



(19)대한민국특허청(KR)
(12) 등록특허공보(B1)

(51) 。 Int. Cl. H03M 7/30 (2006.01)	(45) 공고일자 (11) 등록번호 (24) 등록일자	2007년07월10일 10-0737606 2007년07월03일
--	-------------------------------------	--

(21) 출원번호	10-2003-7010295	(65) 공개번호	10-2003-0085527
(22) 출원일자	2003년08월05일	(43) 공개일자	2003년11월05일
심사청구일자	2003년08월05일		
번역문 제출일자	2003년08월05일		
(86) 국제출원번호	PCT/EP2002/001333	(87) 국제공개번호	WO 2002/63775
국제출원일자	2002년02월04일	국제공개일자	2002년08월15일

(81) 지정국

국내특허 : 아랍에미리트, 안티구와바부다, 알바니아, 아르메니아, 오스트리아, 오스트레일리아, 아제르바이잔, 보스니아 헤르체고비나, 바베이도스, 불가리아, 브라질, 벨라루스, 벨리제, 캐나다, 스위스, 중국, 콜롬비아, 코스타리카, 쿠바, 체코, 독일, 덴마크, 도미니카, 알제리, 에쿠아도르, 에스토니아, 스페인, 핀란드, 영국, 그라나다, 그루지야, 가나, 감비아, 크로아티아, 헝가리, 인도네시아, 이스라엘, 인도, 아이슬란드, 일본, 케냐, 키르키즈스탄, 북한, 대한민국, 카자흐스탄, 세인트루시아, 스리랑카, 리베이라, 레소토, 리투아니아, 룩셈부르크, 라트비아, 모로코, 몰도바, 마다가스카르, 마케도니아공화국, 몽고, 말라위, 멕시코, 모잠비크, 노르웨이, 뉴질랜드, 필리핀, 폴란드, 포르투갈, 루마니아, 러시아, 수단, 스웨덴, 싱가포르, 슬로베니아, 슬로바키아, 시에라리온, 타지키스탄, 투르크멘, 터키, 트리니다드토바고, 탄자니아, 우크라이나, 우간다, 우즈베키스탄, 베트남, 세르비아 앤 몬테네그로, 남아프리카, 짐바브웨, 잠비아, 오만, 미국, 튀니지,

AP ARIPO특허 : 가나, 감비아, 케냐, 레소토, 말라위, 모잠비크, 수단, 시에라리온, 스와질랜드, 탄자니아, 우간다, 짐바브웨, 잠비아,

EA 유라시아특허 : 아르메니아, 아제르바이잔, 벨라루스, 키르키즈스탄, 카자흐스탄, 몰도바, 러시아, 타지키스탄, 투르크멘,

EP 유럽특허 : 오스트리아, 벨기에, 스위스, 독일, 덴마크, 스페인, 프랑스, 영국, 그리스, 아일랜드, 이탈리아, 룩셈부르크, 모나코, 네덜란드, 포르투갈, 스웨덴, 터키, 핀란드, 사이프러스,

OA OAPI특허 : 부르키나파소, 베닌, 중앙아프리카, 콩고, 코트디부아르, 카메룬, 가봉, 기니, 적도 기니, 기니 비사우, 말리, 모리타니, 니제르, 세네갈, 차드, 토고,

(30) 우선권주장 60/265,901 2001년02월05일 미국(US)

(73) 특허권자 이엑스피웨이
프랑스공화국, 에프-51100 행스, 뒤 가브리엘 부와쟁, 8비스, 뿔 떼끄놀로지끄 파르망, 르 크리스탈리드

(72) 발명자 셰이라 끌로드
12, 뒤 롤링, 에프-75005 파리 프랑스

띠에노 셰드리
115, 뒤 오베르캅뽀, 75011 파리, 프랑스

(74) 대리인 허용록

(56) 선행기술조사문헌
EP1069715
WO9734240

US05794030 A1

심사관 : 강윤석

전체 청구항 수 : 총 15 항

(54) 구조화된 문서의 부호화 방법 및 복호화 방법

(57) 요약

복호기에 접근할 수 없는 적어도 제1 스키마에서 정의되는 구조를 가지며, 적어도 상기 복호기에 접근할 수 있는 제2 스키마의 변화로부터 기인하는 구조화된 문서를 상기 복호기가 적어도 부분적으로 복호화할 수 있도록 하는 부호화 방법에 있어서, 상기 구조화된 문서는 서로 포위되는 정보 요소들을 포함하며, 상기 문서의 상기 정보 요소들은 정보 요소들의 각각의 요소구조를 각각 정의하는 각각의 요소 유형과 적어도 제1 및 제2 스키마에서 관련되며, 상기 제1 스키마는 복호기에 접근할 수 없고 상기 제2 스키마는 상기 복호기에 접근할 수 있으며, 상기 제1 스키마는 상기 제2 스키마에서 정의되는 해당 요소로부터 유도되는 적어도 하나의 유도되는 정보 요소를 정의하며, 상기 부호화 방법은, 상기 제1 및 제2 스키마를 이용하여, 상기 문서의 각 정보 요소에 대하여 상기 정보 요소를 부호화하는 이진 시퀀스를 포함하는 이진 스트림으로 상기 문서를 부호화하는 단계; 및 상기 유도된 정보 요소를 부호화하는 상기 이진 시퀀스에 상기 유도된 요소의 구조를 정의하는 상기 제1 스키마를 가리키는 참조를 삽입하는 단계;가 포함되고, 상기 문서의 각각의 요소를 부호화하는 상기 이진 시퀀스는 상기 요소의 부호화된 값을 포함하는 내용 필드와, 상기 내용 필드의 길이를 정의하는 길이 필드로 이루어지고, 상기 제1 스키마를 가리키는 상기 참조는 상기 문서를 부호화하는 데에 사용되는 모든 스키마에 대한 참조를 포함하는 스키마 참조 리스트에서 정의되고, 상기 스키마 참조 리스트는 상기 복호기에 접근할 수 있도록 만들어진다.

대표도

도 1

특허청구의 범위

청구항 1.

복호기에 접근할 수 없는 적어도 제1 스키마에서 정의되는 구조를 가지며, 적어도 상기 복호기에 접근할 수 있는 제2 스키마의 변화로부터 기인하는 구조화된 문서를 상기 복호기가 적어도 부분적으로 복호화할 수 있도록 하는 부호화 방법에 있어서, 상기 구조화된 문서는 서로 포위되는 정보 요소들을 포함하며, 상기 문서의 상기 정보 요소들은 정보 요소들의 각각의 요소구조를 각각 정의하는 각각의 요소 유형과 적어도 제1 및 제2 스키마에서 관련되며, 상기 제1 스키마는 복호기에 접근할 수 없고 상기 제2 스키마는 상기 복호기에 접근할 수 있으며, 상기 제1 스키마는 상기 제2 스키마에서 정의되는 해당 요소로부터 유도되는 적어도 하나의 유도되는 정보 요소를 정의하며,

상기 부호화 방법은,

상기 제1 및 제2 스키마를 이용하여, 상기 문서의 각 정보 요소에 대하여 상기 정보 요소를 부호화하는 이진 시퀀스를 포함하는 이진 스트림으로 상기 문서를 부호화하는 단계; 및

상기 유도된 정보 요소를 부호화하는 상기 이진 시퀀스에 상기 유도된 요소의 구조를 정의하는 상기 제1 스키마를 가리키는 참조를 삽입하는 단계;가 포함되고,

상기 문서의 각각의 요소를 부호화하는 상기 이진 시퀀스는 상기 요소의 부호화된 값을 포함하는 내용 필드와, 상기 내용 필드의 앞에 위치되어 상기 내용 필드 길이의 부호화된 값을 포함하는 길이 필드로 이루어지고,

상기 제1 스키마를 가리키는 상기 참조는 상기 문서를 부호화하는 데에 사용되는 모든 스키마에 대한 참조를 포함하는 스키마 참조 리스트에서 정의되고, 상기 스키마 참조 리스트는 상기 복호기에 접근할 수 있도록 만들어지는 것을 특징으로 하는 부호화 방법.

청구항 2.

제 1 항에 있어서,

상기 유도되는 정보 요소는 상기 제 2 스키마에서 상기 해당 정보 요소의 구조유형에 대하여 제한된 구조유형에 제 1 스키마에서 연관되며, 상기 유도된 요소들을 부호화하는 상기 이진 시퀀스는 내용필드를 포함하며, 상기 내용필드와 상기 제 1 스키마를 가리키는 참조와 상기 유도된 요소의 상기 구조유형을 가리키는 참조는 상기 내용 필드에 부착되고 상기 제 2 스키마에서 정의되는 것을 특징으로 하는 부호화 방법.

청구항 3.

제 1 항 또는 제 2 항에 있어서,

상기 유도되는 정보 요소는 상기 제2 스키마에서 상기 해당 정보 요소의 구조유형에 대하여 확장된 구조유형에 제1 스키마에서 연관되며, 상기 유도된 정보 요소의 구조유형은 제2 스키마에서 정의된 해당 정보 요소의 구조유형을 가지는 제1 부분과 상기 유도된 정보 요소에 대하여 특별하고 상기 제1 스키마에서 정의된 구조유형을 가지는 제2 부분을 포함하며, 상기 이진 시퀀스는 내용 필드를 포함하는 상기 유도된 요소들을 부호화하며, 상기 내용 필드는,

상기 제2 스키마를 가리키는 참조를 포함하는 필드와,

상기 제2 스키마에서 상기 해당 요소의 구조 유형을 가리키는 구조 유형 참조를 포함하는 필드와,

상기 제1 부분의 부호화된 값을 포함하는 필드와,

상기 제1 스키마를 가리키는 참조를 포함하는 필드와,

상기 제2 부분의 상기 구조 유형을 가리키는 구조 유형 참조를 포함하는 필드와,

상기 제2 부분의 부호화된 값을 포함하는 필드를 포함하는 것을 특징으로 하는 부호화 방법.

청구항 4.

