

---

# 다양한 응용의 스트림 데이터 모니터링을 위한 연속질의 모델링

조대수\*

Continuous Query Modelling for Various Kinds of Monitoring Services for Stream Data

Dae-Soo Cho\*

---

이 논문은 2010년도 동서대학교 "Dongseo Frontier Project" 지원에 의하여 이루어진 것임

---

## 요 약

다양한 센서들로부터 실시간으로 수집되는 데이터를 분석, 처리해야 하는 유비쿼터스 환경의 다양한 응용 서비스(모니터링 서비스)를 개발하기 위해서는 반드시 연속질의 처리 방법이 요구된다. 기존 연구에서 연속질의는 대부분 간격질의와 영역질의만으로 표현되고, 이러한 질의에 대한 처리 기법들이 제시되었다. 그러나 연속질의의 종류는 매우 다양하며, 속성 조건, 공간 조건, 시간 조건의 결합을 통해 표현될 수 있다.

이 논문에서는 연속질의의 종류를 분류하고, 여러 조건들의 결합을 통해 표현 가능한 연속질의의 모델을 제안하고자 한다. 이 논문은 연속질을 표현하는 질의의 모델을 제안함으로써, 연속질의 처리를 위한 다양한 후속 연구에 대한 방향을 제시하는데 기여하였다.

## ABSTRACT

Techniques for processing continuous queries are required to developing the various types of application services (monitoring services) in ubiquitous environment where the real-time data acquisition from a lot of sensors, analysis, and processing are required. In the previous works of the continuous queries, they have represented all of the continuous queries as the interval queries or region queries, and proposed some methods for processing theses queries. The types of continuous queries, however, are very various, and could be presented by combining the attribute conditions, spatial conditions, and temporal conditions. In this paper, I have classify the types of continuous queries, and have proposed the continuous query model which could be presented by combining those conditions. The contributions of this paper include that it proposes the query model representing the continuous queries and suggests future research directions.

## 키워드

연속질의의 모델링, 모니터링 서비스, 센서 네트워크, 스트림데이터

## Key word

Continuous Query Modelling, Monitoring Service, Sensor Network, Stream Data

---

\* 정회원 : 동서대학교 (교신저자, dscho@dongseo.ac.kr)

접수일자 : 2011. 05. 13

심사완료일자 : 2011. 06. 16

## I. 서론

스마트그리드, 유비쿼터스 헬스케어, 텔레메틱스/위치기반서비스(LBS), 텔레메트릭스(스마트SOC) 등의 유비쿼터스 환경에서 다양한 서비스들은 다음과 같은 공통적인 특징을 갖는다. 첫째, 데이터를 수집할 수 있는 방대한 수의 센서들이 있다. 둘째, 수집된 데이터를 유무선 망을 사용하여 실시간으로 관리서버에 전송하므로, 서버에서는 실시간으로 수집되는 스트림 데이터에 대한 처리를 효과적으로 수행해야 한다. 셋째, 서버에서는 전송된 데이터를 기반으로 사용자에게 맞춤형 서비스를 제공한다.

스트림 데이터를 효율적으로 관리하기 위해서 데이터 스트림 관리 시스템(DSMS, Data Stream Management System)이 도입되었다. 스트림 데이터는 다음과 같은 특징을 갖는다. 첫째, 시간이 지남에 따라서 계속해서, 빈번히 수집되며, 수집되는 양을 예측하기 어렵다. 둘째, 데이터의 수정, 삭제, 변경이 없으며, 계속해서 추가되는 특징이 있다.

DSMS에서는 스트림 데이터에 대한 연속질의를 효과적으로 수행할 수 있어야 한다. 기존 DBMS에서는 사용자가 요청하는 질의는 한번 수행(one-time query)하는 특징이 있으나, DSMS에서의 연속질의는 일정 기간 동안 시스템에서 관리되며, 주기적으로 혹은 특정 이벤트에 의해 반복적으로 수행(long-running continuous query)되는 특징이 있다.

