

# 시공간 EPC 데이터 처리를 위한 선택률 기반 효율적인 연속질의 처리 기법

Efficient continuous query processing technique based on  
selectivity for EPC data with time and location<sup>+</sup>

추병조\* · 홍봉희 · 김기홍

Byung-jo Chu\*, Bonghee Hong, Gi-hong Kim

부산대학교 컴퓨터공학과

{cbj1004\*, bhhong, buglist}@pusan.ac.kr

## 요 약

EPCglobal은 기업 간의 물류 활동 촉진과 글로벌 유통물류 시스템 구축을 위하여 EPCglobal Architecture Framework을 제시 하였다. EPCglobal Architecture Framework의 한 구성요소인 EPCIS(Electronic Product Code Information Services)는 EPC, 시간, 위치와 같은 물류 관련 정보에 대해 저장 및 검색 서비스를 제공한다. EPCIS는 단발성 질의(poll)와 연속 질의(subscribe) 검색 서비스를 제공한다. EPCIS의 연속 질의는 시스템 자동화 및 재고 관리, 공급망 관리를 위해 다양한 응용에서 활용이 가능하다. 일반적으로 연속 질의 처리를 위해서는 등록된 연속 질의와 입력된 데이터를 순차적으로 비교하는 Sequential Matching 기법을 사용한다. Sequential Matching기법은 등록된 연속 질의 수가 증가 할 경우 많은 부하를 발생 시키고, 이로 인해 시스템 처리 지연이 발생한다.

본 논문에서는 EPCIS의 시공간 EPC 데이터의 연속질의 처리 성능 향상을 위해 선택률 기반 효율적인 연속질의 처리 기법을 제안한다. 13차원의 도메인을 여러 개의 질의 색인으로 구성하고, 등록된 질의 정보를 기반으로 선택률을 계산한다. 선택률에 의해 변경되는 동적 질의 실행 계획을 제안함으로써, EPCIS에서 시공간 EPC 데이터의 연속질의 처리에 대해 평균 60%의 성능이 향상이 가능하도록 하였다.

## 1. 서론

RFID(Radio Frequency Identification)는 각종 물품에 소형 칩을 부착해 사물의 정보와 주변 환경정보를 무선주파수로 전송·처리하는 비접촉식 자동 인식시스템이다. 이러한 RFID는 빠른 인식 속도, 인식률, 데이터 저장능력으로 인해 여러 응용 분야에서 주목 받고 있는 기술이다. RFID의 응용 분야는 재고관리, 물류의 공급망 관리, 공장 자동화 등으로 다양하게 확산되고 있다[1].

EPC(Electronic Product Code)는 기존의 바코드와 마찬가지로 상품을 식별하는 코드이다. 바코드는 상품 단위 식별이 가능했지만, EPC는 상품의 단품 레벨까지 식별

이 가능하다. EPCglobal은 RFID 관련 기술을 보급 및 표준화를 담당하는 비영리 국제기구이다. EPCglobal은 개체에 EPC 고유 식별자를 넣은 RFID 태그를 붙여 관리하고 인터넷을 통해 이에 대한 정보획득 절차를 표준화하여 글로벌한 유통물류 시스템 구축을 목표로 하고 있다. 이를 위해 EPCglobal은 EPCglobal Architecture Framework를 제안하였다. EPCglobal Architecture Framework의 한 구성요소인 EPCIS(EPC Information Services)는 EPC 관련 정보를 저장 및 검색 서비스를 제공한다. EPCIS는 원하는 검색 조건을 기술하고, 그 결과를 즉시 받는 단발성 질의(poll)와 검색 결과를 주기적으로 보고 받

+ “이 논문은 2008년 정부(교육인적자원부)의 재원으로 한국학술진흥재단의 지원을 받아 수행된 연구임”(지방연구중심대학육성사업/차세대IT기술연구사업단)

는 연속 질의(subscribe) 검색 서비스를 제공한다. 일반적으로 연속 질의 처리를 위해서는 등록된 연속 질의와 입력된 데이터를 순차적으로 비교하는 Sequential Matching 기법을 사용한다. 그러나 Sequential Matching 기법은 등록된 연속 질의 수가 증가 할 경우 많은 부하를 발생시키고, 이로 인해 시스템 처리 지연이 발생한다.