제 1 항 또는 제 2 항에 있어서,

정보 요소를 부호화하는 이진 시퀀스는 정보 요소의 이름이 바뀌는지를 나타내는 대체 플래그를 포함하는 대체 필드를 포함하며, 상기 대체 플래그가 변화를 나타내면 요소이름 참조 필드는 상기 정보 요소의 새로운 이름을 지칭하는 참조를 포함하고, 스키마 참조 필드는 상기 새로운 이름 참조가 정의되는 스키마를 가리키는 참조를 포함하는 것을 특징으로 하는 부호화 방법.

청구항 5.

제 1 항 또는 제 2 항에 있어서,

상기 부호화된 문서에서 하나의 정보 요소를 부호화하는 상기 이진 시퀀스는 스키마 상태모드 필드를 포함하며, 상기 스키마 상태모드 필드는,

상기 정보 요소가 상기 제2 스키마에서 해당 요소에 대한 제1 스키마에서 변하지 않는다는 것을 나타내는 제1 상태와,

상기 정보 요소의 어떠한 부요소들도 상기 제2 스키마에서 상기 해당요소에 대하여 상기 제1 스키마에서 변하지 않는다는 것을 나타내는 제2 상태와,

상기 제2 스키마에서 상기 해당 요소에 대하여 상기 제1 스키마에서 상기 정보 요소가 변한다는 것을 나타내는 제3 상태를 포함하며,

상기 스키마 상태 모드 필드가 제1 상태에 있을 때에 상기 부호화된 정보 요소는 어떤 스키마 참조와 어떤 다른 변화 정보라도 포함하고, 상기 스키마 상태 모드 필드가 제2 상태에 있을 때에 상기 부호화된 정보 요소의 어떤 부요소도 어떤 스키마 참조와 어떤 다른 변화 정보라도 포함하지 않는 것을 특징으로 하는 부호화 방법.

청구항 6.

제 1 항 또는 제 2 항에 있어서,

상기 부호화된 문서에서 적어도 하나의 정보 요소를 부호화하는 상기 이진 시퀀스는 스키마 상태 모드 필드를 포함하고, 상기 스키마 상태 모드 필드는,

상기 제2 스키마에서 해당 요소에 대하여 상기 제1 스키마에서 상기 정보 요소가 변하지 않는다는 것을 나타내는 제1 상태와,

상기 정보 요소의 어떠한 부요소들도 상기 제2 스키마에서 상기 해당요소에 대하여 상기 제1 스키마에서 변하지 않는다는 것을 나타내는 제2 상태와,

상기 제2 스키마에서 상기 해당 요소에 대하여 상기 제1 스키마에서 상기 정보 요소가 변한다는 것을 나타내는 제3 상태와,

상기 제2 스키마에서 상기 해당 요소에 대하여 상기 제1 스키마에서 상기 정보 요소가 변한다는 것과 상기 정보 요소의 어떠한 부요소들도 상기 제2 스키마에서 상기 해당요소에 대하여 상기 제1 스키마에서 변하지 않는다는 것을 나타내는 제4 상태를 포함하며,

상기 스키마 상태 모드 필드가 상기 제1 상태일 때에 상기 부호화된 정보요소는 어떤 스키마 참조와 다른 어떤 변화 정보라도 포함하고, 상기 스키마 상태 모드 필드가 상기 제2 상태이거나 제4 상태일 때에 상기 정보요소의 어떤 부요소도 스키마 참조와 다른 어떤 변화 정보라도 포함하지 않는 것을 특징으로 하는 부호화 방법.

청구항 7.

제 1 항 또는 제 2 항에 있어서,

상기 구조화된 문서를 부호화하는 데에 사용되는 모든 스키마들을 가리키는 참조들을 포함하는 상기 스키마 참조 리스트는 상기 구조화된 문서를 부호화하는 상기 이진 스트림과 연관된 헤더에 삽입되는 것을 특징으로 하는 부호화 방법.

청구항 8.

복호기에 접근할 수 없는 적어도 제1 스키마에서 정의되는 구조를 가지며, 적어도 상기 복호기에 접근할 수 있는 제2 스키마의 변화로부터 기인하는 구조화된 문서를 부호화하는 이진 스트림을 적어도 부분적으로 복호화하기 위한 복호화 방법에 있어서, 상기 구조화된 문서는 서로 포위되는 정보 요소들을 포함하며, 상기 문서의 상기 정보 요소들은 정보 요소들의 각각의 요소구조를 각각 정의하는 각각의 요소 유형과 적어도 제1 및 제2 스키마에서 관련되며, 상기 제1 스키마는 복호기에 접근할 수 없고 상기 제2 스키마는 상기 복호기에 접근할 수 있으며, 상기 제1 스키마는 상기 제2 스키마에서 정의되는 해당 요소로부터 유도되는 적어도 하나의 유도되는 정보 요소를 정의하며,

상기 복호화 방법은,

상기 문서의 각 정보 요소를 부호화하는 이진 시퀀스를 상기 이진 스트림에서 검출하기 위하여, 상기 제2 스키마를 이용하여 상기 구조화된 문서를 부호화하는 이진 스트림을 순차적으로 읽는 단계;

상기 복호기에 의해 알려진 스키마 참조에서 정의된 바와 같이 상기 제1 스키마를 가리키는 참조를 부호화된 정보 요소의 각각의 검출된 이진 시퀀스에서 검출하는 단계;

상기 검출된 이진 시퀀스에서 상기 제1 스키마를 가리키는 상기 참조가 검출되지 않으면, 상기 검출된 이진 시퀀스로 부호화된 상기 정보 요소의 상기 제2 스키마에서 정의된 상기 구조에 따라 상기 검출된 이진 시퀀스를 복호화하는 단계; 및

상기 검출된 이진 시퀀스에서 상기 제1 스키마를 가리키는 상기 참조가 검출되면, 상기 제1 스키마에서 정의된 구조를 갖는 구조화된 문서의 정보 요소를 부호화하는 상기 제1 스키마 이진 데이터를 상기의 검출된 참조로부터 구별하고, 상기 이진 스트림을 순차적으로 읽는 동안에 상기 구별된 이진 데이터를 건너뛰고, 상기 정보 요소를 부호화하는 이진 시퀀스의 끝을 결정하기 위한 길이 부호화된 값을 사용하는 단계;가 포함되고,

상기 문서의 각각의 요소를 부호화하는 상기 이진 시퀀스는 상기 요소의 부호화된 값을 포함하는 내용 필드와, 상기 내용 필드의 앞에 위치되어 상기 내용 필드 길이의 부호화된 값을 포함하는 길이 필드로 이루어지는 것을 특징으로 하는 복호화 방법.

청구항 9.

제 8 항에 있어서,

상기 제1 스키마를 가리키는 참조를 포함하는 상기 이진 시퀀스 내에서 길이 코드 값을 읽고 복호화하는 단계와,

상기 제1 스키마를 가리키는 상기 참조의 상기 이진 시퀀스 내에서 상기 복호화된 길이 값과 상기 위치의 기능으로서 건너뛰기 위한 이진 데이터의 길이를 결정하는 단계를 더 포함하는 것을 특징으로 하는 복호화 방법.

청구항 10.

제 8 항에 있어서,

상기 유도되는 정보 요소는 상기 제2 스키마에서 상기 해당 정보 요소의 구조 유형에 대하여 제한된 구조 유형에 제1 스키마에서 연관되며, 상기 유도된 요소들을 부호화하는 상기 이진 시퀀스는 내용필드를 포함하며, 상기 내용필드와 상기 제1 스키마를 가리키는 참조와 상기 유도된 요소의 상기 구조유형을 가리키는 참조는 상기 내용필드에 부착되고 상기 제2 스키마에서 정의되는 것을 특징으로 하는 복호화 방법.

청구항 11.

제 8 항 또는 제 10 항에 있어서,

상기 유도되는 정보 요소는 상기 제 2 스키마에서 상기 해당 정보 요소의 구조유형에 대하여 확장된 구조유형에 제 1 스키마에서 연관되며, 상기 유도된 정보 요소의 구조유형은 제 2 스키마에서 정의된 해당 정보 요소의 구조유형을 가지는 제1 부분과 상기 유도된 정보 요소에 대하여 특별하고 상기 제1 스키마에서 정의된 구조유형을 가지는 제2 부분을 포함하며, 상기 이진 시퀀스는 내용 필드를 포함하는 상기 유도된 요소들을 부호화하며, 상기 내용 필드는,

상기 제 2 스키마를 가리키는 참조를 포함하는 필드와,

상기 제 2 스키마에서 상기 해당 요소의 구조 유형을 가리키는 구조 유형 참조를 포함하는 필드와,

상기 제 1 부분의 부호화된 값을 포함하는 필드와,

상기 제 1 스키마를 가리키는 참조를 포함하는 필드와,

상기 제 2 부분의 상기 구조 유형을 가리키는 구조 유형 참조를 포함하는 필드와,

상기 제 2 부분의 부호화된 값을 포함하는 필드를 포함하는 것을 특징으로 하는 복호화 방법.

청구항 12.

제 8 항 또는 제 10 항에 있어서,

상기 유도된 정보 요소는 상기 제 2 스키마에서 상기 해당 정보 요소의 이름에 대하여 변하는 이름을 제 1 스키마에서 가지고, 상기 유도된 요소를 부호화하는 상기 이진 시퀀스는 상기 유도된 정보 요소의 이름이 바뀌는지를 나타내는 대체 플래그를 포함하는 대체 필드를 포함하며, 상기 대체 플래그가 변화를 나타내면 스키마 참조 필드는 상기 제 1 스키마를 가리키는 참조를 포함하고, 상기 제1 스키마에서 상기 유도 정보 요소의 이름을 지칭하는 요소 이름 참조를 포함하는 것을 특징으로 하는 복호화 방법.

청구항 13.