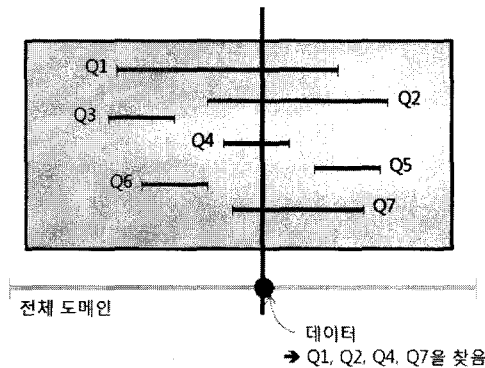


그림1. 간격질의의 예  
Fig 1. Example of Interval Query

현재까지 연속질의의 처리에 대한 다양한 연구들이 수행되어 왔으나, 대부분의 연구에서 연속질의는 간격질의(Interval Query) 또는 영역질의(Range Query)에 한정되어 있다. 예를 들면, 그림1과 같이 간격으로 표현된 질의들에 대해서, 특정 데이터를 포함하는 모든 간격 질의를 빠르게 검색하는 연구가 수행되었다. 즉, 그림에서 주어진 데이터에 대해서 Q1, Q2, Q4, Q7를 빠르게 검색하는 방법에 대한 연구가 수행되었다.

스트림 데이터에 대한 연속질의에서 간격질의와 영역질의가 차지하는 비중이 매우 높은 것은 사실이다. 그러나 연속질의에 대한 응용 분야가 다양해짐에 따라서, 다양한 종류의 연속질의에 대한 연구가 요구되고 있다. 따라서 이 논문에서는 연속질의의 종류를 분류하고, 여러 조건들의 결합을 통해 표현 가능한 연속질의 모델을 제안하고자 한다. 제안하는 연속질의 모델을 기반으로 질의 색인 기법, 질의처리 알고리즘 등의 다양한 연구가 파생될 수 있을 것으로 기대된다.

이 논문의 구성은 다음과 같다. 2장에서 관련연구를 살펴보고, 3장에서는 다양한 종류의 질의를 표현할 수 있는 연속 질의의 모델을 제안한다. 4장에서는 제안한 모델에 의해 연속질의를 표현하는 예제를 살펴보고, 5장에서 결론을 맺는다.

## II. 관련연구

스트림 데이터를 효과적으로 처리하기 위한 시스템으로, DSMS(Data Stream Management System)의 필요성에 대한 요구가 있었으며, 다양한 DSMS에 대한 연구가 수행되었다[1,2,3,4,5]. 브라운대의 Aurora[1]는 사각형과 화살표를 배열함으로써 질의의 계획을 생성할 수 있는 워크플로 기반의 DSMS 시스템을 제안하였다. 위스콘신대의 NiagaraCQ[2]에서는 동적인 웹 콘텐츠를 모니터링하기 위해 설계된 연속질의의 시스템으로서 스트림 데이터에 대해서 다수개의 연속질의를 효율적으로 수행하고 있다. 스탠포드대의 STREAM[3,4]는 메모리 관리 및 근사질의의 처리에 중점을 둔 일반적인 목적의 관계형 시스템을 제안하였다. 버클리대의 Telegraph[5] 질의평가 공유(질의색인) 및 적응적 질의 처리(상황에 따라 질의처리 순서가 변경됨)에 중점을 둔 연속질의의 처리 목적

의 시스템을 제안하였다.

스트림 데이터에 대한 연속질의 분야에 대한 연구는 프레디카트 매칭[6], 연속질의 처리[2,7], 이동객체 연속질의 처리[8,9]에 대한 연구가 있다. 프레디카트 매칭은 능동 데이터베이스, 규칙기반 시스템 등에서 ECA(Event-Condition-Action) 규칙을 효율적으로 처리하기 위해 연구되었다. 연속질의 처리는 새로운 데이터가 있을 경우(변화-기반 연속질의)에 또는 일정 시간간격마다 실행(타이머-기반 연속질의)되는 연속질을 효율적으로 수행하기 위해서 연구되었다. 이동객체 연속질의는 지속적으로 위치가 변동하는 이동 객체를 효과적으로 저장하고, 연속질을 효율적으로 수행하기 위해 연구되었으며, 질의영역이 고정되어 있는 정적 연속질의와 질의영역도 이동하는 동적 연속질의 처리로 구분된다.

연속질을 잘 처리하기 위해서는 색인 갱신 문제가 해결되어야 하며, 질의 재수행 문제를 해결해야 한다. 색인 갱신 문제는 질의색인을 이용하는 방법과 색인 갱신을 유예하는 방법으로 해결하고 있다. 질의 재수행 문제는 점진적 접근방법(Incremental evaluation)으로 해결하고 있다. 즉, 질의 결과에 변화를 주는 객체에 대해서만 질의를 수행함으로써 문제를 해결[8]하고 있다.

