

생물 정보원에 대한 통합 질의를 지원하는 그래픽 사용자 인터페이스 개발

정채영[†], 조찬제^{**}, 박홍원^{***}, 박은경^{****}, 김현주^{*****}, 배종민^{*****}

요 약

XML 기반의 미디어이터 시스템은 분산된 이질의 정보원들을 가상적으로 통합하는 방법 중의 하나이다. 이는 각 정보원을 하나의 가상의 XML 데이터베이스로 간주하고 XML 질의어로서 정보원에 접근한다. 본 논문은 사용자가 각 생물 정보원에 대한 XML 뷰를 융통성 있게 정의하고, 또한 정의한 XML 뷰를 기반으로 통합된 XML 질의를 쉽게 생성할 수 있는 사용자 인터페이스에 대한 설계 개념과 개발 결과를 제시한다. 개발된 인터페이스는 통합 스키마에 대하여 XML 뷰나 XML 질의를 손쉽게 생성할 수 있을 뿐 아니라, 하나 이상의 생물 정보원에 대한 질의 결과의 일부를 다른 생물 정보원의 입력 자료로 사용하는 워크플로우 형태의 질의 인터페이스 기능을 제공한다.

Development of an XML Query Interface to Support Integrated Queries for Biological Information Sources

Chai-Young Jung[†], Chan-Je Cho^{**}, Hong-Won Park^{***}, Eun-Kyung Park^{****},
Hyun-Ju Kim^{*****}, Jong-Min Bae^{*****}

ABSTRACT

An XML-based mediator system is the one of the methods to virtually integrate distributed heterogeneous information sources. It considers each information source as the virtual XML database and accesses the information source by XML query language. In this paper, we present a design concept and a development result of the graphic user interface that generates XML views for each biological information sources with flexibility and XML queries on the views. In addition to the easy generation of XML views and queries for the integrated schema, the system supports the form of the workflow, in which the part of the query result against biological information sources is used as input data of another source.

Key words: Query Interface(질의 인터페이스), Mediator(미디어이터), Data Integration(데이터 통합), Bioinformatics(바이오인포매틱스)

※ 교신저자(Corresponding Author) : 배종민, 주소 : 경남 진주시 가좌동 900번지(660-701), 전화 : 055)751-5995, FAX : 055)762-1944, E-mail : jmbae@gnu.ac.kr
접수일 : 2006년 2월 24일, 완료일 : 2006년 11월 27일

[†] 정희원, 경상대학교 컴퓨터과학과
(E-mail : chaeng@paran.com)

^{**} 경상대학교 컴퓨터과학과
(E-mail : cjcho@thinkfree.com)

^{***} 경상대학교 컴퓨터과학과

(E-mail : hongsi80@hanmail.net)

^{****} 경상대학교 컴퓨터과학과

(E-mail : pek1028@hanmail.net)

^{*****} 정희원, 진주산업대학교 컴퓨터공학부 조교수

(E-mail : khj@jinju.ac.kr)

^{*****} 정희원, 경상대학교 컴퓨터과학부/컴퓨터정보통신연구
구원 교수

※본 연구는 학술진흥재단(KRF-2004-002-D00380)의 지원으로 수행되었음.

1. 서 론

생물학적 실험환경의 발전으로 인하여 대량의 생물 데이터들이 생성되었고 이를 분석하고 처리하기 위한 도구들과 이들 데이터를 관리하는 수많은 생물 정보원들이 생겨났다[1]. 생물학적 실험 과정에서 의미 있는 결과를 도출하기 위해서는 여러 생물 정보원들이 제공하는 결과를 통합분석해서 또 다른 의미 있는 지식을 추론해 나가는 과정이 필요한데 이 과정을 효율적으로 지원하기 위해서는 생물 정보원을 물리적, 혹은 가상적으로 통합하거나, 혹은 특정 실험 환경에 적합한 작업의 흐름에 따라서 특화된 통합 시스템이 필요하다.

생물 정보원의 통합의 어려움은 각 정보원에 저장된 데이터의 타입이 다양하고, 데이터를 저장하는 모델 또한 다양하며, 접근하는 방법과 질의하는 방법, 그리고 질의에 대한 결과 양식이 다양하기 때문이다 [2]. 분산된 생물 정보원을 물리적으로 혹은 가상적으로 통합하기 위해서는 이와 같은 정보원들의 이질성을 해결해야 한다. 본 연구에서는 다수의 이질 생물 정보원을 통합하기 위해 미디어이터 시스템을 이용한다. 미디어이터 시스템은 데이터의 물리적 이동 없이 가상의 뷰를 이용하여 정보원을 통합한다.

본 연구에서 제시하는 미디어이터 시스템은 각 정보원에 대한 뷰를 XML을 기반으로 정의한다. 각 생물 정보원들의 스키마는 XML Schema로써 표현하며, 질의어는 XQuery를 사용하고 사용자가 새로운 뷰를 정의할 때도 XQuery를 사용한다. 질의에 대한 결과는 XML 문서이다. 따라서 사용자의 다양한 요구를 수용할 수 있는 XQuery 생성기의 개발이 필요하다.

정보원 통합 시스템에서 질의 인터페이스의 기본적인 기능은 질의의 생성과 조작이라고 할 수 있다. 기존의 대부분의 질의 인터페이스는 통합 스키마에 대하여 질의를 생성하는 과정을 도와주며, 질의 결과를 브라우징하는 기능을 제공한다. 이는 정보원 통합 시스템이 제공해야 할 기본적인 도구라고 할 수 있다. 그런데, 생물 정보원 통합 시스템도 이와 같은 기능이 기본적으로 제공되어야 하지만, 생물 정보원 사용자의 작업과정에는 임의의 정보원으로부터 받은 질의 결과의 일부를 다른 생물 정보원의 입력 자료로 활용하고, 계속해서 그 정보원으로부터 받은 질

의 결과의 일부를 또 다른 생물 정보원의 입력 자료로 활용하는 사례가 빈번히 발생한다. 따라서 생물 정보원 통합 시스템에는 보통의 정보원 통합 시스템에서 제공하는 질의 인터페이스 뿐 아니라, 이와 같이 워크플로우를 따르는 정보원 통합 시스템의 기능도 함께 있으면 사용자에게 보다 강력한 질의 인터페이스를 제공할 수 있다.