제 8 항 또는 제 10 항에 있어서,

상기 부호화된 문서에서 하나의 정보 요소를 부호화하는 상기 이진 시퀀스는 스키마 상태모드 필드를 포함하며, 상기 스키마 상태모드 필드는,

상기 정보 요소가 상기 제2 스키마에서 해당 요소에 대한 제1 스키마에서 변하지 않는다는 것을 나타내는 제 1 상태와,

상기 정보 요소의 어떠한 부요소들도 상기 제2 스키마에서 상기 해당요소에 대하여 상기 제1 스키마에서 변하지 않는다는 것을 나타내는 제 2 상태와,

상기 제2 스키마에서 상기 해당 요소에 대하여 상기 제1 스키마에서 상기 정보 요소가 변한다는 것을 나타내는 제 3 상태를 포함하며,

상기 스키마 상태 모드 필드가 제 1 상태에 있을 때에 상기 부호화된 정보 요소는 어떤 스키마 참조와 어떤 다른 변화 정보도 포함하지 않고, 상기 스키마 상태 모드 필드가 제 2 상태에 있을 때에 상기 부호화된 정보 요소의 어떤 부요소도 어떤 스키마 참조와 어떤 다른 변화 정보라도 포함하지 않는 것을 특징으로 하는 복호화 방법.

청구항 14.

제 8 항 또는 제 10 항에 있어서,

상기 부호화된 문서에서 적어도 하나의 정보 요소를 부호화하는 상기 이전 시퀀스는 스키마 상태 모드 필드를 포함하고, 상기 스키마 상태 모드 필드는,

상기 제2 스키마에서 해당 요소에 대하여 상기 제1 스키마에서 상기 정보 요소가 변하지 않는다는 것을 나타내는 제 1 상태와,

상기 정보 요소의 어떠한 부요소들도 상기 제2 스키마에서 상기 해당요소에 대하여 상기 제1 스키마에서 변하지 않는다는 것을 나타내는 제 2 상태와,

상기 제2 스키마에서 상기 해당 요소에 대하여 상기 제1 스키마에서 상기 정보 요소가 변한다는 것을 나타내는 제 3 상태와,

상기 제2 스키마에서 상기 해당 요소에 대하여 상기 제1 스키마에서 상기 정보 요소가 변한다는 것과 상기 정보 요소의 어떠한 부요소들도 상기 제2 스키마에서 상기 해당요소에 대하여 상기 제1 스키마에서 변하지 않는다는 것을 나타내는 제 4 상태를 포함하며,

상기 스키마 상태 모드 필드가 상기 제 1 상태일 때에 상기 부호화된 정보요소는 어떤 스키마 참조와 다른 어떤 변화 정보라도 포함하고, 상기 스키마 상태 모드 필드가 상기 제 2 상태이거나 제 4 상태일 때에 상기 정보요소의 어떠한 부요소도 스키마 참조와 다른 어떤 변화 정보라도 포함하지 않는 것을 특징으로 하는 복호화 방법.

청구항 15.

제 8 항 또는 제 10 항에 있어서,

상기 구조화된 문서를 부호화하는 데에 사용되는 모든 스키마들을 가리키는 참조들을 포함하는 상기 스키마 참조 리스트는 상기 구조화된 문서를 부호화하는 상기 이전 스트림과 연관된 헤더 내에서 읽히는 것을 특징으로 하는 복호화 방법.

청구항 16.

삭제

청구항 17.

삭제

명세서

기술분야

본 발명은 일반적인 컴퓨터 시스템의 분야에 관한 것으로서, SGML(Standard Generalized Markup Language) 및 XML(eXtensible Markup Language)과 같은 일반화된 마크업 언어로 된 문서 표현을 이용한 구조화된 문서를 부호화하는 방법 및 복호화하는 방법에 관한 것이다. 그러한 문서는 미디어 정보를 포함할 수도 있다.

배경기술

수 년 내에 컴퓨터 네트워크는 통신을 위한 주 미디어가 되었다. 컴퓨터는 이제 공유 네트워크라고 해도 과언이 아니며, 운영체제는 응용 프로그램들이 메시지들을 쉽게 교환할 수 있게 해주며, 인터넷 인프라구조는 컴퓨터들이 그들의 대리자를 발견하게 하고, 응용 프로그램들은 복잡한 알고리즘들을 이용하여 자신들을 동기화한다.

상호 작동성의 배경에서는, 일반화된 마크업 언어들이 문서처리를 다루는 솔루션을 제공한다. 사실은 문서의 구조는 문서 사용에 있어서 주된 역할을 수행한다. 문서를 포매팅하고 인쇄하고 색인하는 것은 그 구조를 따라서 실제적으로 이루어진

다. SGML은 애초부터 문서 표현을 문서 구조와 내용으로부터 쉽게 분리시키도록 만들어졌다. 구조들을 부호화하는 능력 때문에 XML은 비문서 응용 프로그램에 관심이 있는 다른 사람들로부터 관심을 끌어들였다. XML 청중은 더 넓어져서 전자 상거래, 데이터베이스 지식 표현 집단을 (다른 것들 중에서) 포함한다.

XML과 더 일반적으로 마크업 언어들은 현재 문서들(메타데이터)을 표현하고 구조화하는 데에 넓게 사용된다. 구조화된 문서는 서로 포위되는 몇 개의 정보 요소들을 포함한다. 정보 요소들은 태그에 의해 서로 구별되고 분리된다. 그 태그들은 정보 요소들의 요소 유형들을 나타낸다. 구조화된 문서는 전체 문서를 나타내고 문서의 시작과 끝을 표시하는 태그로 구분된 제1 정보 요소 즉 기본 요소를 일반적으로 포함한다. 이 제1 요소는 예를 들어 텍스트의 패러그래프와 같은 정보 부요소를 포함한다. 각 정보 부요소는 요소의 시작과 끝을 나타내는 태그에 의해 구별된다. 태그들은 정보 요소의 하나 이상의 특성을 나타내는 태그 속성과 관련될 수도 있다.

태그 내용은 일반적으로 사용자에게 의해 표시되거나 조작되도록 의도되는 정보를 나타낸다. 태그 내용은 선택적이거나 태그의 유형에 따라 요구될 수도 있고, 태그들에 의해 차례로 제한되는 다른 포위되는 정보 부요소를 포함하고 내용과 속성을 가질 수도 있다.

구조화된 문서는 "유효"라고 생각되도록 구조화된 문서가 검증해야 하는 규칙을 반영하는 스키마와 연관될 수도 있다. 그것은 디폴트 값들, 요소와 속성들 유형들 및 유형 계층들에 대한 정보를 포함한다. 유효성은 수신된 문서가 스키마에 맞는 것을 보장한다. 따라서 그 수신된 문서는 의도하는 의미를 갖는다. 더욱이, 그것은 특성 즉, 각 기술 아이템(정보 요소 또는 속성들)의 유형을 결정한다. XML 표준은 같은 구조를 가지는 XML 문서의 클래스를 위한 문법을 특정하도록 설계된 XML 스키마 언어를 포함한다.

그러나 XML은 장황한 언어이다. 따라서 비효과적으로 처리되고 값 비싸게 전송된다. 이러한 이유로, ISO/IEC 15938-1과 더 자세히는 MPEG-7(Moving Picture Expert Group)은 구조화된 문서의 기술을 부호화(압축)하고 그러한 이진 포맷을 복호화하는 방법과 이진 포맷을 제안한다. 이 표준은 더욱이 멀티미디어 데이터와 같이 고도로 구조화된 데이터를 다루도록 설계된다.

압축 효율을 얻기 위하여, 이 방법은 스키마 분석 상태에 기초한다. 이러한 상태 동안에 내부 표들은 이진 코드를 각각의 XML 요소들과 유형들과 속성들에 관련시키도록 계산된다. 이 방법은 부호기와 해당 복호기에 의해 같은 스키마에 대한 충분한 지식을 요구한다.

구조화된 문서를 부호화하는 데에 쓰이는 스키마는 확장되기를 요구할 때에 최선의 솔루션은 복호기에 유용한 확장 스키마를 만들어야 한다. 그러나 특별한 경우에, 확장 스키마에 접근하기 위하여 복호기를 쉽게 갱신하는 것은 가능하지 않다.

발명의 상세한 설명

발명의 요약

본 발명의 목적은 스키마에 필요한 모든 것이 복호기에 의해 알려지지 않더라도 문서가 부분적으로 복호화될 수 있는 방법으로 구조화된 문서를 부호화하는 방법을 제공하는 데에 있다.

본 발명의 다른 목적은 전방 및 후방 호환성을 보장하는 즉, 복호기에 접근할 수 없는 적어도 제1 스키마에서 정의되는 구조를 가지며, 적어도 상기 복호기에 접근할 수 있는 제2 스키마의 변화로부터 기인하는 구조화된 문서를 상기 복호기가 적어도 부분적으로 복호화할 수 있도록 하는 부호화 방법으로서, 상기 구조화된 문서는 서로 포위되는 정보 요소들을 포함하며, 상기 문서의 상기 정보 요소들은 정보 요소들의 각각의 요소구조를 각각 정의하는 각각의 요소 유형과 적어도 제1 및 제2 스키마에서 관련되며, 상기 제1 스키마는 복호기에 접근할 수 없고 상기 제2 스키마는 상기 복호기에 접근할 수 있으며, 상기 제1 스키마는 상기 제2 스키마에서 정의되는 해당 요소로부터 유도되는 적어도 하나의 유도되는 정보 요소를 정의하는 부호화 방법을 제공하는 데에 있다.

본 발명에 따르면, 부호화 방법은 상기 제1 및 제2 스키마를 이용하여, 상기 문서의 각 정보 요소에 대하여 상기 정보 요소를 부호화하는 이진 시퀀스를 포함하는 이진 스트림으로 상기 문서를 부호화하는 단계와, 상기 유도된 정보 요소를 부호화하는 상기 이진 시퀀스에 상기 유도된 요소의 구조를 정의하는 상기 제1 스키마를 가리키는 참조를 삽입하는 단계를 포함하며, 상기 제1 스키마를 가리키는 상기 참조는 상기 문서를 부호화하는 데에 사용되는 모든 스키마에 대한 참조를 포함하는 스키마 참조 리스트에서 정의되고 상기 스키마 참조 리스트는 상기 복호기에 접근할 수 있도록 만들어진다.

본 발명의 관점에 따르면, 상기 문서의 각각의 요소를 부호화하는 상기 이진 시퀀스는, 상기 요소의 부호화된 값을 포함하는 내용 필드와, 상기 내용 필드 앞에 위치하고 상기 내용 필드의 길이의 부호화된 값을 포함하는 길이 필드를 포함한다.