본 논문에서는 EPCIS의 시공간 EPC 데이터의 연속질의 처리 성능 향상을 위해 선택률 기반 효율적인 연속질의 처리 기법을 제안하였다. 이를 통해 EPCIS에서 시공간 EPC 데이터의 연속질의 처리에 대해 평균 60%의 성능이 향상이 가능하도록 하였다.

본 논문의 구성은 다음과 같다. 2장은 관련 연구로써 EPCglobal Architecture Framework와 연속 질의를 위한 질의 색인을 소개 한다. 3장에서는 이 논문의 대상 환경과 문제를 정의 한다. 4장에서는 문제를 해결하기 위한 방법으로써, 여러 개의 색인 구성 방법과 동적 질의 실행 계획을 제안한다. 5장에서는 실험결과에 대해 살펴보고 마지막으로 6장에서는 결론 및 향후 연구 방향에 대해 기술한다.

## 2. 관련 연구

### 2.1 EPCglobal Network

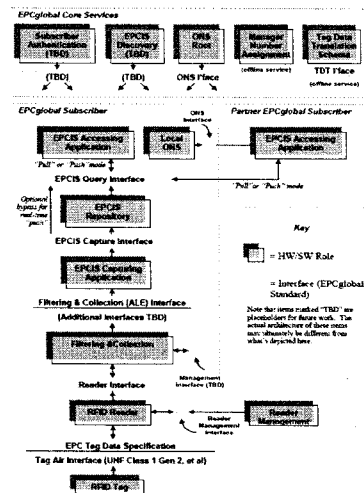


그림 2 EPCglobal Network Architecture

EPCglobal은 개체에 EPC 고유 식별자를 넣은 RFID 태그를 붙여 관리하고, 이를 사용하여 공급, 유통망을 강화를 통한 물류 활동 촉진시키기 위해 그림 1과 같은 EPCglobal Architecture Framework를 제안하였다.[2]

EPCglobal Network Architecture의 주요 구성 요소의 역할은 다음과 같다.

- RFID Tag: 상품을 유일하게 구분할 수 있는 EPC를 보유
  - RFID Reader: 하나 이상의 안테나 환경 내에 있는 RFID 태그의 EPC를 인식
  - Filtering & Collection (ALE): 하나 이상의 RFID Reader로부터 수신된 데이터를 중복 제거 및 그룹화
  - EPCIS Capturing Application: EPC 관련 비즈니스 이벤트 생성 및 EPCIS로 이벤트 전송
  - EPCIS: EPC 관련 정보를 저장하고, 검색 서비스를 제공
  - ONS: 특정 EPC에 관련 자세한 정보를 제공해주는 EPCIS 주소를 검색
- 위의 주요구성 요소 중에서 우리는 정보 저장소의 역할을 하는 EPCIS의 검색 서비스에 대해서 연구해보고자 한다.

### 2.2 EPCIS(EPC Information Services)

EPCIS의 특징은 ALE, Reader와 같은 하위 시스템과 달리 과거 데이터를 저장하고, 비즈니스 정보를 다룬다. EPCIS 데이터는 크게 4개의 기본 이벤트로 구성이 된다. 이를 통해 비즈니스 상에 일어나는 사건을 표현한다.

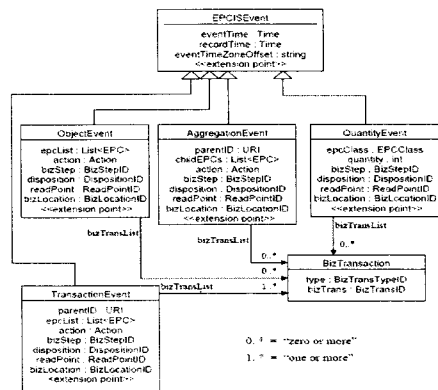


그림 3 EPCIS 기본 이벤트

그림 2는 EPCIS의 기본 이벤트 타입을 보여준다.[3]

4가지 이벤트는 다음과 같은 의미를 가진다.