이에 따라 본 연구에서는 먼저 통합 스키마에 대한 질의 생성을 도와주는 질의 인터페이스에 대하여 논한 뒤에, 생물학자의 워크플로우를 효과적으로 지원하기 위한 인터페이스의 설계 개념과 개발 결과를 제시한다.

본 논문의 구성은 다음과 같다. 2장에서는 기존의 정보원 통합 시스템에서의 사용자 인터페이스에 대해 살펴보고 3장에서는 질의 처리를 수행하는 통합 질의 시스템의 전체적인 구조와 사용자 인터페이스에 대해 소개한다. 4장에서는 통합 스키마에 대한 질의 인터페이스와 워크플로우를 지원하는 인터페이스에 대하여 논한다. 5장에서는 구현 결과를 제시하고 6장에서 결론 및 향후 과제에 대해 논한다.

2. 관련연구

여기에서는 기존의 미디어이터 시스템에서의 사용자 인터페이스에 대해 살펴본다.

TAMBIS(Transparent Access to Multiple Bioinformatics Information Sources)[3, 4]는 GRAIL 언어를 이용하여 가상의 통합 데이터베이스를 만들고 사용자는 통합 스키마를 바탕으로 그래픽 도구를 이용하여 질의를 하면 BioKleisli 시스템의 램퍼를 이용하여 질의를 수행한다. TABMIS의 지식 기반 사용자 인터페이스(Knowledge-Driven Graphical User Interface)는 크게 온톨로지(Ontology) 기반의 뷰를 보여주는 온톨로지 브라우저와 질의 생성과 조작을 담당하는 질의 생성기로 구성된다. 온톨로지 브라우저는 특정 뷰를 중심으로 관련된 여러 내용들과 그들 사이의 관계를 시각적으로 표현한다. 질의 생성기는 GUI를 통해 검색하고자 하는 내용들의 관계를 표현하면 선언적 언어인 GRAIL 질의를 생성하고 이를 미디어이터에 전달한다.

MIX(Mediation of Information Using XML)[5]는 범용 정보원 통합 시스템이다. MIX는 BBQ(Blended

Browsing and Querying)[6, 7]라는 그래픽 사용자 인터페이스를 제공한다. BBQ는 미디어이터가 제공하는 뷰 DTD를 기반으로 통합된 뷰 DTD를 조회하고 XMAS 질의를 생성하는 기능을 제공한다. BBQ는 뷰 DTD를 트리 형태로 표현하고 이를 이용하여 사용자는 드래그 앤 드랍 등을 이용하여 필요한 부분을 표현하면 XMAS 질의를 생성하고 미디어이터를 통해 이를 수행한다.

TIMBIS와 MIX가 제공하는 사용자 인터페이스의 기능은 범용 정보원 통합 시스템이 제공해야 할 기본적인 기능이지만, 생물 정보원 통합에 특화된 통합 질의 인터페이스로는 그 기능이 제한적이다.

한편 SDM(Scientific Data Management Center project)[8, 9]에서는 XWRAP[10] 랩퍼 시스템을 이용하여 사용자가 작성한 워크플로우를 기반으로 생물학적 질의를 수행하는 방법을 제공한다. 사용자는 WF-Pilot이라는 GUI 인터페이스를 이용해 자신이 찾고자 하는 정보와 관련된 추상 뷰들을 연결하여 AWF(abstract workflow)를 생성하고 WF-Composer에서 생성한 AWF를 EWF(executable workflow)를 변환하고 사용자는 WF-Pilot를 이용하여 변환된 EWF를 조회하거나 다시 수정할 수 있다. 변환된 EWF는 WF-Engine를 이용하여 관련된 XWRAP 랩퍼와 통신하여 원하는 결과를 얻을 수

있다. XWRAP은 범용 통합 시스템이 아니고 웹 정보원에 대한 랩퍼이기 때문에 대상이 되는 정보원에 제약이 있다.

본 논문에서 제시하는 인터페이스는 관계형 혹은 객체관계형 데이터베이스, 웹 자원 등의 통합에 모두 적용되는 범용 정보원 통합 시스템에서 사용할 수 있도록 통합 스키마에 대한 질의 인터페이스를 통하여 XML 질의와 XML 뷰의 생성을 효과적으로 지원하며, 생물 정보원 사용자의 특성에 맞도록 사용자 워크플로우를 효과적으로 지원하기 위한 기능도 함께 지원한다.

3. 정보원 통합 시스템 구조

사용자 인터페이스의 설계개념을 설명하기 위하여 먼저 개발된 정보원 통합 시스템 구조에 대하여 논한다.

3.1 시스템 구조

통합 시스템은 그림 1과 같이 크게 이질의 생물 정보원과 직접 대화하는 랩퍼와, 랩퍼를 관리하는 미디어이터 그리고 사용자 질의를 생성하는 사용자 인터페이스로 구성된다.

