



(19) 대한민국특허청(KR)
(12) 등록특허공보(B1)

(45) 공고일자 2011년11월30일
(11) 등록번호 10-1087766
(24) 등록일자 2011년11월22일

(51) Int. Cl.

G06F 17/21 (2006.01) G06F 17/30 (2006.01)

(21) 출원번호 10-2010-0093879

(22) 출원일자 2010년09월28일

심사청구일자 2010년09월28일

(56) 선행기술조사문헌

KR1020090002244 A

US20110191323 A1

류병걸 외 3명, "스트리밍 XML 데이터에 대한 빠른 트윅 질의 처리 기법", 한국정보처리학회 추계 학술대회 논문집, pp.65-68, 2010년 11월

(73) 특허권자

고려대학교 산학협력단

서울 성북구 안암동5가 1

(72) 발명자

이상근

서울특별시 동대문구 제기동 153번지 안암골 벽산 아파트 101동 203호

박상현

서울특별시 성북구 안암동2가 126-2번지 해송빌딩 305호

(74) 대리인

특허법인엠에이피에스

전체 청구항 수 : 총 15 항

심사관 : 신유철

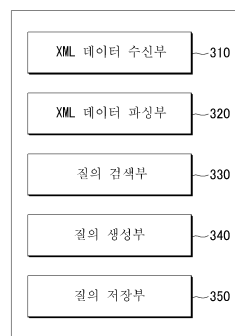
(54) 스트리밍 데이터에 대한 질의 처리장치 및 처리방법

(57) 요약

본 발명은 스트리밍 XML 데이터에 대하여 트윅 패턴 질의에 해당하는 데이터를 검색하는데 소요되는 자원 낭비를 최소화하는 스트리밍 데이터 처리장치 및 그를 이용한 처리방법을 제공한다.

본 발명의 스트리밍 데이터 처리장치는 스트리밍 XML 데이터를 파싱하여 상기 스트리밍 XML 데이터에 포함된 객체에 대한 시작 이벤트 또는 종료 이벤트를 검출하는 XML 데이터 파싱부 및 상기 시작 이벤트 또는 종료 이벤트가 검출된 객체와 트윅(twig) 패턴 질의를 비교하여, 트윅 패턴 질의에 해당하는 데이터를 검색하는 질의 검색부를 포함한다.

대표도 - 도3



특허청구의 범위

청구항 1

스트리밍 데이터에 대한 질의 처리장치에 있어서,

스트리밍 XML 데이터를 파싱하여 상기 스트리밍 XML 데이터에 포함된 객체에 대한 시작 이벤트 또는 종료 이벤트를 검출하는 XML 데이터 파싱부 및

상기 시작 이벤트 또는 종료 이벤트가 검출된 객체와 트위그(twig) 패턴 질의를 비교하여 트위그 패턴 질의에 해당하는 데이터를 상기 스트리밍 XML 데이터로부터 검색하는 질의 검색부

를 포함하는 스트리밍 데이터에 대한 질의 처리장치.

청구항 2

제 1 항에 있어서,

상기 질의 검색부는,

상기 검출된 객체에 대한 정보를 임시 저장부에 저장하고,

상기 트위그 패턴 질의의 루트 노드에 해당하는 객체와 동일한 객체에 대한 종료 이벤트가 검출된 경우, 상기 임시 저장부에 저장된 객체와 상기 트위그 패턴 질의를 비교하는 것인 스트리밍 데이터에 대한 질의 처리장치.

청구항 3

제 2 항에 있어서,

상기 질의 검색부는,

상기 임시 저장부에 저장된 객체와 상기 트위그 패턴 질의의 비교결과, 상기 임시 저장부가 상기 트위그 패턴 질의에 포함된 객체들을 모두 저장하고 있지 않은 경우, 상기 임시 저장부에 저장된 데이터를 삭제하는 것인 스트리밍 데이터에 대한 질의 처리장치.

청구항 4

제 2 항에 있어서,

상기 질의 검색부는,

상기 트위그 패턴 질의의 각 객체별로 상기 스트리밍 XML 데이터의 각 객체를 구분하여 상기 임시 저장부에 저장하는 스트리밍 데이터에 대한 질의 처리장치.

청구항 5

제 2 항에 있어서,

상기 질의 검색부는,

상기 스트리밍 XML 데이터의 각 객체의 계층관계에 대한 정보를 해시 포인트를 이용하여 상기 임시 저장부에 저장하는 스트리밍 데이터에 대한 질의 처리장치.

청구항 6

제 1 항에 있어서,

사용자에 의하여 입력된 검색 조건에 기초하여 상기 트위그 패턴 질의를 생성하는 질의 생성부를 더 포함하는 스트리밍 데이터에 대한 질의 처리장치.

청구항 7

제 1 항에 있어서,

상기 스트리밍 XML 데이터를 수신하여 상기 XML 데이터 파싱부로 전달하는 XML 데이터 수신부를 더 포함하는 스트리밍 데이터에 대한 질의 처리장치.

청구항 8

스트리밍 데이터에 포함된 특정 데이터를 추출하여 표시하는 사용자 단말 장치에 있어서,

스트리밍 XML 데이터를 수신하는 XML 데이터 수신부,

사용자에 의하여 입력된 검색 조건에 기초하여 트위그(twig) 패턴 질의를 생성하는 질의 생성부,

상기 수신된 스트리밍 XML 데이터를 파싱하여 상기 스트리밍 XML 데이터에 포함된 객체를 검출하는 XML 데이터 파싱부 및

상기 스트리밍 XML 데이터로부터 상기 트위그 패턴 질의에 해당하는 데이터를 검색하여 출력하는 질의 검색부를 포함하는 사용자 단말 장치.

청구항 9

제 8 항에 있어서,

상기 질의 검색부는,

상기 검출된 객체에 대한 정보를 임시 저장부에 저장하고,

상기 트위그 패턴 질의의 루트 노드에 해당하는 객체와 동일한 객체에 대한 종료 이벤트가 검출된 경우, 상기 임시 저장부에 저장된 객체와 상기 트위그 패턴 질의를 비교하는 것인 사용자 단말 장치.