본 발명의 다른 관점에 따르면, 상기 유도되는 정보 요소는 상기 제2 스키마에서 상기 해당 정보 요소의 구조유형에 대하여 제한된 구조유형에 제1 스키마에서 연관되며, 상기 유도된 요소들을 부호화하는 상기 이진 시퀀스는 내용필드를 포함하며, 상기 내용필드와 상기 제1 스키마를 가리키는 참조와 상기 유도된 요소의 상기 구조유형을 가리키는 참조는 상기 내용 필드에 부착되고 상기 제2 스키마에서 정의된다.

본 발명의 또 다른 관점에 따르면, 상기 유도되는 정보 요소는 상기 제2 스키마에서 상기 해당 정보 요소의 구조유형에 대하여 확장된 구조유형에 제1 스키마에서 연관되며, 상기 유도된 정보 요소의 구조유형은 제2 스키마에서 정의된 해당 정보 요소의 구조유형을 가지는 제1 부분과 상기 유도된 정보 요소에 대하여 특별하고 상기 제1 스키마에서 정의된 구조유형을 가지는 제2 부분을 포함하며, 상기 이진 시퀀스는 내용 필드를 포함하는 상기 유도된 요소들을 부호화하며, 상기 내용 필드는, 상기 제2 스키마를 가리키는 참조를 포함하는 필드와, 상기 제2 스키마에서 상기 해당 요소의 구조 유형을 가리키는 구조 유형 참조를 포함하는 필드와, 상기 제1 부분의 부호화된 값을 포함하는 필드와, 상기 제1 스키마를 가리키는 참조를 포함하는 필드와, 상기 제2 부분의 상기 구조 유형을 가리키는 구조 유형 참조를 포함하는 필드와, 상기 제2 부분의 부호화된 값을 포함하는 필드를 포함한다.

본 발명의 또 다른 관점에 따르면, 정보 요소를 부호화하는 이진 시퀀스는 정보 요소의 이름이 바뀌는지를 나타내는 대체 플래그를 포함하는 대체 필드를 포함하며, 상기 대체 플래그가 변화를 나타내면 요소이름 참조 필드는 상기 정보 요소의 새로운 이름을 지칭하는 참조를 포함하고, 스키마 참조 필드는 상기 새로운 이름 참조가 정의되는 스키마를 가리키는 참조를 포함한다.

본 발명의 또 다른 관점에 따르면, 상기 부호화된 문서에서 하나의 정보 요소를 부호화하는 상기 이진 시퀀스는 스키마 상태 모드 필드를 포함하며, 상기 스키마 상태 모드 필드는, 상기 정보 요소가 상기 제2 스키마에서 해당 요소에 대한 제1 스키마에서 변하지 않는다는 것을 나타내는 제1 상태와, 상기 정보 요소의 어떠한 부요소들도 상기 제2 스키마에서 상기 해당 요소에 대하여 상기 제1 스키마에서 변하지 않는다는 것을 나타내는 제2 상태와, 상기 제2 스키마에서 상기 해당 요소에 대하여 상기 제1 스키마에서 상기 정보 요소가 변한다는 것을 나타내는 제3 상태를 포함하며, 상기 스키마 상태 모드 필드가 제1 상태에 있을 때에 상기 부호화된 정보 요소는 어떤 스키마 참조와 어떤 다른 변화 정보라도 포함하고, 상기 스키마 상태 모드 필드가 제2 상태에 있을 때에 상기 부호화된 정보 요소의 어떤 부요소도 어떤 스키마 참조와 어떤 다른 변화 정보라도 포함하지 않는다.

본 발명의 또 다른 관점에 따르면, 상기 부호화된 문서에서 적어도 하나의 정보 요소를 부호화하는 상기 이진 시퀀스는 스키마 상태 모드 필드를 포함하고, 상기 스키마 상태 모드 필드는, 상기 제2 스키마에서 해당 요소에 대하여 상기 제1 스키마에서 상기 정보 요소가 변하지 않는다는 것을 나타내는 제1 상태와, 상기 정보 요소의 어떠한 부요소들도 상기 제2 스키마에서 상기 해당 요소에 대하여 상기 제1 스키마에서 변하지 않는다는 것을 나타내는 제2 상태와, 상기 제2 스키마에서 상기 해당 요소에 대하여 상기 제1 스키마에서 상기 정보 요소가 변한다는 것을 나타내는 제3 상태와, 상기 제2 스키마에서 상기 해당 요소에 대하여 상기 제1 스키마에서 상기 정보 요소가 변한다는 것과 상기 정보 요소의 어떠한 부요소들도 상기 제2 스키마에서 상기 해당 요소에 대하여 상기 제1 스키마에서 변하지 않는다는 것을 나타내는 제4 상태를 포함하며, 상기 스키마 상태 모드 필드가 상기 제1 상태일 때에 상기 부호화된 정보 요소는 어떤 스키마 참조와 다른 어떤 변화 정보라도 포함하고, 상기 스키마 상태 모드 필드가 상기 제2 상태이거나 제4 상태일 때에 상기 정보 요소의 어떤 부요소도 스키마 참조와 다른 어떤 변화 정보라도 포함하지 않는다.

본 발명의 또 다른 관점에 따르면, 상기 구조화된 문서를 부호화하는 데에 사용되는 모든 스키마들을 가리키는 참조들을 포함하는 상기 스키마 참조 리스트는 상기 구조화된 문서를 부호화하는 상기 이진 스트림과 연관된 헤더에 삽입된다.

본 발명의 다른 목적은 복호기에 접근할 수 없는 적어도 제1 스키마에서 정의되는 구조를 가지며, 적어도 상기 복호기에 접근할 수 있는 제2 스키마의 변화로부터 기인하는 구조화된 문서를 부호화하는 이진 스트림을 적어도 부분적으로 복호화하기 위한 복호화 방법으로서, 상기 구조화된 문서는 서로 포위되는 정보 요소들을 포함하며, 상기 문서의 상기 정보 요소들은 정보 요소들의 각각의 요소구조를 각각 정의하는 각각의 요소 유형과 적어도 제1 및 제2 스키마에서 관련되며, 상기 제1 스키마는 복호기에 접근할 수 없고 상기 제2 스키마는 상기 복호기에 접근할 수 있으며, 상기 제1 스키마는 상기 제2 스키마에서 정의되는 해당 요소로부터 유도되는 적어도 하나의 유도되는 정보 요소를 정의하는 복호화 방법을 제공하는 데에 있다.

본 발명의 또 다른 관점에 따르면, 상기 복호화 방법은, 상기 문서의 각 정보 요소를 부호화하는 이진 시퀀스를 상기 이진 스트림에서 검출하기 위하여, 상기 제2 스키마를 이용하여 상기 구조화된 문서를 부호화하는 이진 스트림을 순차적으로 읽는 단계와, 상기 복호기에 의해 알려진 스키마 참조에서 정의된 바와 같이 상기 제1 스키마를 가리키는 참조를 부호화된 정보 요소의 각각의 검출된 이진 시퀀스에서 검출하는 단계와, 상기 검출된 이진 시퀀스에서 상기 제1 스키마를 가리키는 상기 참조가 검출되지 않으면, 상기 제2 스키마에서 상기 해당 요소와 상기 제2 스키마에서 상기 해당 요소의 구조 유형을 가리키는 구조 유형 참조를 포함하는 필드와 상기 제1 부분의 부호화된 값을 포함하는 필드와 상기 제1 스키마를 가리키는 참조를 포함하는 필드와 상기 제2 부분의 상기 구조 유형을 가리키는 구조 유형 참조를 포함하는 필드에 따라 상기 검출된 이진 시퀀스를 복호화하는 단계와, 상기 이진 스트림을 순차적으로 읽고 복호화하는 동안, 상기 제1 스키마에 관련된 상기 이진 데이터를 건너뛰는 단계를 포함한다.

본 발명의 또 다른 관점에 따르면, 상기 문서의 각각의 요소를 부호화하는 상기 이진 시퀀스는, 상기 요소의 부호화된 값을 포함하는 내용 필드와, 상기 내용 필드 앞에 위치하고 요소를 부호화하는 상기 이진 시퀀스의 끝을 결정하기 위한 복호기에 의해 사용되는 상기 길이 부호화 값을 포함하는 길이 필드를 포함한다.

본 발명의 또 다른 관점에 따르면, 상기 복호화 방법은, 상기 제1 스키마를 가리키는 참조를 포함하는 상기 이진 시퀀스 내에서 길이 코드 값을 읽고 복호화하는 단계와, 상기 제1 스키마를 가리키는 상기 참조의 상기 이진 시퀀스 내에서 상기 복호화된 길이 값과 상기 위치의 기능으로서 건너뛰기 위한 이진 데이터의 길이를 결정하는 단계를 더 포함한다.

본 발명의 또 다른 관점에 따르면, 상기 유도되는 정보 요소는 상기 제2 스키마에서 상기 해당 정보 요소의 구조유형에 대하여 제한된 구조유형에 제1 스키마에서 연관되며, 상기 유도된 요소들을 부호화하는 상기 이진 시퀀스는 내용필드를 포함하며, 상기 내용필드와 상기 제1 스키마를 가리키는 참조와 상기 유도된 요소의 상기 구조유형을 가리키는 참조는 상기 내용 필드에 부착되고 상기 제2 스키마에서 정의된다.

본 발명의 또 다른 관점에 따르면, 상기 유도되는 정보 요소는 상기 제2 스키마에서 상기 해당 정보 요소의 구조유형에 대하여 확장된 구조유형에 제1 스키마에서 연관되며, 상기 유도된 정보 요소의 구조유형은 제2 스키마에서 정의된 해당 정보 요소의 구조유형을 가지는 제1 부분과 상기 유도된 정보 요소에 대하여 특별하고 상기 제1 스키마에서 정의된 구조유형을 가지는 제2 부분을 포함하며, 상기 이진 시퀀스는 내용 필드를 포함하는 상기 유도된 요소들을 부호화하며, 상기 내용 필드는, 상기 제2 스키마를 가리키는 참조를 포함하는 필드와, 상기 제2 스키마에서 상기 해당 요소의 구조 유형을 가리키는 구조 유형 참조를 포함하는 필드와, 상기 제1 부분의 부호화된 값을 포함하는 필드와, 상기 제1 스키마를 가리키는 참조를 포함하는 필드와, 상기 제2 부분의 상기 구조 유형을 가리키는 구조 유형 참조를 포함하는 필드와, 상기 제2 부분의 부호화된 값을 포함하는 필드를 포함한다.

