.
- [0066] 도 6은 본 발명의 일 실시예에 따른 가역 정보 추출 장치의 구성을 나타낸 블록도이다.
- [0067] 도 6에서와 같이, 본 발명의 일 실시예에 따른 가역 정보 추출 장치(200)는, 압축 문서 입력부(210), 복원 전처리부(220), 정보 분석부(230), 정보 분석 기준 데이터베이스(231), 은닉 정보 추출부(240) 및 문서 복원부(250)를 포함한다.
- [0068] 압축 문서 입력부(210)는 가역 정보 은닉 장치(100)로부터 압축 문서 파일에 은닉 정보 데이터가 인코딩된 정보 은닉 문서 파일을 입력받는다.
- [0069] 이때, 압축 문서 입력부(210)는 가역 정보 은닉 장치(100)로부터 원본 문서 파일의 비트 블록 당 정보 지수에 기초하여 설정된 정보 지수 임계값 및 비트 블록 단위에 대한 길이 정보를 더 입력 받을 수 있다.
- [0070] 복원 전처리부(220)는 정보 은닉 문서 파일을 기설정된 복원 포맷으로 디코딩하되, 기설정된 길이(즉, 'L')의 비트 블록 단위로 디코딩(이하, '1차 디코딩'이라고 지칭함)한다. 이때, 기설정된 복원 포맷은 가역 정보 은닉 장치(100)가 원본 문서 파일을 압축하는데 사용한 압축 포맷과 대응된 포맷이다.
- [0071] 정보 분석부(230)는 1차 디코딩된 비트 블록 별로 포함된 문자 및 기설정된 정보 분석 기준에 기초하여 정보 지수를 산출한다. 참고로, 정보 지수는 비트 블록에 포함된 문자 또는 문자의 집합 별 속성에 따른 값으로서, 유의미 지수, 무의미 지수 및 문서 지수를 포함한다.
- [0072] 구체적으로, 정보 분석부(230)는 '사전(dictionary)' 데이터를 정보 분석 기준으로 사용하여, 1차 디코딩된 비트 블록에 포함된 문자에 의한 단어 중 '사전'에 포함된 단어(이하, '유의미 단어'라고 지칭함)를 추출한다. 이때, 정보 분석부(230)는 정보 분석 기준 데이터베이스(231)에 저장되어 있는 사전 데이터를 사용하거나, 외부 사전 데이터베이스로부터 정보 분석 기준을 획득하여 사용할 수 있다. 또한, 정보 분석부(230)는 비트 블록의 문자에서 공백에 기초하여 단어를 구분할 수 있다. 즉, 비트 블록의 문자 중 두 빈칸 사이에 위치한 적어도 하나의 문자를 단어로 판단한다.
- [0073] 그리고, 정보 분석부(230)는 추출된 유의미 단어의 개수에 기초하여 유의미 지수를 산출한다. 예를 들어, 정보 분석부(230)는 해당 비트 블록의 문자에 포함된 단어 중 '사전'에 포함된 유의미 단어의 개수를 비트 블록의 문자에 포함된 전체 단어의 개수로 나눈 값을 유의미 지수로 산출할 수 있다.
- [0074] 또한, 정보 분석부(230)는 '컴퓨터 표준 코드'를 정보 분석 기준으로 사용하여, 1차 디코딩된 비트 블록의 문자 별 속성에 따른 무의미 지수 및 문서 지수를 산출한다. 이때, 정보 분석부(230)는 해당 비트 블록에 포함된 문자 별로 대응된 컴퓨터 표준 코드를 추출하고, 추출된 컴퓨터 표준 코드 중 기설정된 특정 코드에 대응된 문자의 개수에 기초하여 무의미 지수 및 문서 지수를 산출할 수 있다.
- [0075] 참고로, 정보 분석부(230)는 정보 분석 기준 데이터베이스(231)에 저장되어 있는 컴퓨터 표준 코드 테이블을 사용하거나, 외부 컴퓨터 표준 코드 데이터베이스로부터 정보 분석 기준을 획득하여 사용할 수 있다. 또한, 컴퓨터 표준 코드는 문자 속성에 따라 글자 코드, 숫자 코드, 부호 코드 및 제어 코드로 분류될 수 있다.
- [0076] 구체적으로, 도 7은 본 발명의 일 실시예에서 문서 파일의 정보 지수를 산출하는 방식을 설명하기 위한 아스키 코드 테이블이다.
- [0077] 아스키 코드(ASCII CODE)는 키보드에서 표현되는 모든 문자(예를 들어, 'A, a, B, b, C, c...' 등의 '글자', '0, 1, 2...' 등의 '숫자', '@, #, \$, ...' 등의 '부호' 및 '제어' 문자)의 코드를 포함한다. 도 7의 아스키 코드 테이블에 따르면, 코드 번호 '65'부터 '90'까지는 알파벳 대문자를 의미하고, 코드 번호 '97'부터 '122'까지는 알파벳 소문자를 의미하는 '글자' 코드이다. 이와 같은, 글자 코드