청구항 10

제 9 항에 있어서,

상기 질의 검색부는,

상기 임시 저장부에 저장된 객체와 상기 트위그 패턴 질의의 비교결과, 상기 임시 저장부가 상기 트위그 패턴 질의에 포함된 객체들을 모두 저장하고 있지 않은 경우, 상기 임시 저장부에 저장된 데이터를 삭제하는 것인 사용자 단말 장치.

청구항 11

스트리밍 데이터에 대한 질의 처리방법에 있어서,

스트리밍 XML 데이터를 파싱하여 상기 스트리밍 XML 데이터에 포함된 객체에 대한 시작 이벤트 또는 종료 이벤

트를 검출하는 단계,

상기 검출된 객체에 대한 정보를 임시 저장부에 저장하는 단계 및

상기 임시 저장부에 저장된 객체에 대한 정보와 트위그(twig) 패턴 질의를 비교하여 트위그 패턴 질의에 해당하는 데이터를 상기 스트리밍 XML 데이터로부터 검색하는 단계

를 포함하는 스트리밍 데이터에 대한 질의 처리방법.

청구항 12

제 11 항에 있어서,

상기 임시 저장부에 저장하는 단계는,

상기 트위그 패턴 질의의 각 객체별로 상기 스트리밍 XML 데이터의 각 객체를 구분하여 저장하는 단계를 포함하는 스트리밍 데이터에 대한 질의 처리방법.

청구항 13

제 11 항에 있어서,

상기 임시 저장부에 저장하는 단계는

상기 스트리밍 XML 데이터의 각 객체의 계층관계에 대한 정보를 해시 포인트를 이용하여 저장하는 단계를 포함하는 스트리밍 데이터에 대한 질의 처리방법.

청구항 14

제 11 항에 있어서,

상기 트위그 패턴 질의에 해당하는 데이터를 검색하는 단계는

상기 트위그 패턴 질의의 루트 노드에 해당하는 객체와 동일한 객체에 대한 종료 이벤트가 검출된 경우, 상기 임시 저장부에 저장된 객체와 상기 트위그 패턴 질의를 비교하는 것인 스트리밍 데이터에 대한 질의 처리방법.

청구항 15

제 11 항 또는 제 13 항에 있어서,

상기 트위그 패턴 질의에 해당하는 데이터를 검색하는 단계는

상기 임시 저장부에 저장된 객체와 상기 트위그 패턴 질의의 비교결과, 상기 임시 저장부가 상기 트위그 패턴 질의에 포함된 객체들을 모두 저장하고 있지 않은 경우, 상기 임시 저장부에 저장된 데이터를 삭제하는 단계를 포함하는 스트리밍 데이터에 대한 질의 처리방법.

명 세 서

기술 분야

[0001] 본 발명은 스트리밍 XML 데이터로부터 트위그 패턴 질의를 효율적으로 처리할 수 있는 스트리밍 데이터에 대한 질의 처리장치 및 처리방법에 관한 것이다.

배경 기술

[0002] XML은 W3C(World Wide Consortium)의 정보의 표현 및 교환을 위한 데이터 표준으로서 현재 다양한 웹상의 정보를 포함한 다양한 정보 시스템 분야에서 사용되고 있다. XML로부터 사용자가 원하는 데이터를 얻기 위한 기존

의 대부분의 기술은 XML을 데이터베이스와 같은 고정된 저장소에 유지하고 필요할 때마다 XPath와 같은 질의를 수행하여, 원하는 검색결과를 얻었다.

[0003] 이와 같이 항상 접근 가능한 형태의 XML과는 달리 스트리밍 XML 데이터는 고정된 저장소에 유지되지 않고, 서버로부터 사용자 단말기로 계속적으로 전송된다는 특성을 지닌다. 따라서 스트리밍 XML로부터 원하는 데이터만을 얻기 위해서는 실시간으로 다양한 형태의 사용자 질의에 대한 결과를 효과적으로 가져와야 하는 것이 필수적인 요소이다. 또한 기존 XML 기술에서 많이 사용되는 단순경로의 XPath 질의보다 원하는 정보에 대한 속성을 보다 상세하게 기술할 수 있는 트위그(Twig) 패턴 질의를 지원하기 위해서는 계산 복잡도를 줄이고 스트리밍 XML의 태그정보에 대해 메모리 상에서의 버퍼링을 최소화하는 것이 무엇보다도 중요하다.

[0004] 스트리밍 XML 데이터로부터 정보를 얻기 위한 기술로서 대한민국 특허 출원(출원 번호: 10-2004-0106611, “XML 형태로 표현된 스트리밍 데이터 처리를 위한 XML 질의어 기반 연속 질의 언어 시스템 및 방법”)에서는 기존의 XML 질의어에 스트리밍 데이터 검색을 위한 기능을 추가하여 스트리밍 데이터에 접근할 수 있도록 하였다. 그러나 XML 질의어에 종류에 따라 스트리밍 데이터를 처리하기 위한 연산 및 버퍼링 기법이 매우 다양해질 수 있으며 이는 시스템의 성능에 중요한 요소로 작용한다. 또한, 상기 특허 출원의 경우 XML 질의어가 가지는 특성을 전혀 고려하지 않았다. 예를 들어, 트위그 패턴 질의의 경우 기존의 XML 질의보다 질의 처리를 위한 많은 연산을 요구하고 질의 결과를 얻기 위한 버퍼링이 많아질 수 있으므로 이를 줄이기 위한 기법이 필요하다.

[0005] 이에 따라, 최근에는 스트리밍 XML 데이터에 대한 다중 질의를 처리하기 위한 기술이 제안되었다. 대한민국 특허 출원(출원 번호: 10-2007-0062064, “XML 데이터 스트림에 대한 다중 질의 간 연산 공유를 이용한 연속 질의 처리 장치 및 그 방법”)에서는 XML 질의 언어중 하나인 XQuery를 기반으로 다중 질의를 처리하는 기술을 제안하였다. 그러나 실제로 내부적으로 연산 및 버퍼링을 최소화하기 위한 구체적인 기법에 대해서는 기술한 바 없으며, 트위그 패턴 질의보다 오히려 연산이 복잡한 XQuery만을 고려하고 있다.