본 발명의 또 다른 관점에 따르면, 상기 유도된 정보 요소는 상기 제2 스키마에서 상기 해당 정보 요소의 이름에 대하여 변하는 이름을 제1 스키마에서 가지고, 상기 유도된 요소를 부호화하는 상기 이진 시퀀스는 상기 유도된 정보 요소의 이름이 바뀌는지를 나타내는 대체 플래그를 포함하는 대체 필드를 포함하며, 상기 대체 플래그가 변화를 나타내면 스키마 참조 필드는 상기 제1 스키마를 가리키는 참조를 포함하고, 상기 제1 스키마에서 상기 유도 정보 요소의 이름을 지칭하는 요소 이름 참조를 포함한다.

본 발명의 또 다른 관점에 따르면, 상기 부호화된 문서에서 하나의 정보 요소를 부호화하는 상기 이진 시퀀스는 스키마 상태 모드 필드를 포함하며, 상기 스키마 상태 모드 필드는, 상기 정보 요소가 상기 제2 스키마에서 해당 요소에 대한 제1 스키마에서 변하지 않는다는 것을 나타내는 제1 상태와, 상기 정보 요소의 어떠한 부요소들도 상기 제2 스키마에서 상기 해당 요소에 대하여 상기 제1 스키마에서 변하지 않는다는 것을 나타내는 제2 상태와, 상기 제2 스키마에서 상기 해당 요소에 대하여 상기 제1 스키마에서 상기 정보 요소가 변한다는 것을 나타내는 제3 상태를 포함하며, 상기 스키마 상태 모드 필드가 제1 상태에 있을 때에 상기 부호화된 정보 요소는 어떤 스키마 참조와 어떤 다른 변화 정보도 포함하지 않고, 상기 스키마 상태 모드 필드가 제2 상태에 있을 때에 상기 부호화된 정보 요소의 어떤 부요소도 어떤 스키마 참조와 어떤 다른 변화 정보라도 포함하지 않는다.

본 발명의 또 다른 관점에 따르면, 상기 부호화된 문서에서 적어도 하나의 정보 요소를 부호화하는 상기 이진 시퀀스는 스키마 상태 모드 필드를 포함하고, 상기 스키마 상태 모드 필드는, 상기 제2 스키마에서 해당 요소에 대하여 상기 제1 스키마에서 상기 정보 요소가 변하지 않는다는 것을 나타내는 제1 상태와, 상기 정보 요소의 어떠한 부요소들도 상기 제2 스키마에서 상기 해당 요소에 대하여 상기 제1 스키마에서 변하지 않는다는 것을 나타내는 제2 상태와, 상기 제2 스키마에서 상기 해당 요소에 대하여 상기 제1 스키마에서 상기 정보 요소가 변한다는 것을 나타내는 제3 상태와, 상기 제2 스키마에서 상기 해당 요소에 대하여 상기 제1 스키마에서 상기 정보 요소가 변한다는 것과 상기 정보 요소의 어떠한 부요소들도 상기

제2 스키마에서 상기 해당요소에 대하여 상기 제1 스키마에서 변하지 않는다는 것을 나타내는 제4 상태를 포함하며, 상기 스키마 상태 모드 필드가 상기 제1 상태일 때에 상기 부호화된 정보요소는 어떤 스키마 참조와 다른 어떤 변화 정보라도 포함하고, 상기 스키마 상태 모드 필드가 상기 제2 상태이거나 제4 상태일 때에 상기 정보요소의 어떤 부요소도 스키마 참조와 다른 어떤 변화 정보라도 포함하지 않는다.

본 발명의 또 다른 관점에 따르면, 상기 구조화된 문서를 부호화하는 데에 사용되는 모든 스키마들을 가리키는 참조들을 포함하는 상기 스키마 참조 리스트는 상기 구조화된 문서를 부호화하는 상기 이진 스트림과 연관된 헤더 내에서 읽힌다.

실시예

도 1을 참조하면, MPEG-7 표준에 따른 부호화된 구조화된 문서(1)의 이진 포맷은 부호화 모드들을 특정하는 문서 헤더(2) 및 적어도 하나의 구조화된 정보 요소 또는 요소들의 집합(3)을 포함한다.

문서 헤더(2)는 다음의 파라미터를 부호화한다.

- "allows_skipping": 이 파라미터는 요소 길이가 부호화된 문서들 내에서 부호화되는지를 특정한다. 이 파라미터는 다음과 같은 값들을 가질 수도 있다.

00- 길이가 부호화되지 않는다.

01- 길이가 선택적으로 부호화된다.

10- 길이의 부호화는 필수적이다.

- "allows_partial_instantiation": 이 파라미터는 모든 서브트리 또는 문서 내의 구조화된 정보 요소들이 온전히 사례화되는지를 특정한다. 이 파라미터는 다음과 같은 값을 갖는다.

0 - 부분적 사례화도 허락되지 않는다. 문서의 모든 요소들은 부호화된 문서에 나타난다.

1 - 부분적 사례화는 허용된다.

- "allows_subtyping": 이 파라미터는 서브트리가 동질이형 즉, 데이터 유형의 다른 가능한 부휴형을 가질 수도 있는 정보 요소들 또는 요소 속성들을 가지는지를 특정한다. 이 파라미터는 다음과 같은 값을 가진다.

0 - 서브트리 내에서는 동질이형이 부호화되지 않는다.

1 - 동질이형이 허용된다.

도 1에서, 구조화된 정보 요소(3)의 집합은 요소집합 헤더(31)와 요소(32) 및/또는 요소들(3)의 집합을 포함한다. 요소 집합 헤더(31)는 복호화 상태 동안 요소의 유형에 대하여 잠재적 모호성을 명확하게 하는 목적을 가지는 키워, 다음의 더 높은 계층 레벨의 요소(3)에서의 유사한 요소들 또는 요소 집합의 발생수 Nb를 포함한다.

문서의 구조를 정의하는 스키마가 요소 또는 요소 집합의 유형에 대한 단일 필수값을 특정하면, 발생수는 부호화된 문서에는 나타나지 않는다.

요소(32)는 다음의 필드들을 포함한다.

- 요소의 부호화된 값의 비트들 내의 길이를 특정하는 길이 파라미터(33),

- 요소값과 요소의 속성의 구조를 특정하는 데이터 유형수를 포함하는 유형(35),

- 예들 들어 알파벳 순서로 순서 지어진, 요소의 속성 값들(36),

- 요소 자체의 값을 부호화하는 값(37), 요소가 정수나 문자열 등과 같은 단일 유형을 가질 때에 이 필드는 요소(32) 또는 요소 집합(3)의 구조를 가지는 하나 이상의 요소 또는 단일 값을 포함할 수도 있다.

부호화된 문서의 헤더(2) 내의 "allows_skipping" 파라미터의 값에 따라 길이 파라미터(33)는 존재하지 않을 수도 있다. 길이가 요소(32) 내에서 부호화될 때에, 문서 내의 특정 요소에의 빠른 접근은 길이 정보를 이용하여 문서 내에서 이전 요소를 건너뛰므로써 가능하다.

필드들(35, 36, 37)은 부호화된 요소에 나타나지 않을 수도 있다. 요소의 가능한 데이터 유형은 문서의 스키마에서 유일하면, 유형수(35)는 부호화되지 않는다. 스키마에서 특정되지 않은 유형이 속성들을 가지면, 속성값들(36)은 부호화되지 않는다. 요소는 예를 들어 문서의 부분 사례화의 경우에 비어있으면, 요소의 값(37)은 부호화되지 않는다.

도 2를 참조하면, MPEG-7 복호기(10)는 복호기(10)의 입력에서 적용되는 부호화된 문서(7)를 복호화하는 데에 실행되는 이진 구문 코드(13)를 얻기 위하여 XML 스키마와 같은 스키마들(9)을 받아서 처리하도록 설계된 스키마 컴파일레이터(11)를 포함한다. 후자는 예를 들어 XML 포맷에서 복호화된 문서(8)를 출력하는 데에 제공한다.

도 2a에서, 스키마 컴파일레이서 처리(11)는 4개의 상태를 포함한다. 스키마 실현의 제1 상태는 유형 상속을 고르게 할 때와 네임스페이스 지지를 해결할 때이다. 이 상태는 실현된 스키마(15)를 생성한다. 제2 상태는 각각의 복잡한 유형으로부터 구문 트리를 생성할 때이다. 이에 따라 생성된 이 구문 트리(16)는 이진 부호화를 쉽게 하고 압축률을 향상시키기 위하여 변형된다. 제3 상태는 각 트리 노드에 대한 신호와 정규화된 구문 트리들(17)을 생성하기 위하여 이전 상태에 의해 생성된 구문 트리를 정규화할 때이다. 이 신호들은 요소 헤더(31)에 나타나는 헤더 키들과 트리 노드 사이의 관계를 생성하는 다음의 상태에 사용된다. 제4 상태는 복호화 처리에 사용되는 유한 상태 오토마타(12)를 생성하는 때이다.

스키마 실현 상태에서는, 스키마는 확장된 이름을 각 요소와 속성과 유형과 요소들의 그룹에 부여하기 위하여 스키마에 사용되는 모든 이름들을 목록화하도록 해석된다. 이 이름들은 소위 네임스페이스를 이룬다. 확장된 이름들은 네임스페이스 구분자와 아이템 이름을 연결하여 구성된다. 실현된 복잡한 유형들은 각각의 복잡한 유형에 요소 참조를 참조된 요소들의 정의로 대체함으로써 생성된다.

