

RLE 압축 기법을 이용한 저장 시스템

김경옥* · 김종찬* · 반경진* · 허수연* · 김응곤*

*순천대학교

Storage systems using RLE compression

Kyeong-Og Kim* · Jong-Chan Kim · Kyeong-Jin Ban* · Su-yeon Heo* · Eung-Kon Kim*

*SunChon National University

E-mail : aib27@hanmail.net

요 약

최근에는 주변 환경에 전자태그 및 센서를 부착하고 부착된 센서를 통한 상황 정보가 활용 되면서 많은 양의 데이터를 효과적으로 저장하고 검색하는 기법이 필요하다. 본 논문에서는 USN/RFID 단말 장치 중에서 연속적인 데이터 수집이 이루어지는 센서에 대해서 RLE(Run Length Encoding) 압축 방식을 응용한다. 데이터에는 시간 정보를 마킹하여 하나의 데이터 블록을 생성한 다음에 저장한다. 연속적인 데이터에 시간 정보를 기록함으로써 원하는 시간 및 특정 장소의 데이터를 빠른 시간 안에 찾을 수 있도록 저장 기법을 제안한다.

ABSTRACT

The supply of context information is increasing with the propagation of ubiquitous computing environment. Recently, as context information is being collected through electronic tags and sensors attached to the environment, we need methods to efficiently store and search large volumes of data. This paper describes the application of the RLE (Run Length Encoding) compression method for sensors that continuously collect data in USN/RFID terminals. Time information is marked on the data and one data block is generated and saved.

This paper proposes a storage method that allows us to quickly search data of the desired time and place by recording time information in continuous data.

키워드

컨텍스트, USN/RFID, RLE, 레코드화, 압축

I. 서 론

컨텍스트란 객체 상황에 대한 특징을 기술한 모든 정보를 의미한다[1]. 상황 정보들은 주변 환경의 상황에[2] USN/RFID 태그를 부착하고 부착된 센서를 통해 온도, 습도, 조도, 중력, 자기 등의 데이터의 정보를 수집한다. 이 정보를 통해 여러 가지 상황을 판단하여, 상황에 맞게 많은 데이터를 저장해야 한다[3][4]. 최근 연속적인 데이터의 활용이 점점 확대되면서 데이터의 양도 점점 더 많아지고 데이터 효율적인 저장, 검색이 필요하다.

대용량 자료의 데이터를 단순히 발생된 시간, 장치정보, 센싱 값으로 구성된 레코드 형식으로 저장을 하면 데이터의 중복 뿐만 아니라 방대한 양의 데이터를 저장하기에는 저장 공간의 효과적

인 이용 측면에서 많은 단점을 갖게 된다. 본 연구에서는 이러한 단점을 보완하기 위해 USN/RFID 단말 장치 중에서 연속적인 데이터 수집이 이루어지는 센서에 대해서 RLE 압축 방식을 응용하여 데이터를 시간 정보를 마킹하여 하나의 데이터 블록을 생성한 다음에 저장한다. 저장된 데이터에 시간 정보를 기록함으로써 원하는 시간 및 특정 장소의 센서 데이터를 빠른 시간 안에 찾을 수 있도록 효과적으로 검색하기 위한 저장 기법을 제안한다.

II. 관련연구

1. Run-Length Coding

런 길이 코딩은 자료 중복성을 이용하는 가장

간단한 자료 압축 기법 중의 하나이다. 어떤 영상은 일정한 색상으로 구성되는 영역을 많이 가진다. 일정한 색상은 동일한 색상의 문자가 반복되어서 표현된다. 이는 반복되는 문자들로 정의한다. 이는 원시 데이터의 개수와 데이터로 표현된다. 예를 들면, 샘플 데이터의 문자열이 표 1과 같을 때

예 1) AAAABBBBBCCCCCCCCDEEEEE(22 bytes) 일 경우에

표 2와 같이 표현된다.

4A5B8C1D4E(10 bytes)

A의 개수가 4이므로 4A로 표현되며, B의 개수가 5이므로 5B로 표현된다. 이와 같은 방법이 계속해서 적용되며, 10 바이트로 22 바이트의 자료를 표현할 수 있게 된다. 이 경우에 압축율은 다음과 같다.

$$Compression\ Ratio = \frac{Source\ Data}{Compressed\ Data}$$

22 bytes/10bytes=2.2 선택된 ASCII 문자열에 적용된 이 방법은 명료한 결과를 만든다. 그러나 위와 같은 순서로 인쇄된 문자열은 일반적인 문서에서는 보기 힘들 것이다. 다음과 같이 영문자로 구성된 실질적인 한 문자열을 선택한다면

예제 2) Iamagooboy

다음과 같이 부화화될 것이다.

111a1m1a1g2o1d1b1o1y

여기서 나오는 압축율은 0.55로서 11바이트를 20 바이트로 표현한다. 실제적으로 원시 자료를 두 배 확대한 것이 된다[5]. 이 경우에는 보다 나은 방법이 필요하다.

III. RLE 압축 기법과 블록의 레코드화

센싱 데이터를 전송하는 USN/RFID 단말장치는 1초당 총 4건의 원시 데이터를 발생한다. 60초라고 하면 240건이 되고, 간단히 10개의 USN/RFID 단말장치를 고려해도 1분에 대략 2,400건의 원시 로그 데이터가 발생하게 된다. 이러한 데이터를 단순히 발생된 시간, 장치정보, 센

싱 값으로 구성된 레코드 형식으로 저장을 하면 데이터의 중복 뿐 아니라 저장 공간의 효과적인 이용 측면에서 많은 단점을 갖게된다.