[0006] 위에서 기술한 바와 같이 기존의 기술의 경우 모두 스트리밍 XML 데이터로부터 원하는 데이터를 얻기 위한 전반적인 시스템을 제안하고 있으나, 질의 특성에 따라 스트리밍 되는 XML 데이터를 어떠한 자료구조상으로 저장하며, 그로부터 질의 결과를 어떻게 추출할 것인지에 관한 기술이 전혀 고려되고 있지 않다.

발명의 내용

해결하려는 과제

[0007] 본 발명의 일부 실시예는 스트리밍 XML 데이터에 대하여 트위그 패턴 질의에 해당하는 데이터를 검색하는데 소요되는 자원 낭비를 최소화하는 스트리밍 데이터 처리장치 및 그를 이용한 처리방법을 제공한다.

과제의 해결 수단

[0008] 상술한 기술적 과제를 달성하기 위한 기술적 수단으로서, 본 발명의 제 1 측면에 따른 스트리밍 데이터 처리장치는 스트리밍 XML 데이터를 파싱하여 상기 스트리밍 XML 데이터에 포함된 객체에 대한 시작 이벤트 또는 종료 이벤트를 검출하는 XML 데이터 파싱부 및 상기 시작 이벤트 또는 종료 이벤트가 검출된 객체와 트위그(twig) 패턴 질의를 비교하여, 트위그 패턴 질의에 해당하는 데이터를 검색하는 질의 검색부를 포함한다.

[0009] 또한, 본 발명의 제 2 측면에 따른 사용자 단말 장치는 스트리밍 데이터에 포함된 특정 데이터를 추출하여 표시하는 것으로, 스트리밍 XML 데이터를 수신하는 XML 데이터 수신부, 사용자에게 의하여 입력된 검색 조건에 기초하여 트위그(twig) 패턴 질의를 생성하는 질의 생성부, 상기 수신된 스트리밍 XML 데이터를 파싱하여 상기 스트리밍 XML 데이터에 포함된 객체를 검출하는 XML 데이터 파싱부 및 상기 스트리밍 XML 데이터로부터 상기 트위그 패턴 질의에 해당하는 데이터를 검색하여 출력하는 질의 검색부를 포함한다.

[0010] 또한, 본 발명의 제 3 측면에 따른 스트리밍 데이터 처리방법은 스트리밍 XML 데이터를 파싱하여 상기 스트리밍 XML 데이터에 포함된 객체에 대한 시작 이벤트 또는 종료 이벤트를 검출하는 단계, 상기 검출된 객체에 대한 정보를 임시 저장부에 저장하는 단계, 상기 임시 저장된 객체와 트위그(twig) 패턴 질의를 비교하여, 트위그 패턴 질의에 해당하는 데이터를 검색하는 단계를 포함한다.

발명의 효과

[0011] 전술한 본 발명의 과제 해결 수단에 의하면, 스트리밍 XML 데이터에 대한 트위그 패턴 질의의 처리를 최적화하여, 검색에 소모되는 리소스 낭비를 최소화할 수 있다.

[0012] 또한, 전술한 본 발명의 과제 해결 수단에 의하면, 스트리밍 XML 문서를 순차적으로 파싱하여 이벤트를 검출하고, XML 객체에 대한 정보와 트위그 패턴 질의를 실시간으로 비교 처리하여, 처리 속도를 향상시킬 수 있다.

도면의 간단한 설명

- [0013] 도 1은 본 발명의 처리 대상이 되는 스트리밍 XML 데이터의 구조를 설명하기 위한 도면이다.
- 도 2는 트위그 패턴 질의를 설명하기 위한 도면이다.
- 도 3은 본 발명의 일 실시예에 따른 스트리밍 데이터 처리장치를 도시한 도면이다.
- 도 4는 본 발명의 일 실시예에 따른 스트리밍 XML 데이터로부터 트위그 패턴 질의를 검색하는 방법을 도시한 도면이다.
- 도 5는 본 발명의 일 실시예에 따라 트위그 패턴 질의를 검색하는 방법을 설명하기 위한 도면이다.
- 도 6은 본 발명의 일 실시예에 따라 트위그 패턴 질의를 검색하는 방법을 설명하기 위한 도면이다.
- 도 7은 본 발명의 일 실시예에 따른 스트리밍 데이터 처리방법의 효과를 설명하기 위한 도면이다.