랩퍼는 실제 생물 정보원과 직접 대화하며 각각의

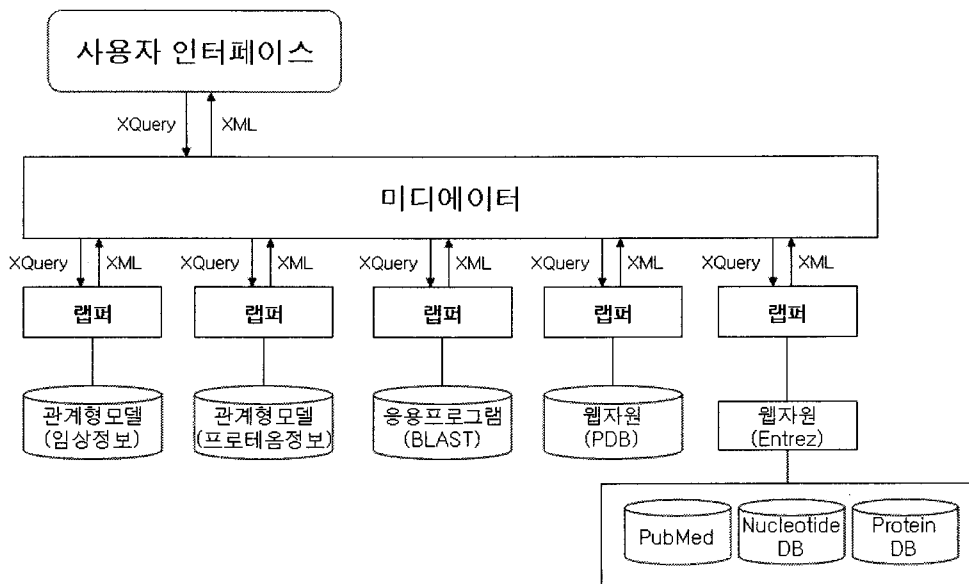


그림 1. 정보원 통합 시스템 구조

정보원이나 응용프로그램별로 별도의 랩퍼가 존재한다. 미디어이터로부터 받은 질의를 정보원에서 수행 가능하도록 하며 생물 정보원에서 수행된 질의 결과를 미디어이터로 전달한다. 미디어이터는 사용자로부터 생성된 질의를 분해하여 랩퍼에게 전달한다. 그리고 각 랩퍼에서 보내온 질의 결과를 통합하여 사용자가 원하는 질의 결과를 생성한다. 사용자 인터페이스는 통합 질의를 효과적으로 생성하기 위한 환경을 제공한다.

3.2 XML 뷰

이질의 정보원을 통합하기 위해서는 이질의 각 정보원들을 XML 데이터 모델로 사상함으로써 모든 정보원들을 XML 관점에서 볼 수 있게 한다. 이를 위해 XML 뷰를 구성하고 이를 기반으로 XML 질의를 행한다. 즉 사용자 관점에서 볼 때 모든 정보원들은 하나의 XML 데이터베이스로 간주된다. 논문에서는 XQuery를 이용하여 XML 뷰와 XML 질의를 작성하고 XML Schema를 이용하여 정보원의 스키마를 표현한다.

그림 2는 지역 XML 뷰와 전역 XML 뷰의 구성에 대한 예이다. 지역 정보원과 직접 대화하는 랩퍼는 각 정보원들에 대한 XML 뷰를 생성하는데 이를 지역 XML 뷰(Local XML view)라 한다. 지역 XML 뷰는 지역 정보원의 데이터 스키마와 XML 스키마 사이의 사상 관계를 표현한 것이라 볼 수 있다.

개발된 시스템은 사용자 정의 XML 뷰를 지원하는데, 이는 사용자의 목적에 맞는 스키마의 설계가 가능하고 사용자 질의 설계에 큰 융통성을 주는 한편, 사용 목적에 맞는 사용자 인터페이스에 대해 필요한 스키마만을 제공하는 방법을 제공할 수 있다. 사용자 정의 XML 뷰는 XQuery로 표현되기 때문에 이를 생성하는 사용자 인터페이스의 지원을 필요로 한다.

3.3 사용자 인터페이스

사용자 인터페이스는 통합 스키마에 대한 질의 인터페이스와 워크플로우 기반 질의 인터페이스 두 가지로 구성된다. 통합 스키마에 대한 질의 인터페이스는 미디어이터가 관리하는 XML 스키마를 트리구조

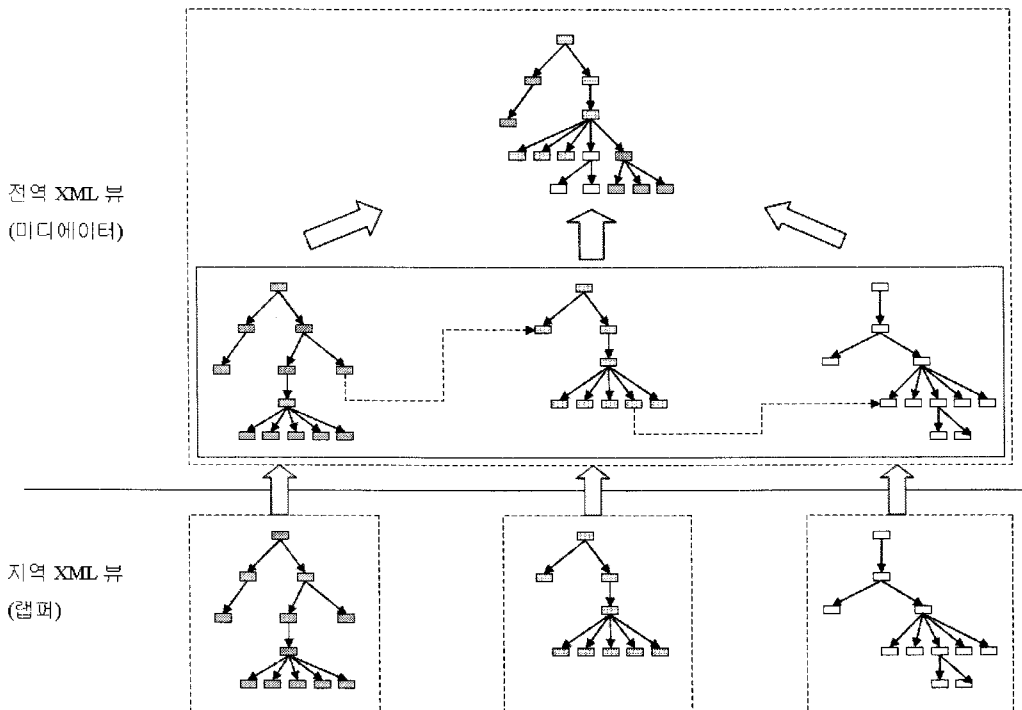


그림 2. 지역 XML 뷰와 전역 XML 뷰

로 제시한다. 사용자는 드래그 앤 드랍 형식으로 계층구조를 가진 다이어그램을 작성하면서 사용자 정의 XML 뷰 또는 사용자 질의를 생성한다.