집합을 'ASCII-A' 라 지칭하도록 한다. 또한, 코드 번호 '48' 부터 '57' 까지는 '숫자' 를 의미하며, 이러한 숫자 코드 집합을 'ASCII-N' 이라 지칭하도록 한다. 또한, 코드 번호 '32' 부터 '126' 까지는 '부호' 를 의미하며, 이러한 부호 코드 집합을 'ASCII-S' 라 지칭하도록 한다. 그리고, 코드 번호 '10' 부터 '31' 까지 및 코드 번호 '127' 은 제어 문자를 의미하며, 이러한 제어 코드 집합을 'ASCII-C' 라 지칭하도록 한다.

[0078] 이때, 정보 분석부(230)는 해당 비트 블록에 포함된 문자 별로 대응하는 아스키 코드를 추출하고, 추출된 아스키 코드 중 ASCII-C에 대응된 문자 개수에 따라 무의미 지수를 산출할 수 있다.

[0079] 한편, 일반적으로 ASCII-C에 대응된 문자(이하, 'ASCII-C 문자' 라고 지칭함)는 일반적인 문서 파일(이하, '보통 문서 파일' 이라고 함)에서는 발견되지 않으며, 길이가 매우 긴 문서 등의 특별한 문서(이하, '특정 문서 파일' 이라고 함)에서 ASCII 코드 번호 '13' (즉, 'CR') 등의 ASCII-C 문자가 사용된다.

[0080] 이에 따라, 정보 분석부(230)는 해당 비트 블록에 포함된 문자 별로 대응하는 아스키 코드를 추출하고, 추출된 아스키 코드 중 ASCII-C 문자 및 ASCII-S 문자의 개수와 ASCII-A 문자 및 ASCII-N 문자의 개수 간의 관계에 기초하여 문서 지수를 산출한다. 예를 들어, 정보 분석부(230)는 ASCII-C 문자 및 ASCII-S 문자의 개수를 ASCII-A 문자 및 ASCII-N 문자의 개수로 나눈 값을 문서 지수로서 산출할 수 있다.

[0081] 참고로, 보통 문서 파일의 경우, ASCII-A와 ASCII-N의 수가 ASCII-C 및 ASCII-S의 수보다 월등히 많을 것으로 예상된다. 예를 들어, '소설' 및 '보고서' 등의 문서 파일에서는 ASCII-A 문자의 개수가 압도적으로 많으며, '회계 장부' 등의 문서 파일에서는 ASCII-N 문자의 개수가 압도적으로 많은 것이 정상적이다. 이에 따라, 정보 분석부(230)는 문서 파일의 종류에 따라 ASCII-C, ASCII-S, ASCII-A 및 ASCII-N 문자의 개수 간의 관계에 따른 연산을 처리할 수 있다.

[0082] 그리고, 정보 분석부(230)는 유의미 지수, 무의미지수 및 문서지수를 혼용하여 정보 지수를 산출하는 것도 가능하다.

[0083] 또한, 도 7에서는 정보 분석 기준으로서 아스키 코드 테이블만을 나타내었으나, 본 발명의 일 실시예에 따른 정보 분석부(230)는 정보 분석 기준으로서 유니코드 등의 다른 표준 코드들을 사용할 수 있다. 즉, 정보 분석 기준으로서 유니코드를 사용할 경우, 부호 코드 집합('UNICODE-C' 로 지칭할 수 있음)과 대응된 문자의 개수를 무의미 지수로 산출할 수 있다. 이때, 유니코드 및 아스키코드 등 복수의 표준 코드를 혼용하여 정보 분석 기준으로서 사용하는 것도 가능하다.

[0084] 다시 도 6으로 돌아가서, 은닉 정보 추출부(240)는 상기 1차 디코딩의 결과 및 상기 산출된 정보 지수에 기초하여 각 비트 블록에 은닉된 은닉 정보 비트를 추출한다. 그리고, 은닉 정보 추출부(240)는 비트 블록 별로 1차 디코딩 이전의 비트 값을 상기 추출된 은닉 정보 비트의 값에 기초하여 변환시키는 2차 디코딩을 처리한다.