### III. 연속질의 모델링

이 논문에서는 연속질을 질의수행시간과 질의조건으로 모델링하고자 한다.

$$\text{연속질의} = \text{질의수행시간} + \text{질의조건}$$

#### 3.1. 질의수행시간

질의수행시간은 수행기간(Duration) QD와 수행간격(Interval) QI로 표현될 수 있다. 수행기간은 연속질의의 생명주기를 의미하며, 수행간격은 연속질의가 반복 수행되는 간격을 의미한다.

예들 들어, “현재 315번 온도센서의 온도를 검색하세요.”와 같은 질의는 한번만 수행되는 질의인 반면에, 다

음과 같은 질의는 반복적으로 수행되는 연속질의로서 질의수행기간 QD와 질의수행간격 QI를 갖는다.

#### (1) 질의수행시간이 구체적으로 명시된 경우

예: 지금부터 315번 온도센서의 온도를 매 1분마다 5시간동안 검색하세요.  
- 질의수행시간이 명시: QD(5시간)  
- 질의수행간격: QI(1분)

연속질의는 5분 동안 메모리에 상주되어 1분마다 주기적으로 실행되며, 5분 후에는 메모리에서 제거된다.

#### (2) 질의수행시간이 명시되지 않은 경우

예: 지금부터 315번 온도센서의 온도를 지속적으로 검색하세요.  
- 질의수행시간이 명시: QD(INF)  
- 질의수행간격: QI(0)

QD(INF)는 별도의 연속질의 종료 명령이 있기 전까지 계속해서 무한대(Infinte)로 실행되는 것을 의미한다. QI(0)의 의미는 연속질의의 반복 주기를 명시하지 않은 경우로서, 이 경우에는 연속질의에 영향을 미치는 스트림 데이터에 변화가 있을 경우에 즉시 실행되는 것을 의미한다. 즉, 특정한 시간 주기로 실행되는 것이 아니라, 새로운 데이터가 입력될 때마다 실행되는 것을 의미한다.

#### 3.2. 질의조건

질의조건은 공간(S, Spatial)조건, 시간(T, Temporal)조건, 속성(A, Attribute)조건으로 표현한다. 또한 각 조건들을 결합하여, 다양하게 표현할 수 있다. 그림 2와 같이 각각의 개별조건들을 결합하여, ST조건, AT조건, SA조건, STAT조건으로 표현이 가능하다.

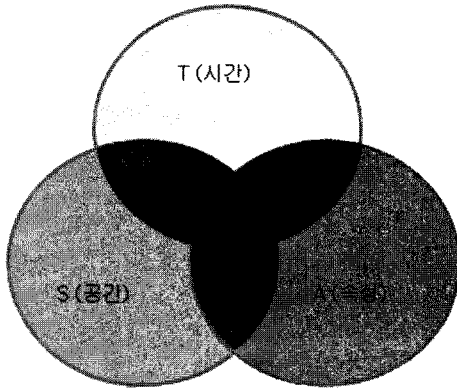


그림2. 다양한 질의조건  
Fig 2. Various Kinds of Query Condition

#### IV. 연속질의 표현 예제

이 논문에서 제안하는 연속질의 모델은 다양한 응용 분야에서 발생할 수 있는 모든 종류의 연속질의를 표현할 수 있다. 이 절에서는 스마트 그리드 응용서비스, 위치기반 광고 서비스, 헬스케어 서비스, 클린룸 및 어린이 보호 구역 모니터링 서비스 등의 실례를 통해 이 논문에서 제안하는 연속질의 모델의 효용성을 검증한다.

기존의 연구에서 다루고 있는 간격질의는 속성조건 질의로 표현이 가능하고, 영역질의는 공간조건 질의로 표현이 가능하므로, 별도의 예제를 제시하지는 않는다. 대신에 이 논문에서 제안하는 속성과 공간이 결합된 형태의 질의와, 시간에 따른 속성 및 공간의 변화관계에 대한 질의에 대한 예제를 다음과 같이 제시한다.

##### (1) 공간(S)조건

- **Region(MBR):** 주어진 MBR(Minimum Boundary Rectangle) 영역 내에 있다.
- **Location():** 현재의 위치를 보고한다.

##### (2) 시간(T)조건

- **Enter():** S(또는 A) 조건의 영역(또는 범위) 내로 진입한다.
- **Leave():** S(또는 A) 조건의 영역(또는 범위) 내에서 밖으로 진출한다.
- **Stay(Value):** S(또는 A) 조건의 영역(또는 범위) 내에 Value 시간(초)동안 머무른다.
- **Difference(OP, Value):** 이전 값과 현재 값의 차이가 Value 대비 OP하다. (OP : EQ | NE | GE | LE | GT | LT)
- **ChangeRate(OP, Value):** 이전 값과 현재 값의 단위 시간(초)당 변화율이 Value 대비 OP하다.