- Object Event: 어떤 시간과 장소에서 어떤 비즈니스 단계에서 관측된 EPC의 리스트를 표현한다. Object Event는 단순 관측 정보를 가진다.
- Aggregation Event: 어떤 시간과 장소에서 어떤 비즈니스 단계에서 부모 단계의 EPC에 물리적으로 결합 또는 해체되는 EPC의 리스트를 가진다.
- Quantity Event: 어떤 시간과 장소에서 관측된 상품단위의 EPC의 수량을 가진다.
- Transaction Event: 비즈니스와 연결된 정보를 저장하는데 사용하는 이벤트이다.

EPCIS는 위와 같은 4가지 이벤트를 다양한 응용에서 검색이 하기 위한 그림 3과 같은 EPCIS Query Control Interface를 제공한다.

```

<<interface>>
EPCISQueryControlInterface
---
subscribe(queryName : String, params : QueryParams, dest :
URL, controls : SubscriptionControls, subscriptionID :
String)
unsubscribe(subscriptionID : String)
poll(queryName : String, params : QueryParams) :
QueryResults
getQueryNames() : List // of names
getSubscriptionIDs(queryName : String) : List // of Strings
getStandardVersion() : string
getVendorVersion() : string
<<extension point>>
    
```

그림 4 EPCIS Query Control Interface

이중 가장 중요한 함수는 poll과 subscribe이다. poll은 원하는 검색 조건을 QueryParams로 기술하고, 그 결과를 즉시 받는 단발성 질의이며, subscribe는 원하는 검색 조건을 기술하고, 검색 결과를 주기적으로 보고 받는 연속 질의이다. 본 논문에서는 EPCIS에서 연속 질의인 subscribe를 효율적으로 처리 하는 기법에 대해서 연구 하고자 한다.

### 2.3 질의 색인

연속 질의 성능을 개선하기 위해 사용하는 방법 중 대표적인 것이 질의 색인이다.

질의 색인의 요구사항은 실시간 연속 질의 수행을 목적으로 하기 때문에 빠른 검색 속도를 위해 기본적으로 메인 메모리 기반의 색인 구조를 가지며, 메인 메모리 환경에서 메모리 사용 공간을 최소화하고 삽입되는 질의 개수에 독립적인 빠른 검색 성능을 보장하는 것이다.

대표적인 다차원 질의 색인 구조로써, VCR(Virtual Construct Rectangle) Index[4]와 CQI(Cell-based Query) Index[5]가 있다.

## 3. 대상 환경 및 문제 정의

### 3.1 대상 환경

EPCIS는 EPC 관련 정보를 저장하는 정보 저장소 역할을 한다. 그리고 저장된 데이터를 단발성 질의(poll)와 연속 질의(subscribe) 형태로 검색 서비스를 제공한다. EPCIS의 연속 질의는 시스템 자동화 및 재고 관리, 공급망 관리를 위해 다양한 응용에서 활용이 가능하다. 그림 4는 EPCIS 연속 질의를 이용한 재고 관리 시스템 응용을 나타낸다.

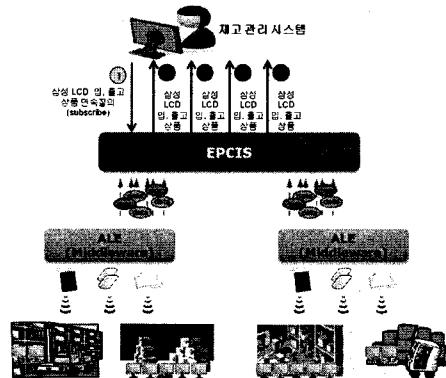


그림 5 EPCIS 연속 질의를 이용한 재고 관리 시스템

재고 관리 시스템 관리자가 작업 진행 상황을 모니터링 하기 위해 매 10분마다 삼성 LCD 입, 출고 상품에 대한 정보가 필요하다 하거나, EPCIS의 연속 질의를 이용하여 이 요구사항을 처리 할 수 있다. 그림 4와 같이 첫째, EPCIS에 삼성 LCD 입, 출고 상품에 대해 10분의 보고 주기를 가

지는 연속 질의를 등록 한다. 그 후 관리자는 매 10분마다 연속 질의의 결과인 삼성 LCD 입, 출고 상품 정보 획득이 가능하다. 그림 4의 예제와 같이 EPCIS 연속 질의는 대량의 데이터에서 사용자가 원하는 데이터만을 효율적으로 검색하여, 제공해주는 역할을 한다. 또한 시스템의 편리함을 제공하고 자동화를 보조해주는 역할을 한다.