압축된 블록의 레코드화는 요청된 데이터를 중심으로 일정 범위 내에 전후 데이터를 모두 포함한다. 이는 일반적으로 특정 데이터를 요청한 경우, 이후에 요청되는 데이터가 초기에 요청된 데이터를 중심으로 일정 범위 내의 데이터일 가능성이 매우 높기 때문에 이러한 전략으로

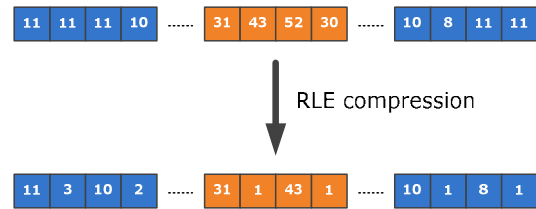


그림 2 원시 로그 데이터의 RLE 압축

개수를 운영하게 된다. 또한 컨텍스트가 인식된 경우 인식에 이용된 원시 데이터에 대한 접근 정보도 같이 저장되는데 이때 로그 ID, TS index가 저장에 이용된다. 그림 2는 원시 로그 데이터의 RLE 블록 압축 기법을 표현한 것이다. 그림 3은 데이터베이스 관련된 개체들을 통합적으로 표현한 스키마이다.

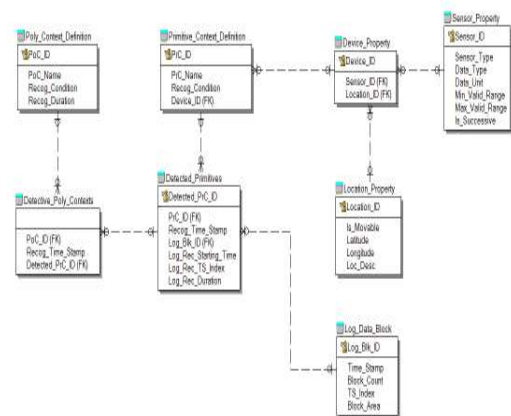


그림 3 통합 데이터베이스 스키마

IV. 컨텍스트 데이터 압축 저장 시스템

1. 컨텍스트 압축 저장

센서 노드에 연결된 각 센서에 특정 상황 즉, 온도 센서의 경우 드라이어로 온도 상승, 습도 센서의 경우 가슴기 가동 등의 조작을 통해 인위적

인 상황을 연출하였다.

센서들에 대해 예상되는 지정된 상황을 무작위로 발생시키고, 1분 동안 반복적으로 실험을 지속한다. 실험 종료 후 데이터베이스에 저장된 컨텍스트의 레코드의 개수를 세서 분당 처리량을 계산하였다. 실험 결과로는 1분 동안 반복적으로 발생시킨 상황 데이터는 상황 인지 프레임워크를 통해 다음 그림과 같이 데이터베이스에 총 54개의 레코드화 되었다. 컨텍스트 압축 데이터 시스템은 데이터 입력, 버퍼링 RLE압축과 시간 정보 마킹을 통해 데이터 블록을 생성한 다음 데이터를 생성하였다. 그림 5는 데이터 베이스에 저장된 컨텍스트 레코드의 분당 처리량을 나타낸 것이다.

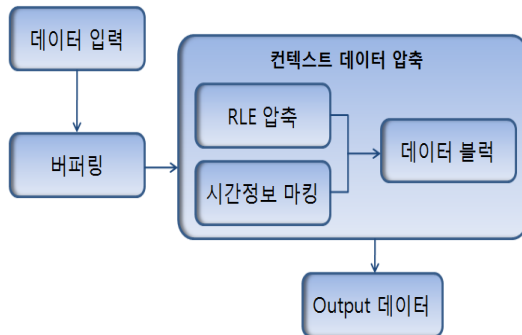


그림 4 컨텍스트 압축 데이터 구성도

V. 결 론

USN/RFID 단말장치들로부터 발생되는 대량의 원시 로그 데이터를 RLE 압축 방식을 응용하여 효과적으로 저장하고 이를 빠르고 편리하게 검색할 수 있는 방법으로 RLE 압축 기법을 제안하였다. 본 논문에서는 여러 장소에서 발생하는 다양한 센서들의 실시간 측정값들을 일정 시간 단위로 블록화한 후 압축 저장하였다. 실시간 데이터의 방대한 양의 데이터를 압축하여 저장함으로써 데이터의 중복 뿐만 아니라 저장 공간을 절약하여 단점을 보완하였다.

감사의 글

본 연구는 교육과학기술부와 한국산업기술재단의 지역혁신 인력양성사업으로 수행된 연구 결과임.

참고문헌

- [1] Daniel Salber, Anind K.Dey and Gregory D. Abowd. The Context Toolkit: Aiding the Development of Context-Enabled Applications. In the Proceedings of the 1999 Conference on Human Factors in Computing Systems (CHI '99), Pittsburgh, PA, May 15-20, pp.434-441,1999.
- [2] 손중수, 정인정, "Fuzzy OWL을 이용한 사용자 Context의 표현 및 추론", 지능정보연구 제14권, 제1호, 3월, pp35~45, 2008.
- [3] 송준영, 이기혁, "센서 노드를 위한 플래시 메모리 저장 시스템에 대한 고찰", 한국정보과학회 가을 학술발표논문집, Vol.34, No.2(C), 2007.
- [4] P. J. Brown, G. J. F. Jones, "Context-aware Retrieval: Exploring a New Environment for Information Retrieval and Information Filtering,"Personal and Ubiquitous Computing, 2001, Volume 5, Issue 4, December, pp. 253-263, 2001.
- [5] 이상진, "Run-Length Coding과 Huffman Coding을 이용한 Entropy Codec의 설계및 구현", 석사 졸업논문, 2004.