[0085] 구체적으로, 정보 은닉 문서 파일을 상기 복원 포맷으로 1차 디코딩한 결과, 해당 비트 블록의 디코딩이 실패한 경우, 은닉 정보 비트는 해당 비트 블록의 비트 값을 반전시키도록 설정된 값을 갖는다.

[0086] 예를 들어, 가역 정보 은닉 장치(100)에서 압축 문서 파일의 비트 블록의 비트 값을 반전시키는 2차 인코딩을 수행하도록 설정된 은닉 정보 비트의 값이 '1' 인 경우, 상기 1차 디코딩 실패 시 해당 비트 블록에 은닉 정보 비트 '1' 이 은닉된 것을 알 수 있다. 이때, 은닉 정보 추출부(240)는 은닉 정보 비트 '1' 을 추출하고 해당 비트 블록의 비트 값을 1 비트 단위로 반전시킨다. 참고로, 해당 비트 블록의 비트 값을 반전시키는 은닉 정보 비트의 값은 '1' 또는 '0' 중 어느 하나로 설정될 수 있다.

[0087] 반면, 상기 1차 디코딩이 성공한 경우, 해당 비트 블록에 은닉된 은닉 정보 비트는 '1' 또는 '0' 일 가능성이 있다. 이에 따라, 은닉 정보 추출부(240)는 1차 디코딩이 성공된 비트 블록에 대해 산출된 정보 지수와 기 설정된 정보 지수 임계값을 비교한 결과에 기초하여 은닉 정보 비트를 추출한다. 참고로, 은닉 정보 추출부(240)는 압축 문서 입력부(210)를 통해 가역 정보 은닉 장치(100)로부터 입력된 정보 지수 임계값을 획득할 수 있다.

[0088] 이때, 은닉 정보 추출부(240)는 1차 디코딩된 비트 블록에 대해 산출된 정보 지수 중 유의미 지수, 무의미 지수 및 문서 지수 중 적어도 하나에 기초하여 은닉 정보 비트를 추출할 수 있다.

[0089] 구체적으로, 은닉 정보 추출부(240)는 산출된 유의미 지수가 기설정된 유의미 지수 임계값보다 일정 범위 이상으로 낮은 경우, 원본 문서 파일에 비해 해당 비트 블록에 유의미 단어가 일정 기준 이하로 적게 포함된 것으로 판단한다.