워크플로우 기반 질의 생성기는 생물 정보원에 대한 검색 과정의 특성을 고려한 것이다. 생물학자들은 하나의 정보원으로부터 얻은 검색 결과의 일부를 바탕으로 다른 정보원에 접속하여 새로운 검색 결과를 얻는 일을 반복한다. 예를 들어 생물학자는 실험을 통해 얻는 단백질의 정보를 얻기 위해 자신의 로컬 데이터베이스에 저장된 정보를 이용하여 Blast 검색을 통해 유사한 단백질 서열을 획득하고 다시 Entrez의 Protein 검색을 통해 해당 단백질에 대한 정보를 얻는다. 또는 PubMed의 검색을 통해 관련된 문헌 검색을 수행하거나 PDB 검색을 통해 단백질의 구조 검색을 수행할 수도 있다. 워크플로우 기반 질의 생성기는 이처럼 일련의 작업 과정을 하나의 인터페이스를 통해 수행할 수 있는 방법을 제공한다. 워크플로우 기반 질의 생성기는 미리 정해진 형태의 질의를 생성하되 사용자가 필요한 조건을 입력하여 원하는 조건에 맞는 결과를 얻을 수 있다.

4. 질의 인터페이스

4.1 통합 스키마에 대한 질의 인터페이스

4.1.1 시스템 구조

그림 3은 통합 스키마에 대한 인터페이스의 개략적인 구조이다.

그림 3에서 뷰 트리 생성기는 미디어이터가 제공

하는 XML 스키마를 이용하여 이를 트리 형태로 생성하고 이를 사용자에게 보여준다. 뷰 트리 생성기는 스키마 파일을 DOM 트리로 파싱한 후 DOM 트리를 분석하여 뷰 트리를 생성한다. 질의 트리 생성기는 생성된 뷰 트리를 기반으로 드래그 앤 드랍 또는 사용자 입력을 통해 질의 트리를 생성한다. 질의 생성기는 완성된 질의 트리를 순회하여 XQuery를 생성하여 미디어이터에게 전달한다. 결과 뷰어는 미디어이터로부터 전달된 질의 결과를 사용자에게 보여준다.

4.1.2 XQuery 생성 방법

그래픽 XQuery 편집기에서 생성되는 질의 트리는 크게 요소 구성자와 FLWOR (FOR - LET - WHERE - ORDER - RETURN) 표현식을 나타내는 노드로 구성된다. XQuery에서 요소 구성자는 엘리먼트를 나타내고 FLWOR 표현식은 중간 결과에 대한 변수 바인딩과 탐색을 지원한다. 엘리먼트는 시작 엘리먼트와 끝 엘리먼트로 나뉘어진다. 질의 트리는 시작 엘리먼트를 나타내는 StartElement와 끝 엘리먼트를 나타내는 EndElement, 그리고 FLWOR 표현식을 나타내는 FlworElement로 이루어지며, 이들의 자료 구조는 서로 유사하다. 질의 트리를 구성하는 3개의 노드의 구조는 표 1과 같다.

StartElement의 intendCnt 필드는 트리의 높이를 나타내는데 이후 XQuery를 생성할 때 들여쓰기를 사용하는데 사용된다. node 필드는 요소 구성자를 구성하는데 필요한 정보를 가지는 Element 객체를 지시한다. construction 필드는 해당 노드가 가지는 자

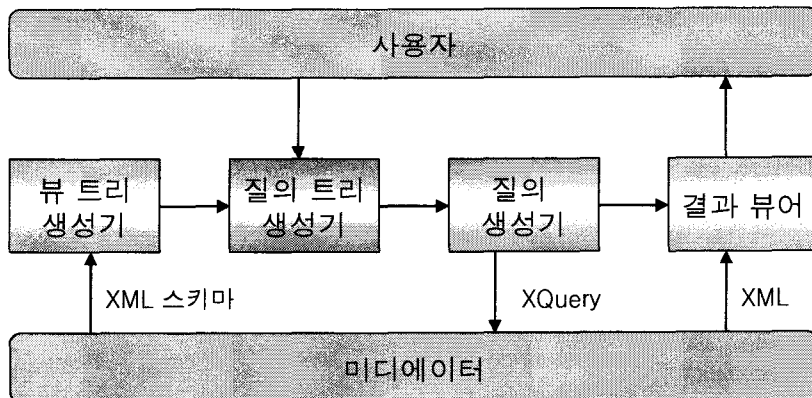


그림 3. 통합 스키마에 대한 질의 인터페이스 구조

표 1. 질의 트리 노드 구조

	StartElement	EndElement	FlworElement
필드	intendCnt	intendCnt	intendCnt
	node	node	flworNode
	construction		construction

식 노드들을 리스트 형태로 가진다. EndElement는 StartElement와 항상 쌍으로 존재하는데 자식 노드를 가지는 경우가 없기 때문에 construction 필드를 가지지 않으며 나머지 필드는 StartElement와 동일하다. FlworElement 노드의 구조 역시 StartElement와 유사하다. FlworElement 노드는 요소 구성자 대신 FLWOR 표현식과 관련된 정보를 유지하기 위한 Flwor 객체를 지시하는 flworNode 필드를 가진다.

요소 구성자에 대한 정보를 유지하는 Element 노드의 구조는 표 2와 같다. nodeName 필드는 표현하고자 하는 엘리먼트의 이름, nodePath는 엘리먼트가 가지는 값을 나타내는데 일반적으로 XPath 경로식이 된다. attr은 엘리먼트가 가지는 애트리뷰트에 대한 정보를 가지고 있는 애트리뷰트 객체를 지시한다.

FLWOR 표현식에 대한 정보를 유지하는 Flwor 노드의 구조는 표 3과 같다. variable 필드는 바인딩을 위한 변수명이고 source 필드는 바인딩을 위한 XML 뷰의 이름 또는 미리 선언된 변수명을 나타낸다. path는 선언한 변수가 바인딩된 경로식이 저장되면 conditionList는 변수 바인딩에 필요한 조건절을 포함한다. arrange 필드에는 결과의 정렬 기준이 저장된다.

그림 4는 제시된 XQuery과 같은 사용자 질의를 작성하고자 할 경우에, 생성되는 질의 트리의 구조와 질의 트리의 노드가 가지는 정보를 나타낸다. 그림에서 S는 StartElement, E는 EndElement, 그리고 F는

표 2. Element 자료 구조

필드	내용
nodeName	엘리먼트 이름
nodePath	엘리먼트가 가지는 경로식
attr	엘리먼트가 가지는 애트리뷰트에 대한 객체

표 3. Element 자료 구조

필드	내용
variable	변수명
source	스키마 이름 또는 변수명
path	변수에 바인딩 된 경로식
conditionList	조건식
arrange	정렬기준

FlworElement를 의미한다.