발명을 실시하기 위한 구체적인 내용

- [0014] 아래에서는 첨부한 도면을 참조하여 본 발명이 속하는 기술 분야에서 통상의 지식을 가진 자가 용이하게 실시할 수 있도록 본 발명의 실시예를 상세히 설명한다. 그러나 본 발명은 여러 가지 상이한 형태로 구현될 수 있으며 여기에서 설명하는 실시예에 한정되지 않는다. 그리고 도면에서 본 발명을 명확하게 설명하기 위해서 설명과 관계없는 부분은 생략하였으며, 명세서 전체를 통하여 유사한 부분에 대해서는 유사한 도면 부호를 붙였다.
- [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 처리 대상이 되는 스트리밍 XML 데이터의 구조를 설명하기 위한 도면이다.
- [0017] 도 1의 (a)를 참조하면, (b)에 도시된 XML 문서에 포함된 객체들의 계층구조를 확인할 수 있다. 즉, (b)에 도시된 XML 문서는 루트 객체(root) 및 총 5 개의 객체(A, B, C, D, E)를 포함하고 있다. 이때, B 객체 및 D 객체는 A 객체의 자식 노드이고, C 객체는 B 객체의 자식 노드이며, E 객체는 D 객체의 자식 노드이다.
- [0018] 이러한 계층 구조를 XML 문서는 (b)와 같이 표현하고 있다. 각각의 객체는 태그와 엘리먼트 표기에 의하여 구분될 수 있다. 예를 들어, A 객체는 시작태그(<A>)와 종료태그()에 의하여 정의된다.
- [0019] 따라서, (b)에 도시된 문서를 순차적으로 읽어들이어 파싱하게 되면, 각 태그가 어떠한 객체를 나타내는지, 시작 태그인지 종료태그인지, 각 객체의 계층관계를 확인할 수 있다. 즉, 문서에 기재된 순서에 따라 각 객체를 확인하고, 객체들의 계층관계를 확인할 수 있다.
- [0020] (b)의 문서에서 각 객체를 읽어들이는 순서(10)에 따라, (a)의 계층 구조가 완성될 수 있으며, 이는 마치 (a)에 도시된 화살표(20)의 진행방향에 따라 계층 구조를 형성하는 것과 같다.
- [0021] 즉, A 객체의 시작태그(<A>)를 읽은 후에 B 객체의 시작태그()를 읽게 됨에 따라, B 객체가 A 객체의 하위 객체임을 확인할 수 있다. 마찬가지로, B 객체의 시작태그()를 읽은 후에 C 객체의 시작태그(<C>)를 읽게 됨에 따라, C 객체가 B 객체의 하위 객체임을 확인할 수 있다. 다음으로, C 객체의 종료태그(</C>)를 읽게 됨에 따라, B 객체에서는 C 객체가 가장 하위 객체임을 확인할 수 있다. 이와 같은 과정에 따라, (a)에 도시된 계층 구조를 형성할 수 있다.
- [0022] 서버에 저장된 XML 데이터와 달리, 스트리밍 XML 데이터의 경우, 수신되는 대로 파싱을 수행하게 된다. 따라서, (b)의 문서와 같은 XML 문서를 순차적으로 읽어 들인 후에, 각 객체간의 계층관계를 확인하는 과정을 거치게 된다.
- [0023] 도 2는 트위그 패턴 질의를 설명하기 위한 도면이다.
- [0024] 본 발명에서는 스트리밍 XML 데이터로부터 트위그 패턴 질의에 해당하는 데이터를 검색한다. 트위그 패턴에서

는 총 3개의 객체가 포함되며, 각 객체간의 계층관계가 정의 된다. 즉, 트위그 패턴 질의는 최상위 객체 하나와 최상위 객체에 대한 두 개의 하위 객체를 포함한다.

- [0025] (a)의 경우, 최상위 객체(A)와 하위 객체(B, C)의 관계가 부모-자식 관계임을 나타내고 있다. 즉, 각 객체간의 관계를 '/' 로 표시하는 경우 부모-자식 관계를 나타낸다.
- [0026] (b)의 경우, 최상위 객체(A)와 하위 객체(B, C)의 관계가 조상-자손 관계임을 나타내고 있다. 즉, 각 객체간의 관계를 '/' 로 표시하는 경우 조상-자손 관계를 나타낸다. 조상-자손 관계는 각 객체 사이에 하나 이상의 객체를 포함할 수 있다. 또한, 조상-자손 관계는 부모-자식 관계를 포함할 수 있다
- [0027] 이와 같이, 검색하고자 하는 질의가 세 개의 객체를 포함하고, 각 객체간에 특정한 계층관계가 설정되어 있는 경우 트위그 패턴을 구성하게 되며, 이러한 트위그 패턴 질의에 해당하는 데이터를 검색하게 된다.
- [0028] 도 3은 본 발명의 일 실시예에 따른 스트리밍 데이터 처리장치를 도시한 도면이다.
- [0029] 스트리밍 데이터 처리 장치(300)는 XML 데이터 수신부(310), XML 데이터 파싱부(320), 질의 검색부(330), 질의 생성부(340), 질의 저장부(350)를 포함한다. 이러한 스트리밍 데이터 처리 장치(300)는 스트리밍 XML 데이터에 포함된 특정 데이터를 추출하여 각종 인터페이스 또는 응용 프로그램에 적용하여 표시하는 사용자 단말 장치의 형태로 구현될 수 있다.
- [0030] 참고로, 본 발명의 실시예에 따른 도 3에 도시된 구성 요소들은 소프트웨어 또는 FPGA(Field Programmable Gate Array) 또는 ASIC(Application Specific Integrated Circuit)와 같은 하드웨어 구성 요소를 의미하며, 소정의 역할들을 수행한다.
- [0031] 그렇지만 '구성 요소들'은 소프트웨어 또는 하드웨어에 한정되는 의미는 아니며, 각 구성 요소는 어드레싱할 수 있는 저장 매체에 있도록 구성될 수도 있고 하나 또는 그 이상의 프로세서들을 재생시키도록 구성될 수도 있다.
- [0032] 따라서, 일 예로서 구성 요소는 소프트웨어 구성 요소들, 객체지향 소프트웨어 구성 요소들, 클래스 구성 요소들 및 태스크 구성 요소들과 같은 구성 요소들과, 프로세스들, 함수들, 속성들, 프로시저들, 서브루틴들, 프로그램 코드의 세그먼트들, 드라이버들, 펌웨어, 마이크로 코드, 회로, 데이터, 데이터베이스, 데이터 구조들, 테이블들, 어레이들 및 변수들을 포함한다.
- [0033] 구성 요소들과 해당 구성 요소들 안에서 제공되는 기능은 더 작은 수의 구성 요소들로 결합되거나 추가적인 구성 요소들로 더 분리될 수 있다.
- [0034] XML 데이터 수신부(310)는 외부 서버로부터 스트리밍 형태의 XML 데이터를 수신한다. 예를 들면, 네트워크는 근거리 통신망(Local Area Network; LAN), 광역 통신망(Wide Area Network; WAN) 또는 부가가치 통신망(Value Added Network; VAN) 등과 같은 유선 네트워크나 이동 통신망(mobile radio communication network) 또는 위성 통신망 등과 같은 모든 종류의 무선 네트워크를 통해 스트리밍 XML 데이터를 수신할 수 있다.
- [0035] XML 파싱부(320)는 수신된 스트리밍 XML 데이터를 파싱하여, XML 데이터에 포함된 각각의 이벤트를 감지한다. 객체의 시작 태그가 독출된 경우 객체의 시작 이벤트가 발생하였음을 감지하게 되고, 각 객체의 종료 태그가 독출된 경우 객체의 종료 이벤트가 발생하였음을 감지하게 된다. 예를 들어, SAX(Simple API for XML) 파서와 같이 이벤트 처리방식의 파싱부를 통해 스트리밍 XML 데이터를 순차적으로 파싱한다.
- [0036] 질의 검색부(330)는 파싱된 스트리밍 XML 데이터의 노드와 질의 저장부(350)에 저장된 트위그 패턴 질의의 노드를 비교하여, 트위그 패턴 질의에 해당하는 데이터를 검색한다.
- [0037] 보다 구체적으로 살펴보면, 질의 검색부(330)는 파싱된 스트리밍 XML 데이터의 노드와 질의 저장부(350)에 저장된 트위그 패턴 질의의 노드를 비교하여 동일 여부를 판단한다. 또한, 트위그 패턴 질의에 해당하는 노드가 검색된 경우 이를 스택과 같은 임시 저장소에 저장하여, 전체 트위그 패턴 질의에 해당하는 데이터를 검색한다.
- [0038] 질의 생성부(340)는 사용자에게 의하여 입력된 검색어에 기초하여 트위그 패턴 질의를 생성한다. 일반 사용자의 경우 트위그 패턴 질의를 인식하지 못한 상태에서 검색어를 입력하게 되므로, 사용자의 검색 요청에 기초하여 트위그 패턴 질의를 직접 생성한다. 예를 들어, 최근 자동차의 네비게이션 등에서 사용되고 있는 TPEG(Transport Protocol Expert Group) 서비스는 디지털 방송매체 및 인터넷 등을 통해 교통 및 여행정보를 전송하는 프로토콜 규격으로서, 무선채널 전송을 위한 이진 규격과 교통정보 서버들간의 메시지 전송 및 인터넷 서비스를 위한 XML 규격으로 구성되어 있다. 이때, 각 사용자의 네비게이션은 스트리밍 XML 데이터를 수신하여