### 3.2 일반적인 질의 색인 처리 기법

일반적으로 연속 질의 처리를 위해서는 등록된 연속 질의와 입력된 데이터를 순차적으로 비교하는 Sequential Matching 기법을 사용한다.

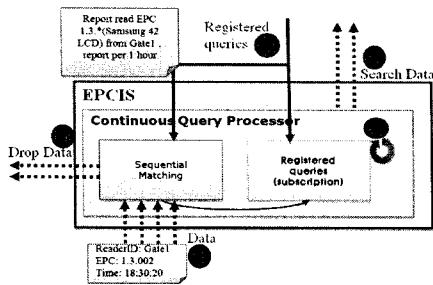


그림 6 Sequential Matching을 이용한 연속질의 처리

동작 방법은 그림 5와 같다. 첫째 질의가 등록이 되고, EPCIS 데이터가 입력이 되면 등록된 연속 질의와 하나씩 비교하여, 매치되는 연속 질의가 있을 경우 해당 연속 질의의 결과로 추가한다. 그리고 연속 질의 처리기(Continuous Query Processor)에서는 해당 질의 결과 반환 시점이 되면 검색 결과를 사용자에서 전송하게 된다. 질의 결과는 메인 메모리에 유지된다.

### 3.3 문제정의

Sequential Matching기법은 등록된 연속 질의 모두와 입력된 데이터를 하나씩 비교하기 때문에, 등록된 연속 질의 수가 증가할 경우 많은 부하를 발생 시키고, 이로 인해 시스템 처리 지연이 발생한다.

일반적으로 많은 수의 연속 질의의 성능을 향상시키기 위해서는 질의 색인을 사용

한다. 그러나 EPCIS는 13차원의 도메인과 모든 데이터 필드가 필수 조건이 아니라는 것으로 인해, 하나의 질의 색인을 사용하여 연속질의를 처리 할 경우, 차원의 저주 및 무한 영역 질의 문제를 발생 시킨다

그림 6은 색인의 차원 증가함에 따라 현격히 증가하는 데이터 중복 비율을 보여준다.[6] 높은 데이터 중복 비율은 검색 시 많은 노드를 탐색하여, 검색 성능을 떨어저 하 시키는 차원에 저주를 발생 시킨다.

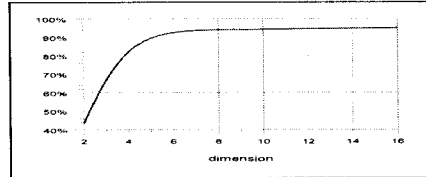


그림 7 차원의 증가에 따른 중복 비율

## 4. 문제해결 방법

### 4.1 Basic Idea

본 논문에서는 13차원의 차원의 저주 및 무한 영역 질의를 해결하기위해 여러 개의 색인으로 EPCIS 질의 색인을 구성한다. 실제적인 동작과정은 그림7과 같다.

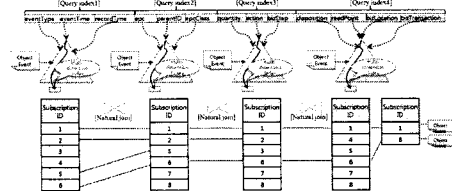


그림 8 여러 개의 색인으로 구성된 질의 색인 동작과정

여러 개의 색인을 구성하고, 각각의 색인을 처리한 결과를 조인하여, 모든 색인 결과에 해당하는 연속 질의가 있을 경우에는 연속 질의의 검색 결과로 데이터(이벤트)를 추가 시키는 것이다.

### 4.2 동적 질의 실행 계획

여러 개의 색인으로 질의 색인을 구성할 경우 그림 8과 같이 색인을 처리하는 순서에 따라 성능의 차이가 발생한다.