구문 트리 생성 상태에서는, 아이템 노드들은 요소 이름을 요소의 유형에 관련시킴으로써 생성된다. 그룹 노드들을 통한 관계를 포함함으로써 연결된 아이템 노드들은 각각의 복잡한 유형을 위하여 생성된 구문 트리를 구성한다. 특별한 경우에, 구문 트리들은 결과되는 이진 포맷의 조밀성을 향상시키기 위하여 줄어든다. 이러한 처리는 비파괴 방법으로 복잡한 유형 정의를 단순화하는 데에 실행된다.

구문 트리 정규화는 유일한 이름을 구문 트리에서 모든 그룹 노드에 주는 것을 목적으로 한다. 이 처리는 노드들의 순서를 짓고 결국 키를 그것들에 주는 동작이다. 이 키는 오토마타 구성 상태 동안 필요하다.

신호는 모든 요소 또는 요소들의 그룹을 위하여 생성된다. 모든 노드 신호는 그것의 자식 노드 신호들을 연결함으로써 생성된다. 다른 그룹 또는 비순서 노드들의 그룹의 경우에는, 노드 신호들은 알파벳 순서로 정렬되고 추가된다. 순서 지어진 노드들의 그룹의 경우에는, 노드 신호들은 스키마 내에 정의의 순서 내에 추가된다. 아이템 노드 신호들은 그것들의 요소 이름과 같다.

유한 상태 오토마타 생성 상태에서는, 복잡한 유형 오토마톤은 재귀적으로 다음의 규칙에 의해 정의된다.

1. 구문 트리의 모든 노드는 오토마톤을 생성한다.
2. 복잡한 유형 오토마톤은 그것의 뿌리 노드에 의해 생성되는 오토마톤이다.
3. 모든 노드 오토마톤은 그것의 자식 오토마타를 합함으로써 생성된다. 이 합하는 것의 특징은 노드의 특징에 따라 다르다.

처리의 마지막에서는, 그들의 천이 부호화 키를 생성하기 위하여 오토마타는 실현된다. 이 오토마타는 2개의 유형 상태들과, 구문 트리에 각각의 아이템 노드를 위한 후자를 연결하는 요소 천이를 포함한다. 해당 요소가 복호화된 문서의 입력 스트림에 나타날 때에 요소 천이는 교차한다. 단순한 유형의 경우에는 특정 데이터 유형 복호기가 되고 복잡한 유형의 경우에는 기본적인 또는 특정한 유형의 복호기가 될 수도 있는 특정 복호기가 활성화될 때에 유형 상태는 트리거된다.

어떤 경우에는, 문서의 구조를 정의하는 스키마는 확장되고 수정될 것을 요구한다. 복호기에 적용되는 부호화된 문서의 새로운 스키마가 유용하면, 후자는 텍스트 포맷과 이진 포맷 모두에 대하여 온전한 전후방 호환성을 제공할 수 있다. 사실, 전체 표현은 유효화 될 수 있고, 네임스페이스들은 각각의 요소와 속성에 부착될 수 있다. 더욱이, 유형의 계층은 유용하고 확장성 포인트들은 애플리케이션에 대하여 외부적으로 만들어질 수 있다. 스키마 갱신은 확장성과 호환성에 의해 제기된 많은 기술적 문제들을 해결한다. 그러나 표준 MPEG-7은 부호화된 문서를 복호화하는 데에 사용되는 스키마에 대하여 어떠한 정보도 제공하지 않는다. 그러므로 스키마를 사용하도록 구성된 복호기는 수정된 스키마로 부호화된 문서를 복호화할 수 없다(전방향 호환성 불가). 이 문제는 부호화된 문서를 복호화하기 전에 사용될 스키마에 대한 정보를 복호기에 제공함으로써 단순히 해결될 수 있다. 또는, 그러한 오토마타 또는 구문 트리는 입력 이진 스트림을 형성하는 동안 문서를 복호화하는 복호기에 의해 실행되는 이진 구문 코드로 변형될 수도 있다.

특별한 경우에, 스키마 갱신에 의해 기인되는 복잡성은 문제가 있을 수도 있다. 그러한 경우는 수정된 스키마(전방향 호환성)에 접근하도록 하기 위하여 쉽게 복호기를 갱신할 수 없을 때에 예를 들어 갱신될 복호기의 개수가 매우 높거나 복호기가 다른 스키마를 사용하도록 설계되지 않았을 때에 발생한다. 이 특정한 상황을 처리하기 위하여 본 발명의 목적은 정보 요소 이진 포맷과, 복호기가 복호기에 예능새 알려진 스키마에 대하여 수정되거나 확장된 구조를 가지는 문서를 적어도 "부분적으로" 복호화하도록 하는 부호화 및 복호화 방법을 제공함에 있다.

이러한 관점에서, 호환성의 레벨이 부호화 상태에서 선택된다. 이 레벨은 호환성이 보장되어야 하는 스키마들의 집합과 관련하여 표현된다. 부호화 처리는 기대되는 호환성을 보장하기 위해 필요한 리던던시를 추가한다. 극단적인 경우에, 하나의 현재의 스키마와의 호환성이 필요하면, 리던던시는 제거되고 결과적인 이진 부호화된 문서는 현재의 스키마를 따른다.

기본적으로, 부호화 처리는 부호화 상태에서 사용되는 스키마에 대하여 부호화된 문서 정보 내에 부가된다. 이 정보는 스키마 구분자들과 길이 정보를 포함하여 복호기에 유용하지 않은 스키마를 사용하여 부호화하는 문서 내에서 요소를 건너 뛰도록 한다.

스키마 구분자들은 복호기로부터 알려진 스키마 구분자 사전에 의해 정의된다. 예를 들어, 스키마 구분자들은 부호화된 문서의 헤더(2)에 삽입되거나 복호기 초기화와 복호기에 의해 사용되는 구성 파일에 삽입된다. 이 스키마 구분자 사전은 받은 부호화된 문서를 복호화하는 복호기에 의해 사용될 모든 가능한 스키마 구분자들을 포함한다. 이 사전은 다음의 이진 구문을 사용하여 복호기에 의해 읽힐 수 있다.

[표 1]

SchemaDictionary() {	Number of bits
NumberOfSchemaIdentifiers	UINTVLC
i = 0	
while (i < NumberOfSchemaIdentifiers) {	
SchemaIdentifier[i]	String
i++	
}	
}	

여기서, "NumberOfSchemaIdentifiers"과 "SchemaIdentifier[]"는 각각 사전에서 스키마 구분자의 개수와 스키마 구분자를 포함하는 배열을 나타낸다. 그리고 "UINTVLC"는 부정 크기의 비부호 정수를 부호화하는 포맷을 특정한다. 포맷 "UINTVLC"에서는 정수는 정수 청크의 부정 집합에 의해 부호화된다. 각각의 청크는 5비트로 구성된다. 각 청크의 제1 비트는 다른 청크가 현재의 청크를 따르는지를 특정하고, 그 청크의 마지막 제4 비트는 그 정수를 부호화한다.

이 이진 구문은 스키마 사전이 개수를 포함한다는 나타낸다. 개수 뒤에는 문자 스트링 형태의 각각의 스키마 구분자의 목록이 뒤따른다. 목록 내의 각 스키마 구분자의 위치는 특정 스키마를 구분하는 부호화된 문서에 사용될 수도 있는 스키마 개수를 자동으로 정의한다. 사전의 읽기 처리는 순차적으로 스키마들의 개수 "NumberOfSchemaIdentifiers"를 사전에서

읽는 단계와, 카운터 i 를 초기화하는 단계와, 배열 "SchemaIdentifier[]" 내의 i 번째 값을 읽는 소단계와 $0 \leq i < \text{NumberOfSchemaIdentifiers}$ 인 카운터 i 를 증가시키는 소단계를 포함하는 루프 명령들을 포함한다. 다음의 표에서 사전은 3개의 스키마 구분자들을 포함한다. 각각의 스키마는 사전에서 스키마 구분자의 순위에 의해 정의되는 이진 스키마 개수와 내부적으로 연관되어 있다.

[표 2]

Schema Identifier	Implicit Number
S_1	00
S_2	01
S_3	10

이진 스키마 개수의 길이(비트의 개수)는 사전 내의 스키마 구분자들의 개수의 기능으로 정의되어 있고 $E(\log_2(\text{NumberOfSchemaIdentifiers}))$ 과 같다. $E(x)$ 는 다음의 더 높은 정수의 반올림한 x 와 같다.

도 3에 나타난 바와 같이, 부호화된 문서 내의 각 유형 수 필드(35)는 스키마 수 필드(38)와 연관되어 있다. 스키마 개수는 스키마 사전에 정의된 스키마 구분자에 해당한다. 이 스키마 개수는 연관된 유형이 정의된 스키마를 구분한다.

이러한 방법으로, 본 발명에 따른 부호기와 복호기는 문서를 부호화하고 복호화하는 여러 개의 스키마들을 이용할 수 있다.

변화의 3개의 유형은 스키마 S_0 와 수정된 스키마 S_1 사이의 XML 언어로 고려된다. 이러한 변화 유형은 함께 조합된다. 변화의 제1 유형은 전역 요소 이름에서 다른 전역 요소 이름을 뺀 것이다. 전역 요소는 다른 요소(다른 요소에 의해 포위되지 않음)에 대하여 독립된 방법으로 스키마에서 정의된다. 그러한 변화는 XML 스키마 언어에서 다음과 같이 정의된다.

XML 스키마 S_0 에서:

...

<complexType name = "S0:t0" base = "string">

...

<element name = "e0" type = "S0:t0"/>

그리고 XML 스키마 S_1 :

...

<element name = "e1" type = "S0:t0" substitutionGroup = "S0:e0"/>

...

이 구문은 복잡한 유형 "t0"과, 유형 "t0"을 가지는 "e0"과 "e1"로 이름 지어진 두 개의 요소들을 정의한다. 요소 "e1"은 요소 "e0"으로 대체된다. 값 "value1"을 가지는 요소 "e1"은 다음과 같이 스키마 S_1 에 기초하여 XML 문서 내에서 나타날 수도 있다.

...

<e1> value1 </e1>

...