표시하게 되는데, 수신되는 모든 데이터를 표시하는 것은 아니며, 사용자에게 의하여 설정된 특정 조건에 해당하는 데이터를 표시하게 된다. 이와 같이, 사용자에게 의하여 설정된 특정 조건을 트위그 패턴 질의로 생성하게 된다.

- [0039] 질의 저장부(350)는 질의 생성부(340)에 의하여 생성된 트위그 패턴 질의를 저장한다. 질의 저장부(350)에 저장된 트위그 패턴 질의는 파싱된 스트리밍 XML 데이터와 비교된다.
- [0040] 이제 본 발명의 일 실시예에 따른 스트리밍 데이터 처리장치(300)에서 수행되는 스트리밍 XML 데이터에서 트위그 패턴 질의를 검색하는 방법을 살펴보기로 한다.
- [0041] 도 4는 본 발명의 일 실시예에 따른 스트리밍 XML 데이터로부터 트위그 패턴 질의를 검색하는 방법을 도시한 도면이고, 도 5와 도 6은 트위그 패턴 질의를 검색하는 방법을 설명하기 위한 도면이다.
- [0042] 먼저, 스트리밍 데이터 처리장치(300)가 스트리밍 XML 데이터를 수신한다(S410).
- [0043] 수신된 스트리밍 XML 데이터를 파싱하여, XML 데이터에 포함된 태그 이벤트를 검사한다(S420). 태그의 형태에 따라 각 객체의 시작 이벤트와 종료 이벤트를 확인할 수 있다. 또한, 시작 이벤트에 의하여 확인된 객체는 스택과 같은 임시 저장소에 저장한다.
- [0044] 예를 들어, 트위그 패턴 질의가 도 5의 (a)에 도시된 바와 같이, 객체(A)와 하위 객체(B, C)를 포함하고, 상위 객체와 하위 객체의 관계가 조상-자손 관계라고 가정한다. 다음으로, 도 5의 (b)에 도시된 바와 같은 계층 구조를 가진 스트리밍 XML 데이터가 수신된 경우, 파싱 결과에 따라 가장 먼저 루트 노드가 독출되고, 다음으로 A1 객체에 대한 시작 이벤트가 독출된다. 그에 따라 도 6의 (a)에 도시된 바와 같이, 스택에 A1 객체가 저장된다.
- [0045] 다음으로, 파싱 결과에 따라 B1 객체에 대한 시작 이벤트가 독출되고, 그에 따라 도 6의 (a)에 도시된 바와 같이, 스택에 B1 객체가 저장된다. 다음으로, 파싱 결과에 따라 B2 객체에 대한 시작 이벤트가 독출되고, 그에 따라 도 6의 (a)에 도시된 바와 같이, 스택에 B2 객체가 저장된다.
- [0046] 이와 같이, 스트리밍 XML 데이터를 파싱하고, 확인된 이벤트에 따라 스택에 각각의 객체를 임시 저장한다.
- [0047] 다음으로, 파싱된 결과에 기초하여, XML 데이터에 포함된 각 객체의 연결관계를 나타내는 해시 포인트를 생성한다(S430).
- [0048] 예를 들어, 도 6(a)에 도시된 바와 같이, 상위 객체(A1)와 하위 객체간의 관계(B1, B2)를 나타내는 해시 포인트(화살표로 표시됨)를 생성한다.
- [0049] 다음으로, 스트리밍 XML 데이터의 파싱 과정에서, 질의의 루트노드와 동일한 노드에 대하여 종료 이벤트가 발생하면, 태그 이벤트 검사 단계로 진행하지 않고, 스택에 포함된 각 객체가 질의에 해당하는지 여부를 판단한다(S440, S440).
- [0050] 예를 들어, 도 5의 (a)와 같은 트위그 패턴 질의에서는 질의의 루트노드는 A가 된다. 이와 동일한 객체에 대한 종료 이벤트가 검사되면, 태그 이벤트의 검사를 일시 중지한다. 그리고 스택에 저장된 결과와 트위그 패턴 질의가 동일한지 여부를 판단한다.
- [0051] 도 5의 (b)와 같은 스트리밍 XML 데이터의 경우, 스택에 A1, B1, B2가 저장되어 있는 상태라면, 사용자에게 의하여 입력된 질의와 질의의 형태는 동일하나, 노드가 상이하므로, 트위그 패턴 질의와 상이한 것으로 판단한다.
- [0052] 그러나, 스택에 A2, A3, B3, C1이 저장되어 있는 상태라면, 사용자에게 의하여 입력된 질의와 동일한 것으로 판단된다.
- [0053] 상기 판단결과 스택에 저장된 데이터가 질의에 포함된 모든 노드를 포함하고 있다면, 그에 기초하여 질의 결과를 도출한다(S460). 그러나, 스택에 저장된 데이터가 질의에 포함된 모든 노드를 포함하고 있지 않다면, 스택에 저장된 노드를 제거하고(S480), 일시 중지하였던 태그 이벤트 검사를 수행한다.
- [0054] 한편, 스택에 저장된 데이터가 질의에 포함된 모든 노드를 포함하고 있어 질의 결과를 도출하는 경우에도, 경우에 따라 스택에서 불필요한 노드를 제거한다.
- [0055] 즉, 도 5의 (b)에 도시된 바와 같이, 최상위 객체가 A2이고, 차상위 객체가 A3, 하위 객체가 B3, C1인 경우, A3, B3 및 C1뿐만 아니라, A2, B3 및 C1도 질의 결과에 해당하는 데이터이다.