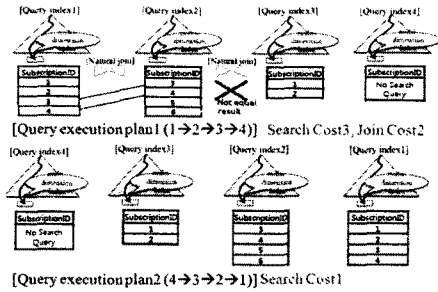


그림 9 질의 실행 계획에 따른 성능차이

이는 각각의 색인을 결과를 조인 후 결과가 있을 경우에 조인의 결과만을 가지고 다음 색인을 처리하기 때문이다. 그림 8의 질의 실행 계획2과 같이 검색 결과를 적게 만드는 색인을 먼저 처리 하는 것이 좋다. 즉 연속질의 선택률이 낮은 색인부터 처리 하는 것이 효율적이라는 것이다.

본 논문에서는 도메인 크기, 도메인 필 수 유무, 등록된 유저의 질의를 고려하여, 각각의 질의 색인 선택률을 계산하고, 이를 기반으로 입력된 이벤트에 대해서 동적 질의 실행 계획을 제안한다. 유저의 질의 조건은 가변적이고, 실시간으로 변경이 되므로, 연속 질의 등록 및 삭제 시 동적으로 정보를 갱신하여, 최적화된 질의 실행 계획이 가능하도록 한다.

#### 4.3 연속 질의 처리기 구조

그림 9는 동적 질의 실행 계획을 적용하기 위한 연속 질의 처리기 구조도를 나타내고 있다.

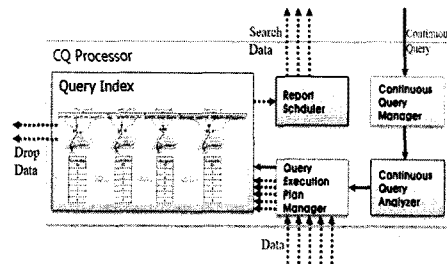


그림 10 연속 질의 처리기 구조도

제안한 연속 질의 처리기는 5개의 모듈로 구성되어 있다. 각 모듈의 역할 및 특징 다음과 같다.

- Query Index: 여러 개의 색인으로 구성

되며, 입력된 데이터를 각각의 색인을 검색하여 조건에 맞는 질의가 있을 경우에 연속 질의 검색 결과로 추가 시키고, 없을 경우에는 삭제된다.

- Report Scheduler: 등록된 모든 연속 질의 보고 주기를 관리한다.
- Continuous Query Manager: 등록된 모든 연속 질의를 관리하고, 연속 질의 결과로 검색된 모든 데이터를 결과 전송 시점까지 관리한다.
- Query Execution Plan Manager: 등록된 연속질의를 분석하여, 각각의 색인의 선택률을 계산하고, 이를 기반으로 입력된 데이터에 대해 최적화된 질의 수행 계획을 생성한다.
- Continuous Query Analyzer: 등록된 연속 질의의 검색 조건의 도메인 크기, 질의 크기 등을 분석하여, Query Execution Plan Manager에게 연속 질의 분석 정보를 제공한다. 연속 질의 등록, 삭제 시점에 질의를 분석한다.

## 5. 성능평가

### 5.1 실험 환경

본 논문에서 제안한 선택률 기반 연속질의 처리 기법은 Java 2 standard edition development kit 1.6을 사용하여 구현하였다. 실험 플랫폼으로는 Windows XP에서 2GB 메인 메모리, CPU Pentium IV 2.4GHz를 장착한 PC를 이용하여 실험하였다.

실험 데이터로는 실제 발생 할 수 있는 모든 데이터 조건에 대해 랜덤으로 생성한 균등분포 데이터를 사용하였다.

### 5.2 성능 비교

성능 비교를 하기 위해 기존의 연속 질의 처리를 위해 사용되는 Sequential Matching과 비교하였다. 연속 질의 처리는 메인 메모리 기반에서 질의 및 데이터를 처리함으로써 CPU 점유율을 비교 통하여, 성능 평가를 하였다.