본 발명에 따르면 그러한 수정은 요소(32)에 감산 필드 4를 추가함으로써 부호화된다. 감산 필드 4는 다음을 포함한다.

- 감산 플래그(41)와,
- 감산 요소 개수(43)와,
- 감산 요소 개수(43)이 정의되는 스키마 개수(42).

감산 플래그가 0이 될 때에 다음의 감산 필드들(42, 43)은 부호화된 요소에 나타나지 않는다.

"e1"의 정의를 포함하는 스키마 S1로 부호화된 요소 "e1"은 도 3에 도시된 바와 같이 다음의 필드 값들을 포함한다.

[length]1[S1][e1][S0][t0][value1]

여기서, [length]는 요소의 이진 부호화된 길이이고, "1"는 감산을 나타내는 감산 플래그의 값이고, [S1]과 [S0]은 스키마 사전에 의해 정의된 스키마 S1과 S0의 부호화된 개수이고, [t0]은 유형 "t0"의 부호화된 유형 개수이고, [value1]은 요소 "e1"의 값 "value1"의 부호화된 값이고, [e1]은 요소 "e1"의 부호화된 값이다. 그러한 이진 코드는 스키마 S0(과 따라서 "e0")을 알기만 하는 복호기에 의해 본 발명에 따라 복호될 것이다. 그리고 복호화된 요소는 다음과 같이 해석될 수 있다.

<e0 DDL:IsSubstituted = "true"> value1 </e0>

따라서 복호기는 [e1]의 의미를 이해할 수 없고, "value1"이 다른 모르는 요소에 의해 대체되는 요소 "e0"의 값이라는 것을 고려한다.

스키마 S0와 새로운 스키마 S1 사이의 변화의 제2 유형은 유형 "t0"으로부터 새로운 유형 "t1"을 제한함으로써 유도하는 것이다. 그러한 제한은 유형 정의에서 적어도 하나의 요소 또는 요소들의 그룹의 최대 발생수를 줄이거나, 최소 발생수를 증가시키거나, 유형 t0를 부유형 t1으로 변화시키거나, 유형 t0에서 미리 정의된 제한을 존중하는 요소들의 새로운 그룹을 정의하는 것이다.

그러한 변화는 다음의 예에서 XML 스키마 언어에서 표현된다. XML 스키마 S0은 다음의 텍스트 구문을 포함한다.

<complexType name = "t0">

<sequence>

<element name = "e1" type = "string" maxOccurs = "unbounded"/>

</sequence>

</complexType>

...

<element name = "e" type = "S0:t0"/>

...

XML 스키마 S1에서:

...

<complexType name = "t1">

<complexContent>

<restriction base = "S0:t0">

<sequence>

<element name = "e1" type = "string" maxOccurs = "1"/>

</sequence>

</restriction>

</complexContent>

</complexType>

...

이 예에서, 유형 "t0"은 요소 "e1"의 어떤 발생수도 포함하는 XML 스키마 언어에서 정의된다. 유형 "t1"은 유형 "t0"의 제한으로서 정의되며 유형 "t0"의 특별한 경우인 0 또는 하나의 단일 요소 "e1"을 포함하는 XML 스키마 언어에서 정의된다. 따라서, 유형 "t1"은 유형 "t0" 제한을 강화하고 같은 이진 포맷이 이용될 수 있다.

부호화될 문서는 다음의 정보를 포함한다.

<S0:e xsi:type = "S1:t1">

<e1> value1 </e1>

</S0:e>

이는 문서가 스키마 S0 내에 정의되는 요소 "e"의 기대되는 유형 "t0"의 부유형인 유형 "t1"을 가지는 요소 "e"를 포함한다는 것을 의미한다.

전방향 호환성이 요구되지 않으면, 이 요소는 다음과 같이 부호화될 수 있다.

[length]0[S1][t1][value1]

여기서, "0"은 감산에 설정되지 않은 대체 플래그의 값이다.

스키마 S1이 아닌 S0을 아는 복호기는 필드 [length]와 [S1]을 복호화한다. S1은 스키마 사전에서 정의되지 않으므로, 그것은 요소 "e"에 관련된 다른 필드들이 스키마 S0을 이용하여 해석될 수 없고 길이 정보를 이용하여 요소 "e"를 건너뛰는 것을 알 수 있다. 복호기가 스키마 S1(과 따라서 S0)을 알면, 그것은 요소 "e"의 원래의 구조를 입력받는다.

전방향 호환성이 요구되면, 요소 "e"는 다음과 같이 도 3에 정의된 구조에 따라 부호화된다.

[length]0[S0][t0][value1][S1][t1]

두 스키마들 S0과 S1을 아는 복호기는 부호화된 요소 "e"의 모든 필드를 복호화할 수 있다. 다음으로, 길이 필드와 복호화된 스트림의 길이를 비교함으로써, 더 이상의 부유형이 없고 요소 "e"의 원래의 구조를 입력받을 것이라는 것을 알 수 있다. 복호기가 스키마 S1이 아닌 S0을 알면, 그것은 필드들의 제1 부분을 복호화하고 (길이 필드를 이용하여) [S1]과 [t1] 필드들을 건너뛸 것이다. 그러므로 그것은 다음의 텍스트 트리를 생성한다.

```
<S0:e xsi:type = "S0:t0" DDL:isSubtype = "true">
```

```
<e1> value1 </e1>
```

```
</S0:e>
```

옛 스키마 S0와 새로운 스키마 S1 사이의 변화의 제3 유형은 유형 "t0"으로부터 새로운 유형 "t1"을 확장함으로써 유도하는 것이다. 전방향 호환성이 보장되어야 하면, 새로운 유형 "t1"은 두 부분에서 정의되어야 하고, 제1 부분은 유형 "t0"을 가지며, 제2 부분은 새로운 스키마 S1에서 정의되어야 하는 유형 "t1"을 가진다.

그러한 변화는 다음의 예에서 XML 스키마 언어에서 표현된다. XML 스키마 S0은 다음의 텍스트를 포함한다.

```
<complexType name = "t0">
```

```
<sequence>
```

```
<element name = "e1" type = "string"/>
```

```
</sequence>
```

```
</complexType>
```

```
...
```

```
<element name = "e" type = "S0:t0"/>
```

```
...
```

XML 스키마 S1에서, 새로운 유형 "t1"은 유형 "t0"의 확장으로서 다음과 같이 정의 된다.

```
...
```

```
<complexType name = "t1">
```

```
<complexContent>
```

```
<extension base = "S0:t0">
```

```
<sequence>
```

```
<element name = "e2" type = "string"/>
```

```
</sequence>
```

```
</extension>
```

```
</complexContent>
```

</complexType>

...

전방향 호환성이 요구되면, 부호화될 XML 문서는 다음의 정보를 포함한다.

<S0:e xsi:type = "S1:t1">

<e1> value1 </e1>

<e2> value2 </e2>

</S0:e>

여기서, value1은 유형 "t1" 내의 유형 "t0"의 부분이다. "value2"는 유형 "t1"에 특정한 부분이고 유형 "t0"에 추가된다. 정보는 다음과 같이 부호화된다.

[length]0[S0][t0][value1][S1][t1][value2]

복호기가 스키마 (S1이 아닌) S0을 알면, 그것은 이진 부호화된 정보의 제1 부분을 복호화하고 길이 정보를 이용하여 t1 부분(즉, [t1][value2])을 건너뛸 것이다. 그러한 처리는 다음의 텍스트 트리에 대하여 수행된다.

<S0:e xsi:type = "S0:t0" ddl:isSubtype = "true">

<e1> value1 </e1>

</S0:e>

그러므로 t1의 특정 부분은 복호화되지 않을 것이다. 복호기가 두 스키마들 S0과 S1을 알면, 그것을 원래의 텍스트 트리를 입력받을 것이다.

상기 설명은 조합될 수도 있다. 따라서 다른 스키마들 내에서 정의될 수도 있는 유형의 복잡한 계층들의 정의는 허용된다. 도 4는 유형 t5와 t6이 유형 t4의 두 확장인 계층의 예를 나타낸다. 유형 t4는 유형 t3의 확장이고, 유형 t3은 유형 t1의 제한이다. 비슷한 방법으로, 유형 t8과 t9는 각각 유형 t7의 제한과 확장이다. 유형 t7은 유형 t2의 확장이다. 유형 t2은 유형 t1의 확장이다. 유형들 t1 내지 t9의 각각은 각각의 다른 스키마 S1 내지 S9에서 정의 된다는 것을 고려하여, 호환성은 스키마들 S1, S3, S4 및 S5에 요구되면, 유형 t5의 요소는 다음 형태로 부호화된다.

[length]0[S1][t1][att1][value1][S3][t3][S4][t4][att4][value4][S5][t5][att5][value5]

여기서, 필드들[att-i]은 유형 t-i의 각각의 부호화된 속성 값들이다. [att-i][value-i]은 유형 t5의 t-i 부분이다(-i = 1, 4, 5). 호환성이 스키마들 S3와 S5에 요구되면, 유형 t5의 요소는 다음의 형태로 부호화된다.

[length]0[S3][t3][att3][value3][S5][t5][att5][value5]

여기서, 필드들[att3][value3]은 유형 t3에 따라 부호화된 유형 t5의 t1 부분이고, [att5][value5]는 유형 t5에 따라 부호화된 유형 t5의 t4 + t5 부분이다.

물론, 스키마들 사이의 변화의 상술한 제1 유형은 대체 필드 4를 부호화된 요소에 더함으로써 변화의 제2 및 제3 유형과 조합될 수도 있다.

상술한 바에서 전방향 호환성은 이진 대체 필드들(필드들(41, 42, 43))과 각각의 요소 유형에 관계된 스키마 개수 필드를 화된 문서에 더함으로써 얻어진다. 그러므로 이 솔루션은 결과적인 부호화된 문서의 크기를 증가시킨다. 부호화를 최적화 하기 위하여, 부호화된 문서의 부분들은 같은 스키마에 자주 연결된다는 것이 나타난다. 이러한 관점에서, 본 발명은 도 3

에 나타난 바와 같이 각각의 부호화된 요소(32)에 스키마 상태 모드를 나타내는 스키마 상태 코드를 포함하는 필드(5)를 추가하는 데에 그 목적이 있다. 이 필드는 해당 요소와 부요소가 같은 스키마로부터 명세서를 요구하는지를 정의한다. 같은 스키마가 요소와 후자의 모든 부요소를 부호화(고 복호화하)는 데에 사용되면, (스키마 수들을 포함하는) 전방향 호환성을 보장하기 위해 부가되는 필드(38, 42)는 부호화된 요소로부터 제거된다. 따라서 결과적인 압축률을 증가시킨다.