- [0090] 예를 들어, 도 4의 (b)의 비트 블록을 상기 1차 디코딩된 비트 블록인 것으로 가정하고, 도 4의 (a)의 비트 블록을 원본 문서 파일의 비트 블록인 것으로 가정할 수 있다. 이때, 원본 문서 파일의 비트 블록에 포함된 문자에 기초하여 산출된 유의미 지수 임계값에 비해, 1차 디코딩된 비트 블록에 포함된 문자에 기초하여 산출된 유의미 지수가 현저히 낮을 것을 예측할 수 있다.
- [0091] 이처럼, 산출된 유의미 지수가 유의미 지수 임계값에 비해 낮은 경우 1차 디코딩의 결과가 비성공적으로 판단할 수 있다. 즉, 해당 비트 블록이 원본 문서 파일의 비트 블록에서의 비트 값이 그대로 유지되지 않고 반전된 것으로 판단한다. 이에 따라, 은닉 정보 추출부(240)는 해당 비트 블록에 은닉된 은닉 정보 비트로서, 비트 값을 반전시키도록 사전에 설정된 값을 갖는 은닉 정보 비트를 추출한다. 그리고, 은닉 정보 추출부(240)는 해당 비트 블록의 1차 디코딩 이전의 비트 값을 1 비트 단위로 반전시킨다.
- [0092] 반면, 은닉 정보 추출부(240)는 산출된 유의미 지수가 유의미 지수 임계값과 동일 또는 일정 범위 이내로 높은 경우, 해당 비트 블록에 원본 문서 파일과 동일 또는 유사한 개수의 유의미 단어가 포함된 것으로 판단한다. 이에 따라, 은닉 정보 추출부(240)는 해당 비트 블록에 은닉된 은닉 정보 비트로서, 비트 블록의 비트 값을 유지시키도록 사전에 설정된 값을 갖는 은닉 정보 비트를 추출한다. 그리고, 은닉 정보 추출부(240)는 해당 비트 블록의 1차 디코딩 이전의 비트 값을 유지시킨다.
- [0093] 한편, 은닉 정보 추출부(240)는 1차 디코딩된 비트 블록에 대해 산출된 무의미 지수가 기설정된 무의미 지수 임계값 보다 일정 범위 이상으로 높은 경우, 해당 비트 블록에 원본 문서 파일에 비해 일정 기준 이상으로 많은 무의미 문자가 포함된 것으로 판단한다.
- [0094] 예를 들어, 상기 도 4의 (b)의 비트 블록은 두 개의 ASCII-N(즉, '5' 및 '4') 문자, 네 개의 ASCII-S(즉, '*', '^', '\$' 및 '?') 문자를 포함하고 있다. 반면, 도 4의 (a)의 비트 블록은 ASCII-N 문자와 ASCII-S 문자를 포함하고 있지 않다. 도 4의 (b)의 비트 블록을 상기 1차 디코딩된 비트 블록인 것으로 가정하고, 도 4의 (a)의 비트 블록을 원본 문서 파일의 비트 블록인 것으로 가정할 때, 1차 디코딩된 비트 블록은 원본 문서 파일의 비트 블록에 비해 다수의 ASCII-C 문자를 포함하고 있으므로, 1차 디코딩이 비성공적이라 판단된다. 이에 따라, 은닉 정보 추출부(240)는 1차 디코딩된 비트 블록으로부터, 비트 블록의 비트 값을 반전시키도록 사전에 설정된 값을 갖는 은닉 정보 비트를 추출할 수 있다. 그리고, 은닉 정보 추출부(240)는 추출된 은닉 정보 비트에 따라 해당 비트 블록의 1차 디코딩 이전의 비트 값을 1 비트 단위로 반전시킨다.
- [0095] 반면, 은닉 정보 추출부(240)는 산출된 무의미 지수가 무의미 지수 임계값 이하이거나 '0' 인 경우, 해당 비트 블록에 원본 문서 파일과 동일 또는 유사한 개수의 무의미 문자가 포함된 것으로 판단한다. 즉, 1차 디코딩이 성공적이라 판단한다. 이에 따라, 은닉 정보 추출부(240)는 1차 디코딩된 비트 블록으로부터, 비트 블록의 비트 값을 유지시키도록 사전에 설정된 값을 갖는 은닉 정보 비트를 추출한다. 그리고, 은닉 정보 추출부(240)는 추출된 은닉 정보 비트의 값에 따라 해당 비트 블록의 1차 디코딩 이전의 비트 값을 유지시킨다.
- [0096] 한편, 은닉 정보 추출부(240)는 1차 디코딩된 비트 블록에 대해 산출된 문서 지수가 기설정된 문서 지수 임계값 보다 일정 범위 이상으로 높은 경우, 해당 비트 블록에 원본 문서 파일에 비해 특정 문서에만 포함되는 특정 문자들이 일정 기준 이상으로 많이 포함된 것으로 판단한다. 즉, 해당 비트 블록이 원본 문서 파일의 비트 블록에서의 비트 값이 그대로 유지되지 않고 반전된 것으로 판단한다. 이에 따라, 은닉 정보 추출부(240)는 해당 비트 블록에 은닉된 은닉 정보 비트로서, 비트 값을 반전시키도록 사전에 설정된 값을 갖는 은닉 정보 비트를 추출한다. 그리고, 은닉 정보 추출부(240)는 해당 비트 블록의 1차 디코딩 이전의 비트 값을 1 비트 단위로 반전시킨다.
- [0097] 반면, 은닉 정보 추출부(240)는 산출된 문서 지수가 문서 지수 임계값 이하이거나 '0' 인 경우, 해당 비트 블록에 원본 문서 파일과 동일 또는 유사한 개수의 상기 특정 문자가 포함된 것으로 판단한다. 즉, 즉, 1차 디코딩이 성공적이라 판단한다. 이에 따라, 은닉 정보 추출부(240)는 1차 디코딩된 비트 블록으로부터, 비트 블록의 비트 값을 유지시키도록 사전에 설정된 값을 갖는 은닉 정보 비트를 추출한다. 그리고, 은닉 정보 추출부(240)는 추출된 은닉 정보 비트의 값에 따라 해당 비트 블록의 1차 디코딩 이전의 비트 값을 유지시킨다.
- [0098] 다시 도 6으로 돌아가서, 문서 복원부(250)는 은닉 정보 추출부(240)를 통해 상기 비트 값이 변환(즉, '반전' 또는 '유지')된 비트 블록을 상기 복원 포맷으로 디코딩하여 원본 문서 파일로 복원한다.
- [0099] 이하, 도 8을 참조하여 본 발명의 일 실시예에 따른 가역 정보 추출 방법에 대해서 상세히 설명하도록 한다.
- [0100] 도 8은 본 발명의 일 실시예에 따른 가역 정보 추출 방법을 설명하기 위한 순서도이다.