질의 트리를 생성하는 과정에서 GUI를 통해 사용자가 필요한 정보를 입력하거나 드래그 앤 드랍을 수행하는 과정에서 생성될 질의 트리가 유효한가를 검사하여 유효하지 않으면 해당되는 노드를 붙이지 않는다. 유효성을 검사하는 방법은 Element 객체일 경우 상위에 반드시 자신이 원하는 정보원의 Flwor 객체가 생성되어 있고 그 Flwor 객체에서 선언된 변수의 경로 이하의 노드일 경우에만 질의 트리에 추가한다. Flwor 객체를 추가하고자 할 경우에는 상위에 Flwor 객체가 항상 존재할 필요는 없으나 상위의 Flwor 객체에서 선언한 변수를 생성할 Flwor 객체에서 사용할 수 있기 때문에 상위의 Flwor 객체에 대한 사용 여부를 검사한다. 그리고 선택하는 스키마에 따라 그것에 해당하는 XML 뷰 트리를 사용자가 선택할 수 있게 Flwor 객체의 경로를 트리 형태로 사용자에게 보여준다.

완성된 질의 트리로부터 XQuery를 생성하는 과정은 단순하다. 루트 노드를 기준으로 깊이 우선탐색을 실시한다. 질의 트리 각 노드의 translate 메소드를 호출함으로써 질의 트리가 가지고 있는 정보를 이용하여 XQuery를 생성한다.

4.2 워크플로우 기반 질의 인터페이스

워크플로우 기반 질의 인터페이스는 통합 질의 시스템이 관리하는 다양한 생물 정보원의 리스트와 각종 편집도구를 제공한다. 사용자는 이를 이용하여 워크플로우를 구성한다.

4.2.1 시스템 개요

그림 5는 워크플로우 기반 질의 인터페이스의 개략적인 시스템 구조이다.

그림 5에서 입출력 제어기는 생물 정보원의 질의

```

<result>
{
for $x in document("nucleotide")/GBSet/GBSeq
where GBSeq_keyword = "cancer"
  <GBSeq_locus>{$x/GBSeq_locus/text()}</GBSeq_locus>
  <GBSeq_primary-accession>{$x/GBSeq_primary-accession/text()}
  </GBSeq_primary-accession>
  <GBSeq_taxonomy>{$x/GBSeq_taxonomy/text()}</GBSeq_taxonomy>
{
for $y in document("pubmed")/PubmedArticleSet/MedlineCitation
where $y/MedlineID = $x//GBReference_medline
  <Title> $y//Title/text() </Title>
  <ISSN> $y//ISSN/text() </ISSN>
}
}
</result>

```

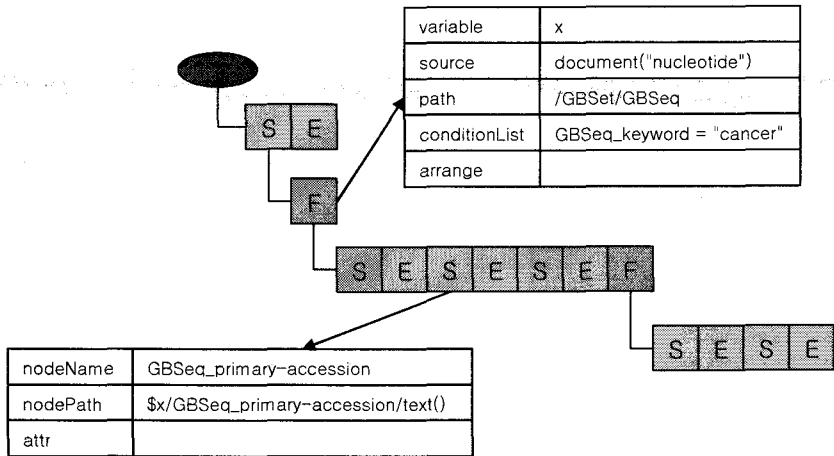


그림 4. 질의 트리

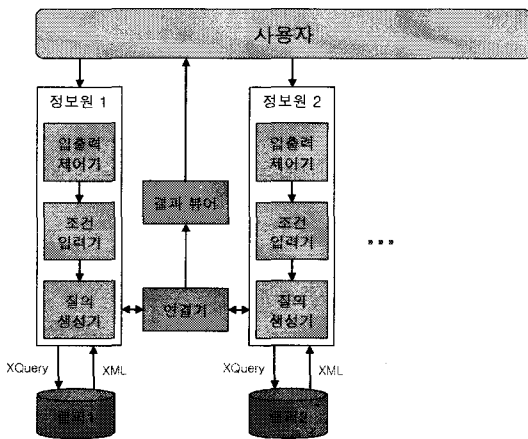


그림 5. 워크플로우 기반 질의 인터페이스 구조

드는 XQuery를 생성하는데 필요한 조건들에 대한 정보를 제공하고 출력 노드는 사용자에게 제공되는 결과를 나타낸다. 기본적으로 워크플로우 기반 질의 생성기는 입출력 제어기에 유지하고 있는 정보에 의해 정해진 형태의 XQuery를 생성한다. 조건 입력기는 랩퍼에 전달하기 위한 질의 생성에 필요한 사용자 입력을 받는 기능을 제공한다. 사용자가 입력한 정보는 이후 생성된 XQuery상의 조건절에 나타난다. 질의 생성기는 입출력 제어기의 기본 정보와 조건 입력기를 통해 입력 받은 사용자 정보를 이용하여 랩퍼에 전달할 XQuery를 생성하고 이를 랩퍼에 전달하고 수행된 결과를 받는다. 정보원에 대한 입력 조건은 사용자 입력 뿐 아니라 다른 정보원의 수행 결과의 일부로부터 전달 받을 수도 있다. 워크플로우를 생성하기 위해서는 한 정보원의 수행 결과의 일부를 다른 정보원의 입력 조건으로 전달해야 하는데 이때 연결

생성에 필요한 입력 노드와 출력 노드에 대한 정보를 유지하고 이를 시각화하는 기능을 담당한다. 입력 노

기를 통해 해당 정보가 전달되며 유지된다. 연결기는 SAX 파서를 이용하여 필요한 정보를 추출하고 이를 다른 정보원으로 전달한다. 또한 사용자는 연결기를 통해 수행된 결과를 결과뷰어에 전달하여 질의 결과를 확인할 수 있다.