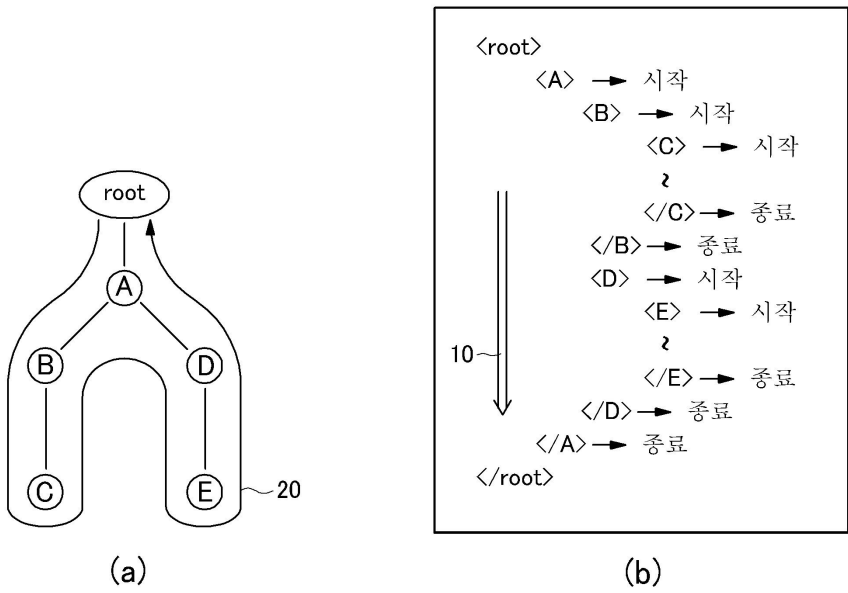
- [0056] 따라서, A3, B3 및 C1에 대한 질의 결과를 출력하는 경우에는 A3 객체만을 스택에서 제거한다. 다음으로, A2, B3 및 C1에 대한 질의 결과를 출력하는 경우에는 스택에 저장된 A2, B3 및 C1에 대한 객체를 모두 제거한다. 태그 이벤트 검사 결과에 따라 A3가 A2의 하위 객체임을 알 수 있고, B3와 상위 객체(A2, A3)와 연결되는 해시 포인트가 2개임을 알 수 있으므로, A3만을 선택적으로 제거할 수 있다.
- [0057] 이와 같이, 스택에 저장된 데이터가 질의와 일치하지 않는 경우 스택에 저장된 데이터를 삭제함으로써, 스트리밍 데이터 처리장치(300)의 리소스 낭비를 최소화할 수 있다. 또한, 스택에 저장된 데이터가 질의와 일치하는 경우에도 이미 출력된 결과에 대한 데이터는 삭제하여 리소스 낭비를 최소화한다.
- [0058] 도 7은 본 발명의 일 실시예에 따른 스트리밍 데이터 처리방법의 효과를 설명하기 위한 도면이다.
- [0059] 도 7의 (a)는 비교실험을 위해 사용한 XML 데이터들과 트위그 패턴 질의를 도시하고 있다. 각 실험 데이터는 기존 질의 성능 측정시 가장 많이 사용되는 DBLP 와 TreeBank XML 데이터이며 각각에 대하여 3개의 트위그 패턴 질의의 성능을 측정하였다.
- [0060] 도 7의 (b)는 질의 처리 시간을 측정한 그래프로서, 본 발명의 처리 방법을 사용할 경우 질의 처리 시간이 현저히 단축되는 것을 확인할 수 있다.
- [0061] 도 7의 (c)는 버퍼 사용량을 측정한 그래프로서, 본 발명의 처리 방법이 기존의 방법에 비해 작은 버퍼를 사용하고 있음을 확인할 수 있다.
- [0062] 이와 같이, 본 발명의 처리 방법을 사용할 경우, 스트리밍 XML 데이터에 대한 트위그 패턴 질의 처리시 질의 처리를 위한 연산비용 및 버퍼링 비용이 작게 드는 것을 실험을 통해 확인할 수 있다.
- [0063] 본 발명의 일 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 메커니즘을 포함하며, 임의의 정보 전달 매체를 포함한다.
- [0064] 본 발명의 방법 및 시스템은 특정 실시예와 관련하여 설명되었지만, 그것들의 구성 요소 또는 동작의 일부 또는 전부는 범용 하드웨어 아키텍처를 갖는 컴퓨터 시스템을 사용하여 구현될 수 있다.
- [0065] 전술한 본 발명의 설명은 예시를 위한 것이며, 본 발명이 속하는 기술분야의 통상의 지식을 가진 자는 본 발명의 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.
- [0066] 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