- [0101] 먼저, 가역 정보 은닉 장치(100)로부터 압축 문서 파일에 은닉 정보 데이터가 인코딩된 정보 은닉 문서 파일을 입력받는다(S810).
- [0102] 이때, 상기 단계 (S810)에서는, 원본 문서 파일의 비트 블록에 포함된 문자에 기초하여 산출된 정보 지수 임계값 및, 비트 블록의 길이 정보 중 적어도 하나를 더 입력 받을 수 있다.
- [0103] 그런 후, 기설정된 복원 포맷을 적용하여 정보 은닉 문서 파일을 기설정된 길이의 비트 블록 단위로 디코딩한다(S820).
- [0104] 다음으로, 상기 디코딩된 비트 블록에 포함된 복수의 문자에 기초하여 정보 지수를 산출한다(S830).
- [0105] 구체적으로, 상기 디코딩된 비트 블록에 포함된 문자를 기설정된 정보 분석 기준에 기초하여 분석하고, 분석한 결과에 따라 유의미 지수, 무의미 지수 및 문서 지수 중 적어도 하나의 정보 지수를 산출한다.
- [0106] 이때, 기설정된 '사전'을 정보 분석 기준으로 사용하여, 비트 블록에 포함된 복수의 문자에 의한 단어 중 '사전'에 포함된 유의미 단어를 추출하고, 유의미 단어의 개수에 기초하여 유의미 지수를 산출할 수 있다.
- [0107] 그리고, 기설정된 글자 코드, 숫자 코드, 부호 코드 및 제어 코드를 포함하는 컴퓨터 표준 코드를 정보 분석 기준으로 사용하여, 비트 블록에 포함된 복수의 문자 별로 대응하는 컴퓨터 표준 코드를 추출하고, 추출된 컴퓨터 표준 코드 중 부호 코드에 대응된 문자의 개수에 기초하여 무의미 지수를 산출할 수 있다.
- [0108] 또한, 기설정된 글자 코드, 숫자 코드, 부호 코드 및 제어 코드를 포함하는 컴퓨터 표준 코드를 정보 분석 기준으로 사용하여, 비트 블록에 포함된 복수의 문자 별로 대응하는 컴퓨터 표준 코드를 추출하고, 추출된 컴퓨터 표준 코드 중 글자 코드 및 숫자 코드 중 적어도 하나에 대응된 문자의 개수와 부호 코드 및 제어 코드 중 적어도 하나에 대응된 문자의 개수에 기초하여 문서 지수를 산출할 수 있다.
- [0109] 다음으로, 산출된 정보 지수와 기설정된 정보 지수 임계값을 비교한 결과에 기초하여 비트 블록에 은닉된 은닉 정보 비트를 추출한다(S840).
- [0110] 이때, 비트 블록 당 한 자릿수 비트 단위의 은닉 정보 비트를 추출하되, 산출한 정보 지수와 정보 지수 임계값을 비교한 결과에 따라 제 1 비트 값을 갖는 제 1 은닉 정보 비트를 추출하거나 또는 제 1 비트 값이 반전된 제 2 비트 값을 갖는 제 2 은닉 정보 비트를 추출할 수 있다.
- [0111] 그런 후, 비트 블록 별로 디코딩 이전의 비트 값을 은닉 정보 비트의 값에 기초하여 변환시킨다(S850).
- [0112] 구체적으로, 추출된 은닉 정보 비트의 값에 따라 비트 블록 별 비트 값을 1 비트 단위로 유지 또는 반전시킨다.
- [0113] 그런 후, 비트 값이 변환된 비트 블록을 상기 복원 포맷으로 디코딩하여 원본 문서 파일로 복원한다(S860).
- [0114] 이상 도 6 내지 도 8에서는, 가역 정보 추출 장치(200)가 정보 은닉 문서 파일의 비트 블록 당 한 자릿수 비트 단위의 은닉 정보 비트를 추출한 후 2차 디코딩한 것을 설명하였다.
- [0115] 그런데, 본 발명의 다른 실시예에 따른 가역 정보 추출 장치(200)는 정보 은닉 문서 파일의 비트 블록 당 두 자릿수 이상의 비트 단위의 은닉 정보 비트를 추출하여 2차 디코딩을 처리하는 것도 가능하다. 이때, 가역 정보 추출 장치(200)는 가역 정보 은닉 장치(100)로부터 은닉 정보 비트의 값 별로 매칭된 난수 비트스트림을 수신한다. 그리고, 가역 정보 추출 장치(200)는 정보 은닉 문서 파일의 각 비트 블록 별로 산출된 정보 지수와 기설정된 정보 지수 임계값을 비교한 결과, 및 난수 비트스트림에 기초하여 비트 블록 별 은닉 정보 비트를 추출한다. 또한, 가역 정보 추출 장치(200)는 추출된 비트 블록 별 은닉 정보 비트에 기초하여 정보 은닉 문서 파일에 대한 2차 디코딩을 수행한다.
- [0116] 본 발명의 실시예는 컴퓨터에 의해 실행되는 프로그램 모듈과 같은 컴퓨터에 의해 실행가능한 명령어를 포함하는 기록 매체의 형태로도 구현될 수 있다. 컴퓨터 판독 가능 매체는 컴퓨터에 의해 액세스될 수 있는 임의의 가용 매체일 수 있고, 휘발성 및 비휘발성 매체, 분리형 및 비분리형 매체를 모두 포함한다. 또한, 컴퓨터 판독 가능 매체는 컴퓨터 저장 매체 및 통신 매체를 모두 포함할 수 있다. 컴퓨터 저장 매체는 컴퓨터 판독가능 명령어, 데이터 구조, 프로그램 모듈 또는 기타 데이터와 같은 정보의 저장을 위한 임의의 방법 또는 기술로 구현된 휘발성 및 비휘발성, 분리형 및 비분리형 매체를 모두 포함한다. 통신 매체는 전형적으로 컴퓨터 판독가능 명령어, 데이터 구조, 프로그램 모듈, 또는 반송파와 같은 변조된 데이터 신호의 기타 데이터, 또는 기타 전송 매커니즘을 포함하며, 임의의 정보 전달 매체를 포함한다.
- [0117] 전술한 본 발명의 설명은 예시를 위한 것이며, 본 발명이 속하는 기술분야의 통상의 지식을 가진 자는 본 발명