Sequential Matching의 경우 그림 10의 그래프와 같이 등록된 연속 질의 수에 비례하여, CPU 점유율이 급격히 증가하는 것을 알 수 있다. 특히 초당 이력 이벤트가 5개 일 경우에는 등록된 연속 질의가 500개

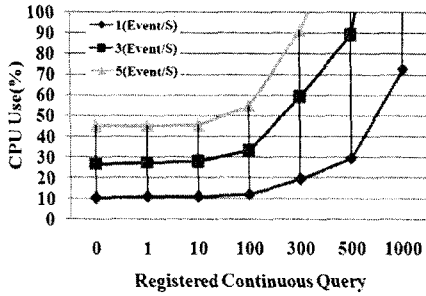


그림 11 Sequential Matching의 연속 질의 처리 성능

이상 일 경우에는 CPU 자원을 100%를 다 사용하여도, 연속 질의를 모두 처리하지 못해, 시스템 처리 지연이 발생 하였다.

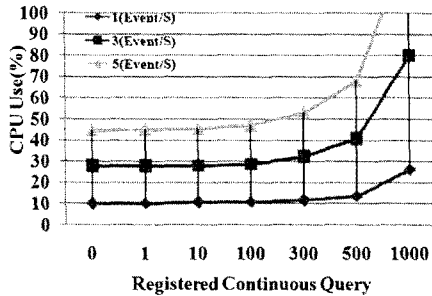


그림 12 선택률 기반 연속 질의 처리 기법 성능

선택률 기반 연속질의 처리 기법의 경우 그림 11의 그래프와 같이 등록된 질의 수에 증가에 따라 완만한 곡선을 이루며, CPU 점유율이 상승 하는 것을 알 수 있다. 전체 적으로 60%이상의 성능 향상이 있었으며, 특히 초당 입력 이벤트 1개이며, 등록된 질의가 1000개의 경우 170% 성능 향상이 되었다. 등록된 연속 질의 수가 증가함에 따라 성능 향상의 폭은 크게 증가함을 알 수 있다.

## 6. 결론 및 향후 연구

EPCIS는 EPCglobal Network Architecture에서 EPC 관련 정보 저장 및 검색을 제공하는 시스템이다. EPCIS에서 제공하는 연

속 질의(subscribe)는 시스템 자동화 및 재고 모니터링, 공급망 관리와 같은 다양한 응용에서 사용 될 수 있다.

본 논문에서는 13차원의 도메인을 여러 개의 질의 색인으로 구성하고, 등록된 질의 정보를 기반으로 선택률을 계산하고, 그에 의해 변경되는 동적 질의 실행 계획을 제안함으로써, EPCIS에서 시공간 EPC 데이터의 연속질의 처리에 대해 평균 60%의 성능이 향상이 가능하도록 하였다.

향후 연구로는 EPC, 시간, 공간 도메인의 선택률을 계산함에 있어, RFID 환경의 특성을 고려하여 계산함으로써, 향상된 연속 질의 처리 성능을 제공하는 것이 필요하다. 또한, 테스트에 사용된 연속 질의 및 이벤트를 균등분포가 아닌, 실제 응용 환경에서 발생 할 수 있는 데이터를 기반으로 테스트 하는 것이 필요하다.

## 6. 참고문헌

- [1] K. Romer, T. Schoch, F. Mattern and T. Dubendorfer, "Smart Identification Frameworks for Ubiquitous Computing Applications" Proc. of Pervasive Computing and Communications, pp. 256-262, 2003
- [2] EPCGlobal, "The EPCglobal Architecture Framework 1.2", 2007
- [3] EPCGlobal, "EPC Information Services (EPCIS) Version 1.0 Specification", 2007
- [4] Kun-Lung Wu, Shyh-Kwei Chen, Philip S. Yu, "Processing Continual Range Queries over Moving Objects Using VCR-Based Query Indexes.", MobiQuitous 2004: 226-235
- [5] D.V.Kalashnikov, S.Prabhakar, W.G. Aref, and S.E.Hambrusch, "Efficient evaluation of continuous range queries on moving objects.", In Proc. of 13th DESA, 2002.
- [6] Stefan Berchtold, Daniel A. Keim, Hans-Peter Kriegel. "The X-tree: An Index Structure for High-Dimensional Data Stefan Berchtold" Proc. 22nd VLDB Conf.. 28-39 1996