다음의 표는 스키마 상태 코드의 가능한 값들을 정의한다.

[표 3]

Schema Status Mode	Value
no change	0
freeze the schema	10
Change the schema	11

여기서, "no change"는 요소의 복호화 처리를 위하여 사용되는 스키마는 요소 자체를 위하여 사용되는 스키마와 같다는 것을 의미한다. 그러므로 요소(32)는 스키마 수 필드(38)와 대체 스키마 필드(42)를 포함하지 않는다.

"freeze the schema"는 요소의 복호화 처리를 위하여 사용되는 스키마와 요소의 모든 부요소(요소의 모든 서브트리)는 요소 자체를 위하여 사용되는 스키마와 같다는 것을 의미한다. 그러므로 요소(32)와 요소의 모든 부요소는 어떤 스키마 수 필드들(38, 42)과 스키마 상태 필드(5)라도 포함하지 않는다.

"change the schema"는 요소의 스키마는 변한다는 것을 의미한다. 그러므로 스키마 수 필드들(38, 42)은 요소(32) 내에 나타난다.

스키마 상태 코드에 의해 제어되는 메커니즘은 도 5에 의해 도시된다. 이 도면은 3개의 부요소(52, 53, 54)를 포함하는 주요 요소(51)을 포함하는 구조화된 문서의 트리 구조를 도시한다. 요소(52)는 2개의 부요소(55, 56)를 포함하고, 요소(54)는 3개의 부요소(57, 58, 59)를 포함한다. 요소(55)는 부요소(60, 61)를 포함한다.

요소들(51, 53, 54, 57, 59)은 0("no change")으로 설정된 스키마 상태 코드(5)로 부호화된다. 이 요소는 어떤 스키마 수 필드들(38, 42)도 포함하지 않는다. 요소(52)의 스키마 상태 코드(5)는 10("freeze the schema")로 설정된다. 그러므로 요소(52)의 모든 부요소들(55, 56, 60, 61)은 요소(52)에서 정의된 같은 스키마로 부호화된다. 이 부요소들은 어떤 스키마 수 필드들(38, 42)와 스키마 상태 코드 필드(5)도 포함하지 않는다.

요소(58)는 2개의 부요소(62, 63)를 포함하는 다른 스키마에서 수정되고 정의된다. 그러므로 요소(58)의 스키마 상태 코드(5)는 11("Change the schema")로 설정된다. 요소(62, 63)가 요소(58)와 같은 스키마에서 정의되면, 요소(62, 63)의 스키마 상태 코드(5)는 0으로 설정된다.

더한 최적화는 가능한 하나의 값을 스키마 상태 코드(5)에 다음과 같이 부가함으로써 수행될 수도 있다.

[표 4]

Schema Status Mode	Value
no change	0
freeze the schema	10
Change the schema	110
Change and freeze the schema	111

요소의 스키마 상태 코드가 110 또는 111로 설정될 때에, 요소의 스키마는 변한다. 이 코드가 110으로 설정되면, 후자의 부요소는 스키마 상태 코드(5)를 포함한다. 이 코드가 111로 설정될 때에, 후자의 스키마를 이용해서 요소의 모든 부요소들은 부호화된다.

도 5의 예에서, 요소들(62, 63)이 요소(58) 같은 스키마에서 정의된다는 것을 특정하기 위하여 요소(58)의 스키마 상태 코드는 111로 설정될 수도 있다. 그러므로 스키마 상태 코드 필드(5)는 요소들(62, 63)에서 요구되지 않는다.

산업상 이용 가능성

본 발명은 스키마에 필요한 모든 것이 복호기에 의해 알려지지 않더라도 문서가 부분적으로 복호화될 수 있는 방법으로 구조화된 문서를 부호화하는 방법을 제공하는데 특징이 있다.

또한 본 발명은 전방 및 후방 호환성을 보장하는 즉, 복호기에 접근할 수 없는 적어도 제1 스키마에서 정의되는 구조를 가지며, 적어도 상기 복호기에 접근할 수 있는 제2 스키마의 변화로부터 기인하는 구조화된 문서를 상기 복호기가 적어도 부분적으로 복호화할 수 있도록 하는 부호화 방법을 제공한다.

도면의 간단한 설명

본 발명의 다양한 형태와 바람직한 실시예들은 다음의 설명과 첨부된 도면을 참조함으로써 더 잘 이해 될 수도 있다.

도 1은 MPEG-7 표준에 따른 구조화된 문서의 트리구조의 이진 포맷.

도 2는 MPEG-7 복호기의 블록도.

도 2a는 도 2에 나타난 복호기의 상세부분의 블록도.

도 3은 본 발명에 따른 부호화된 정보 요소의 이진 포맷을 나타낸 도면.

도 4는 트리 구조에서 서로에 대하여 정의된 요소 유형을 나타낸 도면.

도 5는 트리 구조에서 조직된 정보 요소들을 포함하는 구조화된 문서를 나타낸 도면.

- 도면의 설명 번역 -

<도 1>

1: 문서

2: 헤더

3: 요소 집합

31: 요소 헤더

32: 요소 바디

33: 길이

35: 유형

36: 속성

37: 값

Integer: 정수

Key: 키

<도 2>

7: 부호화된 문서

8: 복호화된 문서

9: 스키마

10: 복호기

11: 스키마 컴파일레이션

12: 유한 상태 오토마타

13: 이진 구문

<도 2a>

9: 스키마

11: 스키마 컴파일레이션

12: 유한 상태 오토마타

15: 실현된 스키마

16: 구문 트리

17: 정규화된 구문 트리

<도 3>

4: 대체

5: 모드

32: 요소 바디

33: 길이

34: 내용

35: 유형

36: 속성

37: 값

38: Sch

41: 대체 폴래그

42: Sch 대체물

43: 요소 번호

<도 4>

extention: 확장

extention: 확장

extention: 확장

restriction: 제한

restriction: 제한

extention: 확장

extention: 확장

extention: 확장

<도 5>

freeze schema: 프리즈 스키마

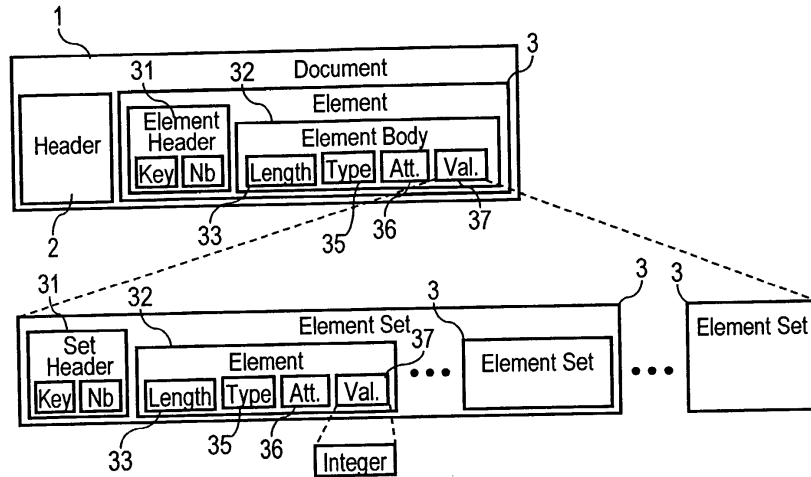
frozen schema: 프로즌 스키마

no schema change: 스키마 무변화

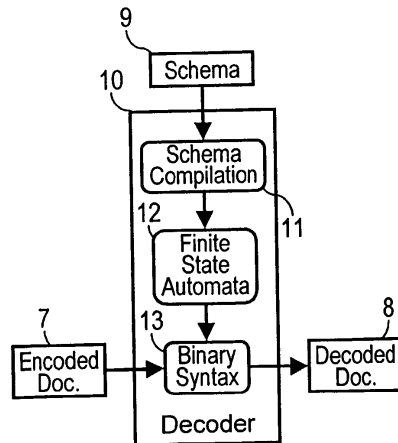
schema change: 스키마 변화

도면

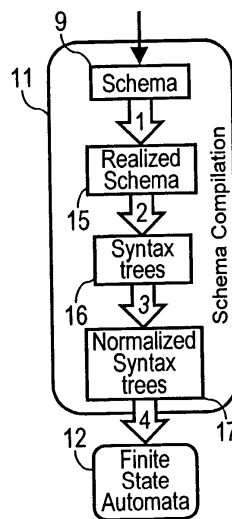
도면1



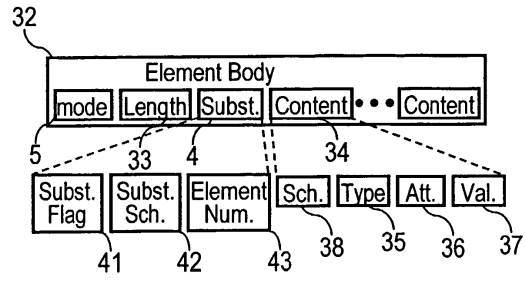
도면2



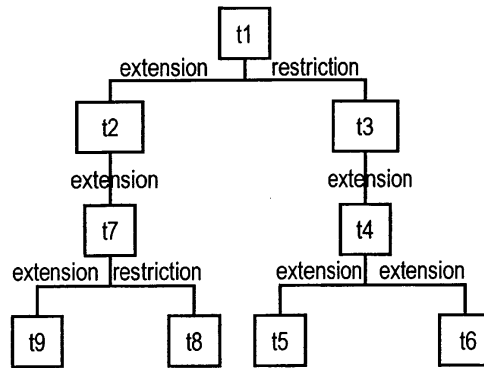
도면2a



도면3



도면4



도면5