의 기술적 사상이나 필수적인 특징을 변경하지 않고서 다른 구체적인 형태로 쉽게 변형이 가능하다는 것을 이해할 수 있을 것이다. 그러므로 이상에서 기술한 실시예들은 모든 면에서 예시적인 것이며 한정적이 아닌 것으로 이해해야만 한다. 예를 들어, 단일형으로 설명되어 있는 각 구성 요소는 분산되어 실시될 수도 있으며, 마찬가지로 분산된 것으로 설명되어 있는 구성 요소들도 결합된 형태로 실시될 수 있다.

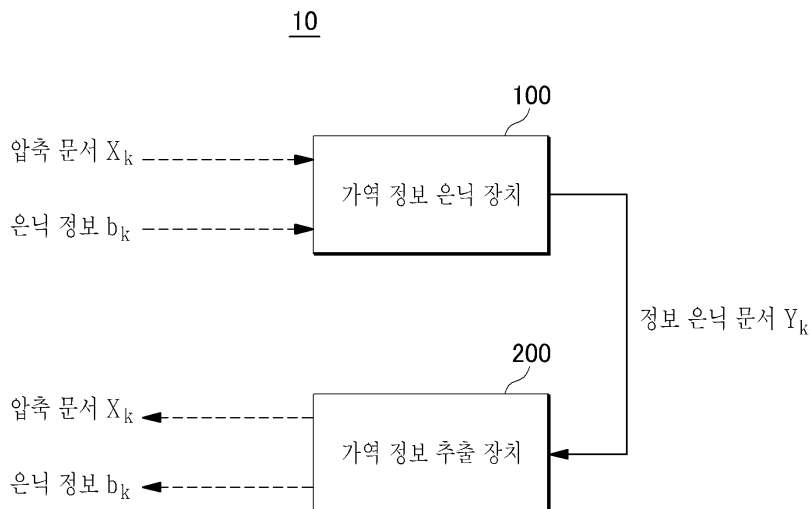
[0118] 본 발명의 범위는 상기 상세한 설명보다는 후술하는 특허청구범위에 의하여 나타내어지며, 특허청구범위의 의미 및 범위 그리고 그 균등 개념으로부터 도출되는 모든 변경 또는 변형된 형태가 본 발명의 범위에 포함되는 것으로 해석되어야 한다.