##### (3) 속성(A)조건

- **Interval(Attribute, Range):** 주어진 Attribute의 값이 주어진 Range 범위 내에 있다.
- **Value(Attribute):** 주어진 Attribute의 현재 값을 보고한다.

##### (1) 공간속성 결합질의 SA

스마트 그리드와 같이 모든 가정에 스마트미터(전기 검침)를 통해 원격으로 전기사용량을 검침할 수 있는 환경에서 발생할 수 있는 질의로서, 다음과 같은 공간 정보가 포함될 수 있다.

- “사직동 지역(R1)내의 스마트 미터 중에서, 전력 사용량이 300KW가 넘는 스마트 미터를 5분마다 지속적으로 검색하세요.”  
⇒ 스마트 미터가 위치한 공간에 대한 영역질의(사직동 지역 내의)와 스마트 미터가 측정하는 속성에 대한 간격질의(“300KW가 넘는”)가 결합된 형태의 연속질의임  
⇒ 연속질의 표현:  $Region(R1) \wedge Interval(\text{전력사용량}, 30KW, INF) \wedge QD(INF) \wedge QI(5분)$

##### (2) 공간관계 질의 ST

위치기반 광고 서비스는 광고 대상을 선정할 때, 위치 정보를 활용해서 광고 대상물 주변에 있는 사람들에게만 광고를 전송하는 것으로, 다음과 같은 연속질의를 표현할 수 있어야 한다.

- “음식점 반경 100미터 영역(R2)으로 진입하는 사람들을 지속적으로 1분마다 검색하세요.”  
⇒ 이전 위치는 음식점 반경 100미터 이상이었으나,

현재 위치는 100미터 이내에 있는 사람을 검색하는 진입 관계 연속질의임

⇒ 연속질의 표현:  $Region(R2) \wedge Enter \wedge QD(INF) \wedge QI(1분)$

- “영화관 반경 500미터 이내(R3)에서 2시간 이상 머무르고 있는 사람들을 10분마다 지속적으로 검색하세요.”

⇒ 2시간 전부터 지금까지의 위치가 계속해서 영화관 반경 500미터 이내에 있는 사람을 검색하는 체류 관계 연속질의임

⇒ 연속질의 표현:  $Region(R3) \wedge Stay(2시간) \wedge QD(INF) \wedge QI(10분)$

### (3) 속성관계 질의 AT

종합병원은 환자 모니터링 시스템을 통해서, 각 병실의 환자 체온을 온도 센서에 의해 측정하여, 실시간으로 모니터링 센터로 전송하는 환경이 요구되며, 다음과 같은 관계 질의가 빈번히 발생한다.

- “체온이 정상범위(36.4~37.2)를 벗어나는 환자를 1분마다 지속적으로 검색하세요.”

⇒ 이전 체온 데이터는 정상범위 내였으나, 현재 체온 데이터가 정상범위 밖에 있는 환자를 검색하는 진출 관계 연속질의임

⇒ 연속질의 표현:  $Interval(체온, 36.4, 37.2) \wedge Leave \wedge QD(INF) \wedge QI(1분)$

### (4) 공간관계 속성관계 결합질의 STAT

클린룸 내에서는 온도의 변화에 대한 지속적인 모니터링이 필요하다.

- “3번 클린룸 내(R4)에 있는 온도센서 중에서 온도 차이가 2 이상 나는 센서를 1분마다 지속적으로 검색하세요.”

⇒ 온도 센서가 위치한 공간에 대한 영역질의(“3번 클린룸 내의”)와 이전 온도와 현재 온도의 차이를 검색하는 차이 관계 질의가 결합된 형태의 연속질의임

⇒ 연속질의 표현:  $Region(R4) \wedge Value(온도) \wedge Difference(GT, 2) \wedge QD(INF) \wedge QI(1분)$

- “3번 클린룸 내(R4)에 있는 온도센서 중에서 초당 온도변화가 1 이상인 센서를 1초마다 지속적으로 검색하세요.”