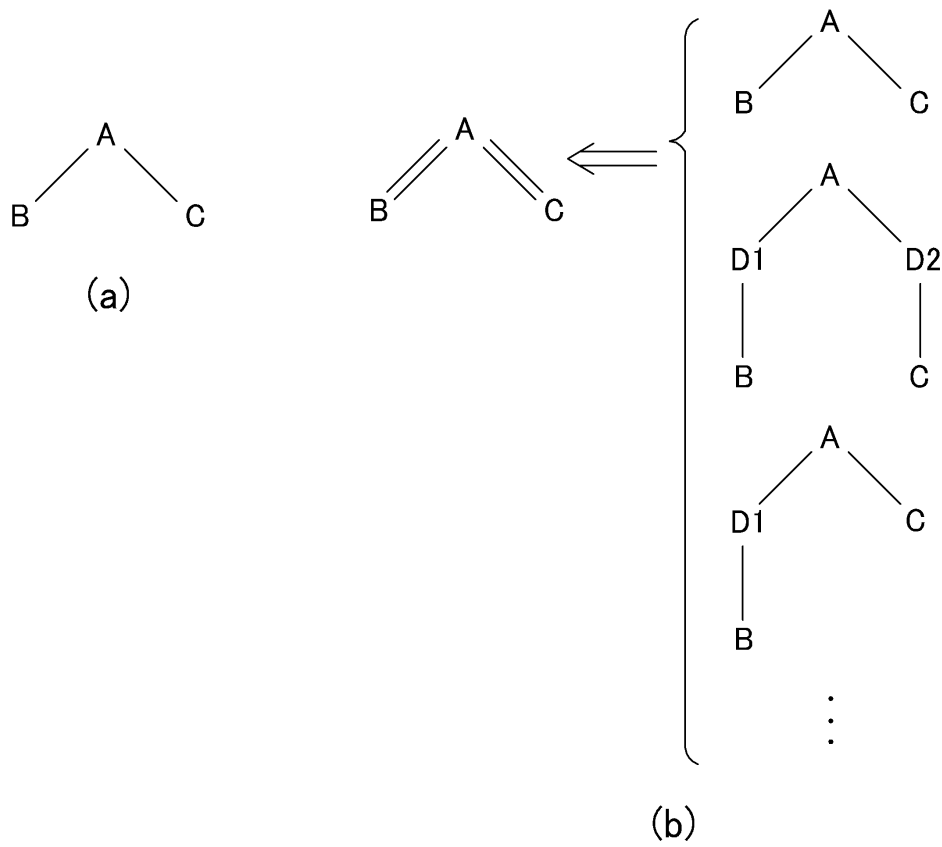
- | | | |
|--------|--------------------|------------------|
| [0067] | 300: 스트리밍 데이터 처리장치 | 310: XML 데이터 수신부 |
| | 320: XML 데이터 파싱부 | 330: 질의 검색부 |
| | 340: 질의 생성부 | 350: 질의 저장부 |