부호의 설명

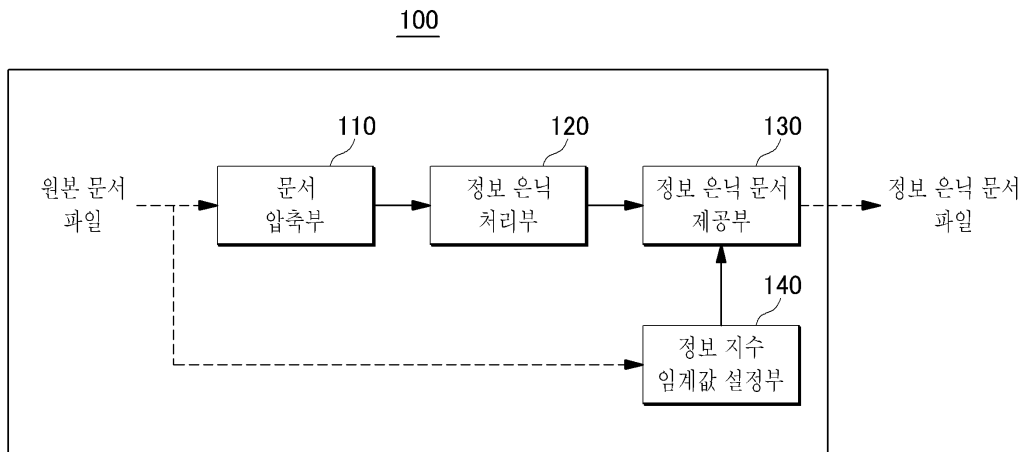
[0119] 10: 가역 정보 은닉 시스템
100: 가역 정보 은닉 장치
200: 가역 정보 추출 장치

도면

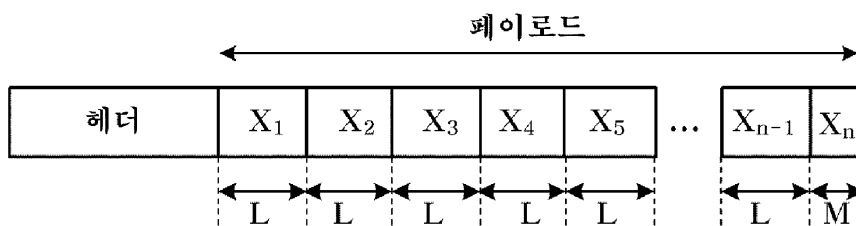
도면1



도면2



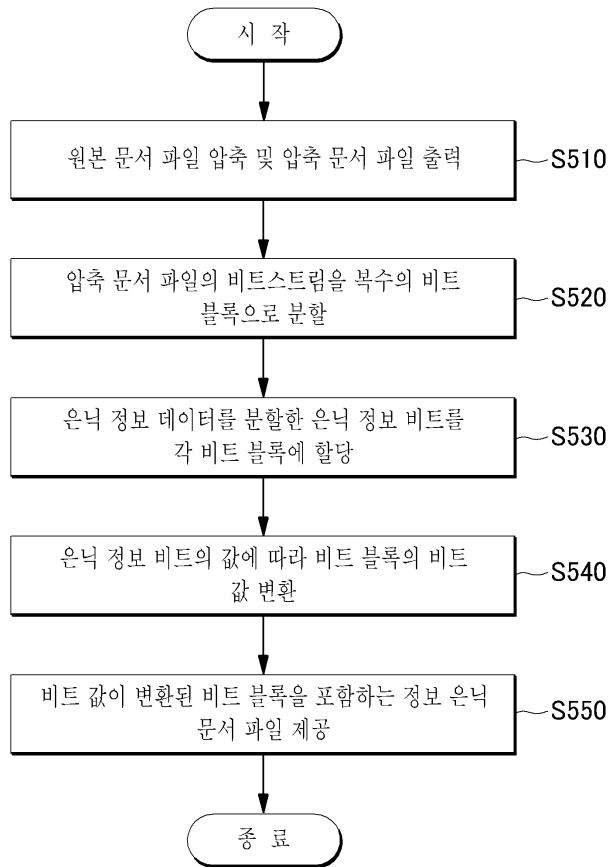
도면3



도면4

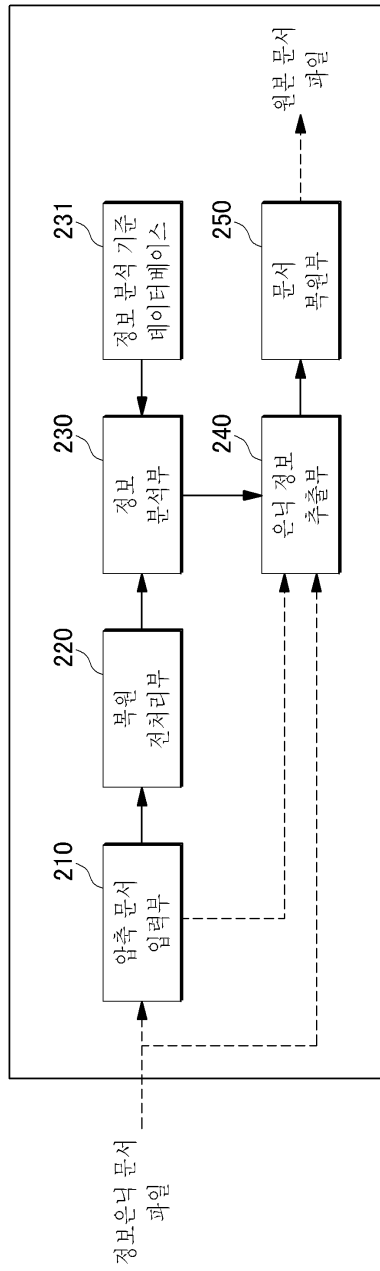
- (a) est friends are coming from London and Newf
- (b) Zisg u*ter las^g isteri&s Tra54os ?ilkHPest odf