⇒ 온도 센서가 위치한 공간에 대한 영역질의(“3번 클린룸 내의”)와 이전 온도와 현재 온도의 변화율을 검색하는 변화율 관계 질의가 결합된 형태의 연속질의임

⇒ 연속질의 표현:  $Region(R4) \wedge Value(온도) \wedge ChangeRate(GT, 1) \wedge QD(INF) \wedge QI(1초)$

어린이 보호구역 내에서는 차량의 속도가 지속적으로 모니터링 되어야 한다.

- “어린이 보호구역(R5)에 진입한 차량 중에서, 속도가 30km 이상인 차량을 1분마다 지속적으로 검색하세요.”

⇒ 이전 위치는 어린이 보호구역 밖이었으나, 현재 위치는 어린이 보호구역에 포함되는 차량을 검색하는 진입관계 질의와 차량의 속도를 측정하는 간격질의(“30km 이상”)가 결합된 형태의 연속질의임

⇒ 연속질의 표현:  $Region(R5) \wedge Enter \wedge Interval(속도, 30, INF) \wedge QD(INF) \wedge QI(1분)$

## V. 결 론

이 논문에서는 유비쿼터스 환경에서 다양한 응용 서비스 제공에 활용될 수 있는 연속질의 모델을 새롭게 제시하였다. 기존 연구에서의 연속질의는 특정속성에 대한 간격질의, 공간에 대한 영역질의 등 특정한 종류만을 다루고 있었으므로, 다양한 종류의 연속질의를 효과적으로 처리하기 위한 연구가 부족하였다. 이 논문에서 제안한 연속질의 모델은 속성조건, 공간조건, 시간조건 등이 결합된 형태로 표현할 수 있다. 특히 시간조건을 통해서 현재 데이터와 이전 데이터간의 관계를 표현할 수 있다. 또한 이 논문에서는 실제 다양한 응용 서비스에서 활용 가능한 연속질의에 대해서 제안한 연속질의 모델을 활용하여 표현한 예제를 보였다. 향후 다양한 종류의 연속질의를 효과적으로 처리할 수 있는 질의처리 알고리즘, 질의색인 기법 등에 대한 연구가 필요하다.

참고문헌

저자소개

- [1] D. Carney, U. Cetintemel, M. Cherniack, C. Convey, S. Lee, G. Seidman, M. Stonebraker, N. Tatbul, and S. B. Zdonik, "Monitoring streams - a new class of data management applications," In Proceedings of VLDB Conference, pp.215-226, 2002.
- [2] J. Chen, D. J. DeWitt, F. Tian, and Y. Wang, "NiagaraCQ: A scalable continuous query system for internet databases," In Proc. of the 2000 ACM SIGMOD Intl. Conf. on Management of Data, pp.379-390, 2000.
- [3] B. Babcock, S. Babu, M. Datar, R. Motwani, and J. Widom, "Models and Issues in Data Stream systems," Proc. of ACM PODS 2002, Madison, Wisconsin, United States, pp.1-16, 2002.
- [4] A. Arasu, B. Babcock, S. Babu, J. Cieslewicz, M. Datar, K. Ito, R. Motwani, U. Srivastava, D. Thomas, R. Varma, and J. Widom, "Stream: The stanford stream data manager," IEEE Data Engineering Bulletin, Vol.26, No.1, pp.19-26, 2003.
- [5] S. Chandrasekaran, O. Cooper, A. Deshpande, M. J. Franklin, J. M. Hellerstein, W. Hong, S. Krishnamurthy, S. Madden, F. Reiss, and M. A. Shah, "TelegraphCQ: Continuous dataflow processing," In Proceedings of ACM SIGMOD Conference, 2003.
- [6] E. N. Hanson and T. Johnson, "Selection Predicate Indexing for Active Database Using Interval Skip Lists," Information Systems, Vol.21, No.3, 1996.
- [7] K. Wu, S. Chen, and S. Yu, "Query indexing with containment-encoded intervals for efficient stream processing," Knowledge and Information Systems, Vol.9, No.1, pp.62-90, 2006.
- [8] B. Gedik, K.-L. Wu, P. S. Yu, and L. Liu, "Processing Moving Queries over Moving Objects using Motion-Adaptive Indexes," In IEEE Trans. Knowledge and Data Eng, Vol.18, pp.651-668, May 2006.
- [9] K.-L. Wu, S.-K. Chen, and P. S. Yu, "Incremental Processing of Continual Range Queries over Moving Objects," In IEEE Trans. Knowledge and Data Eng, Vol.18, pp.1560-1575, Nov. 2006.

조대수(Dae-Soo Cho)

한국해양정보통신학회논문지  
제13권 제8호 참조