4.2.2 XQuery 생성 방법

워크플로우 기반의 인터페이스에서 XQuery 생성은 크게 두 가지 관점으로 볼 수 있다. 첫 번째 관점은 하나의 정보원에 대한 질의어를 생성하는 것이고 두 번째 관점은 둘 이상의 정보원 사이에서 앞 정보원의 검색 결과의 일부가 다음 정보원의 검색 조건으로 추가되어 질의어를 생성하는 것이다. 그림 6은 워크플로어 기반 질의 생성기에서 XQuery를 생성하는 과정이다.

사용자는 선택한 생물 정보원의 조건 입력기를 통해 검색 조건과 검색어를 입력한다. 이들은 질의 생성기에서 XQuery 상의 where 절에 나타난다. 이때 일부 검색 조건과 검색어는 앞서 질의 과정을 수행한 정보원으로부터 연결될 수 있다. 연결기를 통해 정보원의 검색 조건에 대한 검색어를 이전 생물 정보원의 검색 결과로부터 추출하여 전달하는데 이때 SAX 파서를 이용하여 검색 결과를 분석하여 필요한 부분을 추출한다. 이때 이전 생물 정보원으로부터 전달되는 검색어가 하나 이상일 수 있기 때문에 이들은 where 절에서 or 연산자를 이용하여 연결한다.

5. 구현

5.1 구현환경

사용자 인터페이스의 구현 환경은 다음과 같다.

운영체제는 Windows XP이고, 구현 언어는 JDK 1.4.1을 사용하였다. XML 처리를 위한 DOM 파서와 SAX 파서는 아파치 그룹에서 제공하는 Xerces 파서를 이용하였다.

5.2 통합 스키마에 대한 질의 인터페이스

그림 7은 통합 스키마에 대한 질의 인터페이스의 실행화면이다.

왼쪽 창은 뷰 트리 창으로서 미디어이터가 관리하는 기본 뷰 트리를 보여준다. 뷰 트리 창 상단의 콤보 박스에는 미디어이터가 관리하는 전역 뷰의 이름이 있고, 사용자가 이 중 하나를 선택하면 이를 트리 형태로 보여준다. 가운데 창은 사용자 질의를 생성하기 위해 질의 트리창이다. 사용자는 뷰 트리 창으로부터 필요한 노드를 드래그 앤 드랍하거나 사용자 입력을 통해 새로운 노드를 추가하며 XQuery 생성을 위해 필요한 작업을 수행한다. 오른쪽 창은 결과 창으로 질의 트리 창에서 생성된 질의 트리로부터 생성한 XQuery를 보여주거나 생성한 질의를 수행한 후 결과를 보여준다. 아래쪽에 있는 창은 정보 출력 창으로 뷰 트리 창이나 질의 트리 창의 트리에 있는 노드를 선택할 때마다 해당 노드가 Element 노드인지 Flwor 노드인지를 조사하여 노드가 관리하고 있는 정보를 보여준다.

그림 8은 질의 트리 생성시 FLWOR 노드 생성을 위한 대화창이다. Flwor 노드를 생성할 때 변수명은 사용자가 입력한 변수명을 따른다. 경로 1은 사용자가 자신이 가져오고자 하는 정보원의 XML 뷰를 선택하는 것인데 만약 상위에 또 다른 Flwor 노드가 존재한다면 그 Flwor 노드에서 선언한 변수도 선택 가능하다. 찾기 버튼은 현재 콤보박스에서 선택한 뷰