도면5



도면6

200



도면7

Dec	Hx	Oct	Char	Dec	Hx	Oct	Html	Chr	Dec	Hx	Oct	Html	Chr	Dec	Hx	Oct	Html	Chr
0	0	000	NUL (null)	32	20	040	 	Space	64	40	100	@	@	96	60	140	`	`
1	1	001	SOH (start of heading)	33	21	041	!	!	65	41	101	A	A	97	61	141	a	a
2	2	002	STX (start of text)	34	22	042	"	"	66	42	102	B	B	98	62	142	b	b
3	3	003	ETX (end of text)	35	23	043	#	#	67	43	103	C	C	99	63	143	c	c
4	4	004	EOT (end of transmission)	36	24	044	$	\$	68	44	104	D	D	100	64	144	d	d
5	5	005	ENQ (enquiry)	37	25	045	%	%	69	45	105	E	E	101	65	145	e	e
6	6	006	ACK (acknowledge)	38	26	046	&	&	70	46	106	F	F	102	66	146	f	f
7	7	007	BEL (bell)	39	27	047	'	'	71	47	107	G	G	103	67	147	g	g
8	8	010	BS (backspace)	40	28	050	(	(	72	48	110	H	H	104	68	150	h	h
9	9	011	TAB (horizontal tab)	41	29	051	)	)	73	49	111	I	I	105	69	151	i	i
10	A	012	LF (NL line feed, new line)	42	2A	052	*	*	74	4A	112	J	J	106	6A	152	j	j
11	B	013	VT (vertical tab)	43	2B	053	+	+	75	4B	113	K	K	107	6B	153	k	k
12	C	014	FF (NP form feed, new page)	44	2C	054	,	,	76	4C	114	L	L	108	6C	154	l	l
13	D	015	CR (carriage return)	45	2D	055	-	-	77	4D	115	M	M	109	6D	155	m	m
14	E	016	SO (shift out)	46	2E	056	.	.	78	4E	116	N	N	110	6E	156	n	n
15	F	017	SI (shift in)	47	2F	057	/	/	79	4F	117	O	O	111	6F	157	o	o
16	10	020	DLE (data link escape)	48	30	060	0	0	80	50	120	P	P	112	70	160	p	p
17	11	021	DC1 (device control 1)	49	31	061	1	1	81	51	121	Q	Q	113	71	161	q	q
18	12	022	DC2 (device control 2)	50	32	062	2	2	82	52	122	R	R	114	72	162	r	r
19	13	023	DC3 (device control 3)	51	33	063	3	3	83	53	123	S	S	115	73	163	s	s
20	14	024	DC4 (device control 4)	52	34	064	4	4	84	54	124	T	T	116	74	164	t	t
21	15	025	NAK (negative acknowledge)	53	35	065	5	5	85	55	125	U	U	117	75	165	u	u
22	16	026	SYN (synchronous idle)	54	36	066	6	6	86	56	126	V	V	118	76	166	v	v
23	17	027	ETE (end of trans. block)	55	37	067	7	7	87	57	127	W	W	119	77	167	w	w
24	18	030	CAN (cancel)	56	38	070	8	8	88	58	130	X	X	120	78	170	x	x
25	19	031	EM (end of medium)	57	39	071	9	9	89	59	131	Y	Y	121	79	171	y	y
26	1A	032	SUB (substitute)	58	3A	072	:	:	90	5A	132	Z	Z	122	7A	172	z	z
27	1B	033	ESC (escape)	59	3B	073	;	;	91	5B	133	[	[	123	7B	173	{	{
28	1C	034	FS (file separator)	60	3C	074	<	<	92	5C	134	\	\	124	7C	174	|	!
29	1D	035	GS (group separator)	61	3D	075	=	=	93	5D	135	]	]	125	7D	175	}	}
30	1E	036	RS (record separator)	62	3E	076	>	>	94	5E	136	^	^	126	7E	176	~	~
31	1F	037	US (unit separator)	63	3F	077	?	?	95	5F	137	_	_	127	7F	177		DEL

도면8