도면

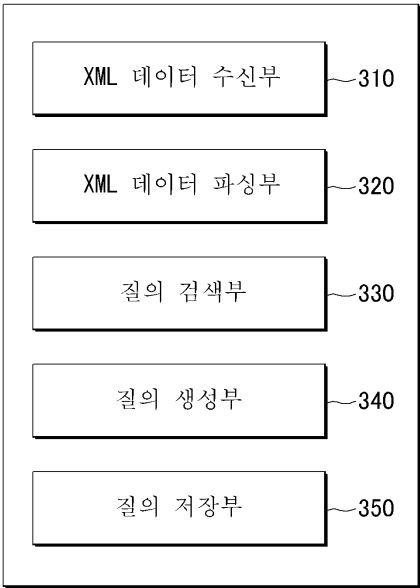
도면1



도면2

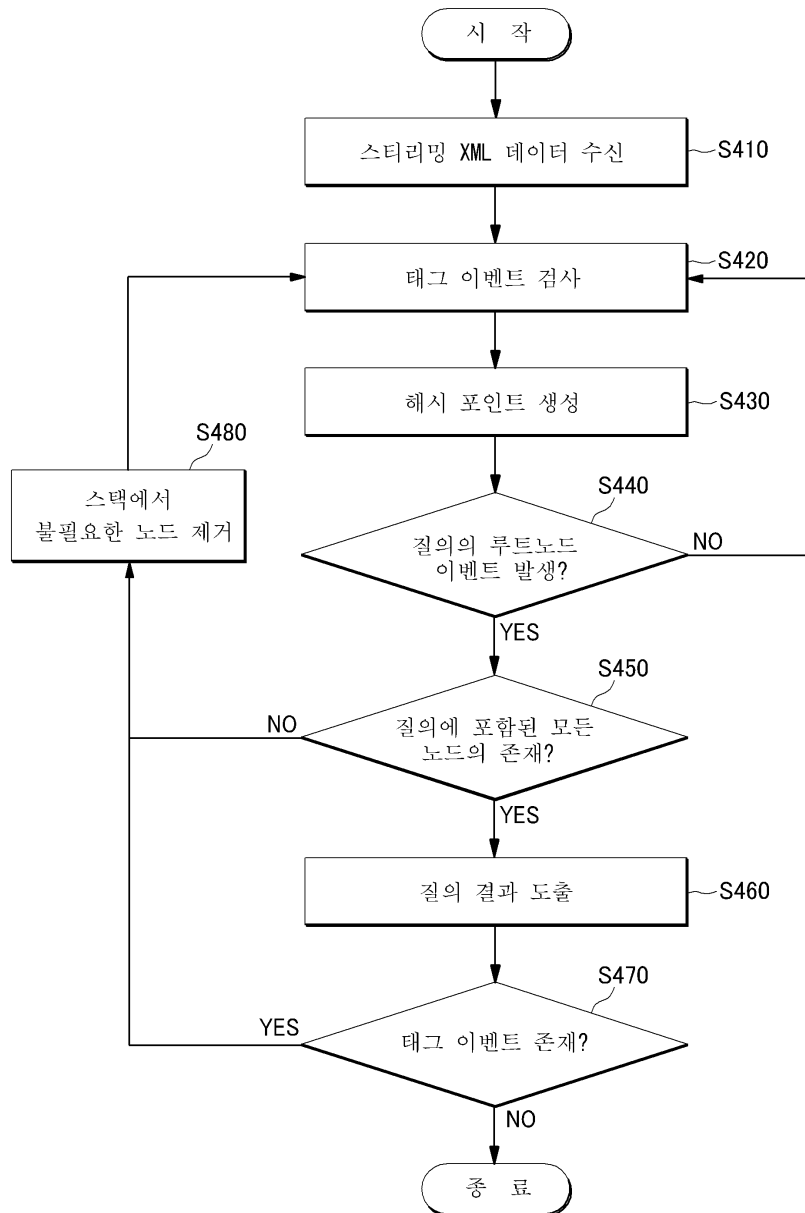


도면3

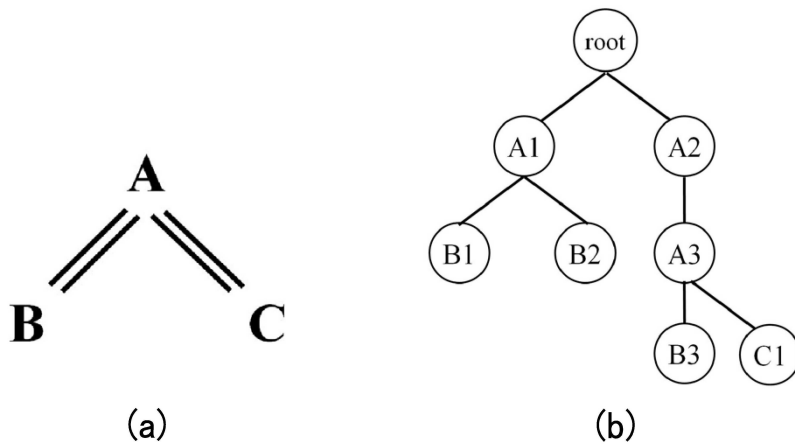


300

도면4



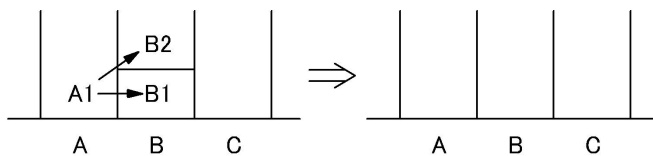
도면5



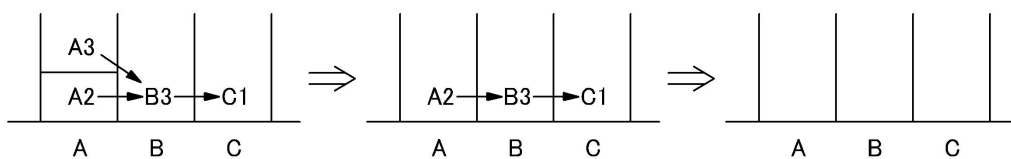
root 시작 → A1 시작 (스택에 A1 저장) → B1 시작 (스택에 B1 저장) → B1 끝
 → B2 시작 (스택에 B2 저장) → B2 끝
 → A1 끝 (질의처리수행, 스택에서 A1, B1, B2 제거)
 → A2 시작 (스택에 A2 저장) → A3 시작 (스택에 A3 저장) → B3 시작 (스택에 B3 저장)
 → B3 끝 → C1 시작 (스택에 C1 저장) → C1 끝
 → A3 끝 (질의처리수행, 질의결과도출. 스택에서 A3만 제거)
 → A2 끝 (질의처리수행, 질의결과도출, 스택에서 A2, B3, C1 제거) → root 끝

(c)

도면6



(a)



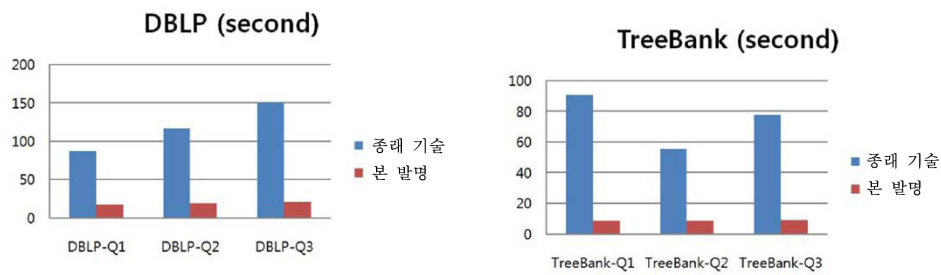
(b)

도면7

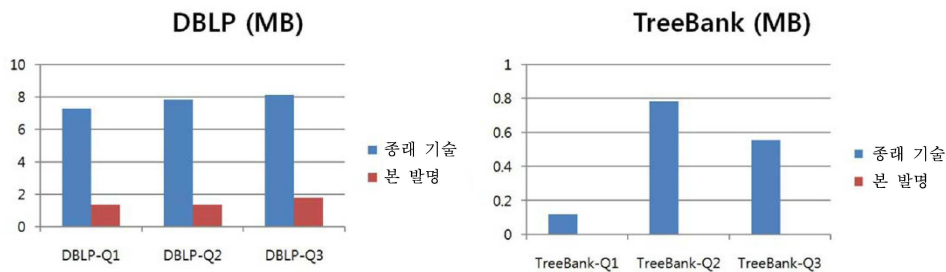
DBLP (180MB)	
Q1	/dblp/inproceedings[title#]/author#
Q2	/dblp/inproceedings[title# and booktitle#]/author#
Q3	/dblp/inproceedings[title# and booktitle# and year#]/author#

TreeBank (160MB)	
Q1	//S[./VP[./JJ[./VBD]]//NP[./WP]//DT#
Q2	//S//NP[./IN][./VBN]//JJ#
Q3	//S[./NP[./DT][./NN]]//PP[./TO]//NN#

(a)



(b)



(c)