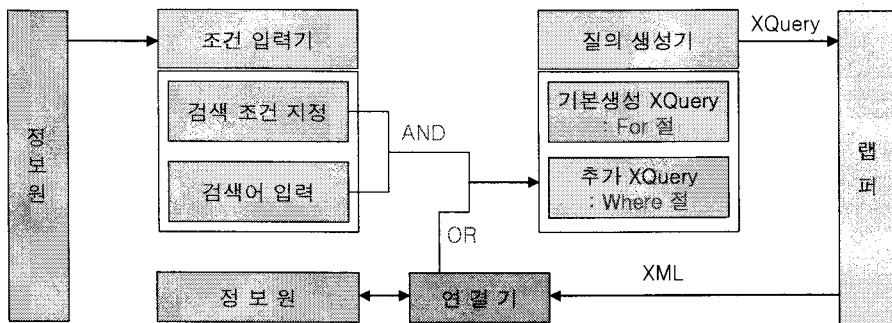


그림 6. 워크플로우 기반 질의 생성기에서 XQuery 생성 과정

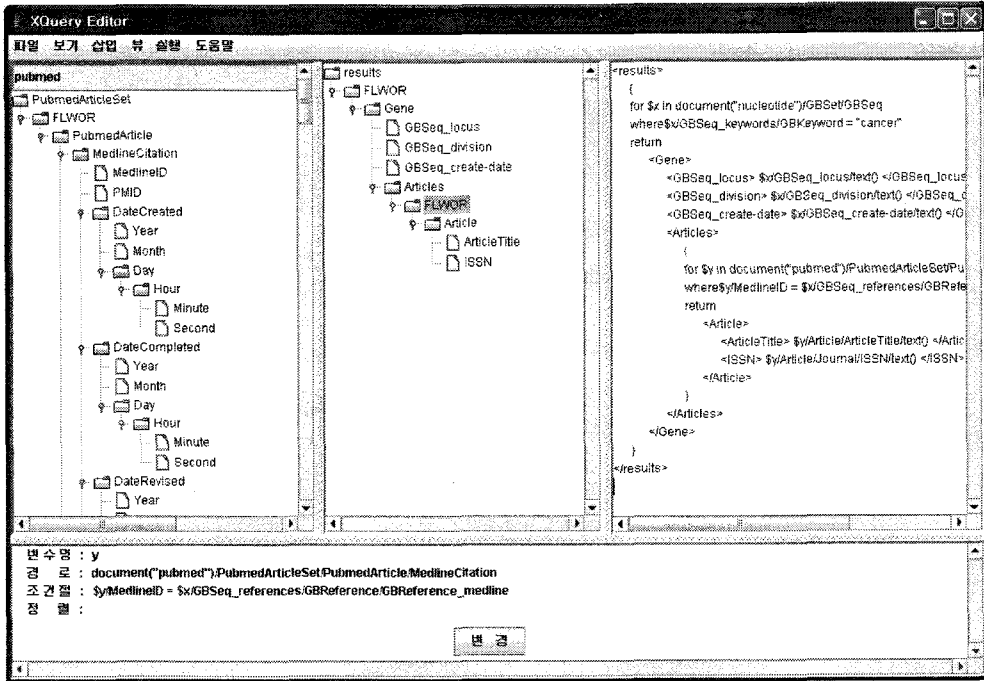


그림 7. 통합 스키마에 대한 질의 인터페이스

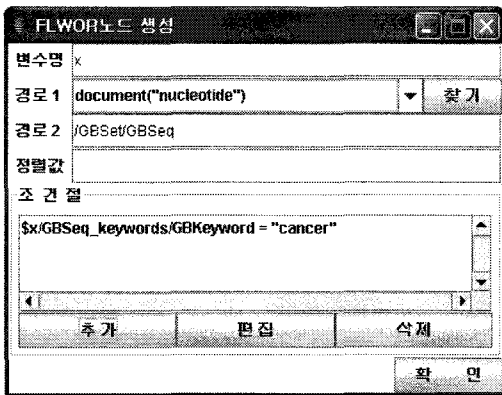


그림 8. Flwor 노드 생성

또는 변수에 대하여 사용자가 선택 가능한 경로를 트리형태로 보여준다. 찾기 버튼 클릭 후 선택한 경로는 경로 2에 표시된다. 사용자가 XML 뷰를 선택하면 XML 뷰 트리 전체를 나타내게 되지만 상위 Flwor 노드에서 선언한 변수를 선택하면 변수에 설정된 경로 이하만을 표시한다. 이는 사용자가 잘못된 경로를 입력하는 것을 사전에 예방하게 한다. 정렬 값은 결과를 받아왔을 때 정렬 상태를 입력하는 곳이

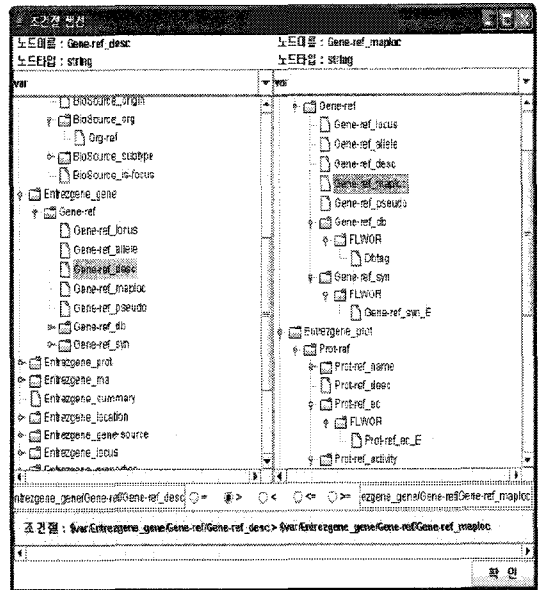


그림 9. 조건절 추가

다. 조건절은 리스트 형식으로 보여준다. 그림 9는 조건절을 추가할 때 생성되는 프레임이다. 가운데를 기준으로 왼쪽 창을 통해 조건절의 원

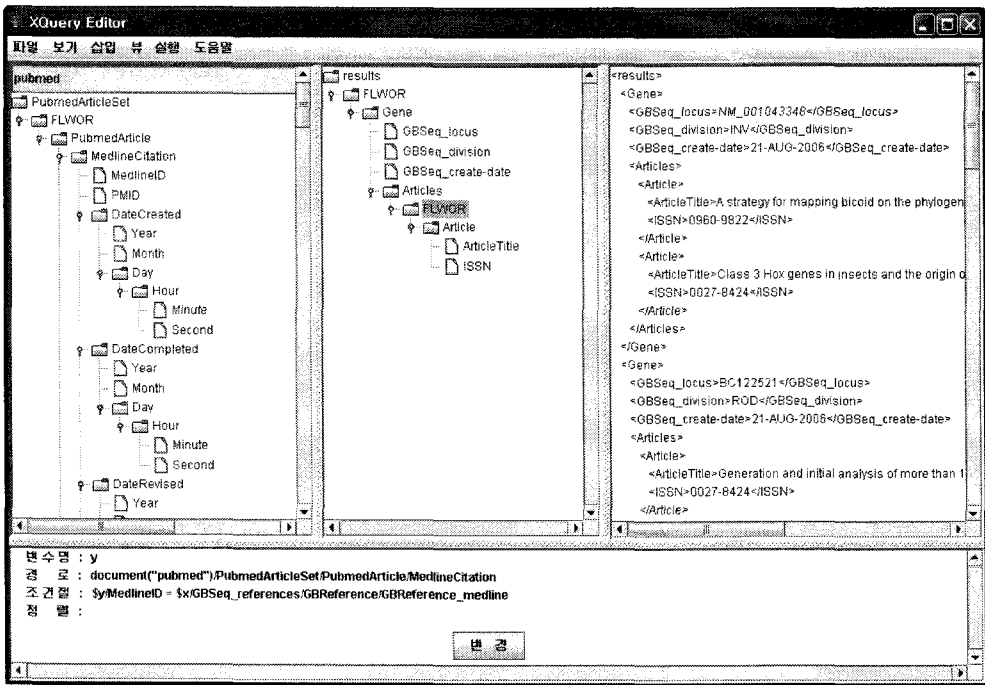


그림 10. 실행 결과

쪽에 들어갈 부분을 결정하고 라디오 버튼을 선택하여 비교 연산자를 선택한다. 그리고 오른쪽 창에서 조건절의 오른쪽에 들어갈 부분을 결정한다. 현재 설정한 변수와 상위에 생성된 FLWOR 노드의 변수 중 자신이 원하는 것을 선택하면 그것에 맞는 트리를 아래에 보여준다. 좌우 모두 트리를 선택해서 조건절을 생성할 수도 있고 아래쪽의 텍스트 필드에 직접 비교 값을 입력 하여 조건절을 생성할 수도 있다. 둘 이상의 조건절을 추가하면 화면 아래쪽에 비교 연산자인 and나 or을 선택하도록 창이 활성화 되어 두 조건절을 연결할 수 있다.

질의 트리 생성이 완료되면 메뉴 상의 실행 메뉴를 통해 질의 트리를 순회하여 통합 질의를 생성한다. 생성된 통합 질의는 미디어이터로 전달되고 미디어이터에서는 수행 결과를 통합하여 반환한다. 수행된 결과는 그래픽 XQuery 편집기의 결과창에 나타난다.

사용자 정의 XML 뷰를 생성하는 과정 역시 질의 트리 생성 과정과 다르지 않다. 다만 사용자 정의 XML 뷰는 질의가 아님으로 미디어이터로 전달된 후 질의 처리 과정을 수행하지 않고 다만 새로운 전역 뷰를 구성한다.

5.3 워크플로우 기반 질의 인터페이스

워크플로우 기반 질의 인터페이스는 사용할 수 있는 생물 정보원 리스트와 각종 도구들을 모아둔 도구 창과 질의 생성 작업을 수행하는 작업창으로 이루어진다. 도구 창에는 사용 가능한 생물 정보원에 대한 뷰와 워크플로우를 작성하는 과정에서 특정 결과만을 전달하기 위한 필터 툴, 질의 결과를 다양한 형태로 사용자에게 보여줄 수 있는 결과 뷰어 툴, 그리고 간단한 데이터마이닝을 수행할 수 있는 데이터마이닝 툴이 존재한다. 사용자는 도구창에서 필요한 아이템을 선택하고 작업창에 끌어다 놓음으로서 질의 작성을 수행하게 된다. 그림 11은 도구 창에서 BLASTN을 선택하여 작업창에 마우스로 드래그 했을 때의 모습이다.

그림 11에서 BLASTN에는 질의 생성에 필요한 다양한 입력 요소가 있으며 검색 결과에서 얻을 수 있는 출력 요소가 있음을 알 수 있다. 사용자가 선택한 생물 정보원에 따라 해당 정보원의 입력 요소나 출력 요소는 다양하다.

질의를 만드는 방법은 질의를 직접 정해서 입력하는 방법과 이전 정보원의 결과를 질의의 입력으로

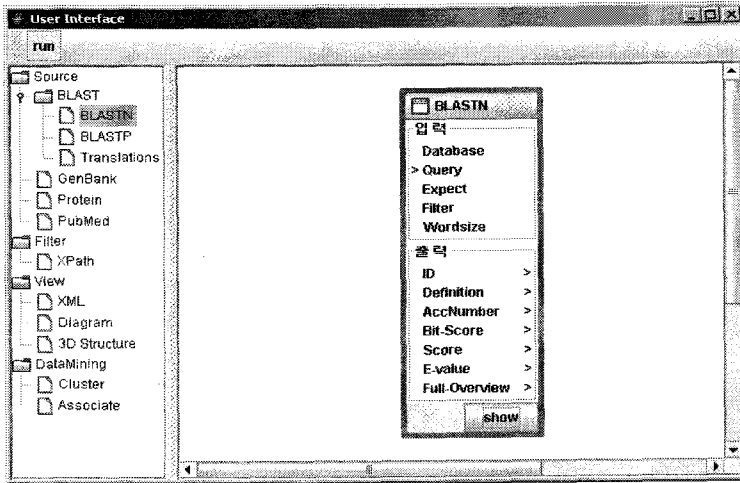


그림 11. BLASTN 정보원을 선택하여 드래그 한 결과

받는 방법 두 가지가 존재한다. 그림 12는 질의 생성에 필요한 조건을 입력할 수 있는 입력창의 모습이다. 질의를 직접 입력하기 위해서는 작업창에 나와 있는 정보원을 더블클릭함으로써 질의를 입력할 수 있는 입력창을 띄울 수 있다. 입력창에서 필요한 정보를 입력하고 'OK' 버튼을 누르면 질의가 작성된다. BLASTN의 경우 단백질 서열을 입력으로 주고자 할 경우 '열기' 버튼을 눌러 단백질 서열의 내용을 입력

한다.

'OK' 버튼을 클릭함으로써 사용자 질의는 완성된다. 생성한 질의는 해당 생물 정보원의 랩퍼에 전달되어 결과를 확인할 수 있다. 그러나 이 생물 정보원에서 얻은 결과를 또 다른 생물 정보원의 입력으로 이용하여 복합적인 결과를 얻고 싶은 때는 또 다른 생물 정보원을 작업창으로 드래그한다. 예를 들어 또 다른 생물 정보원인 GenBank를 작업창으로 드래그

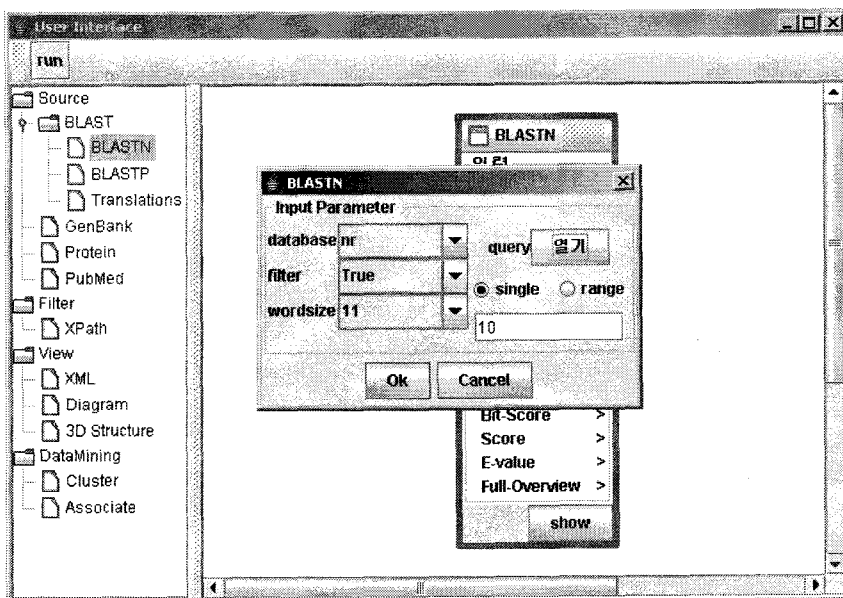


그림 12. BLASTN의 질의 조건 입력

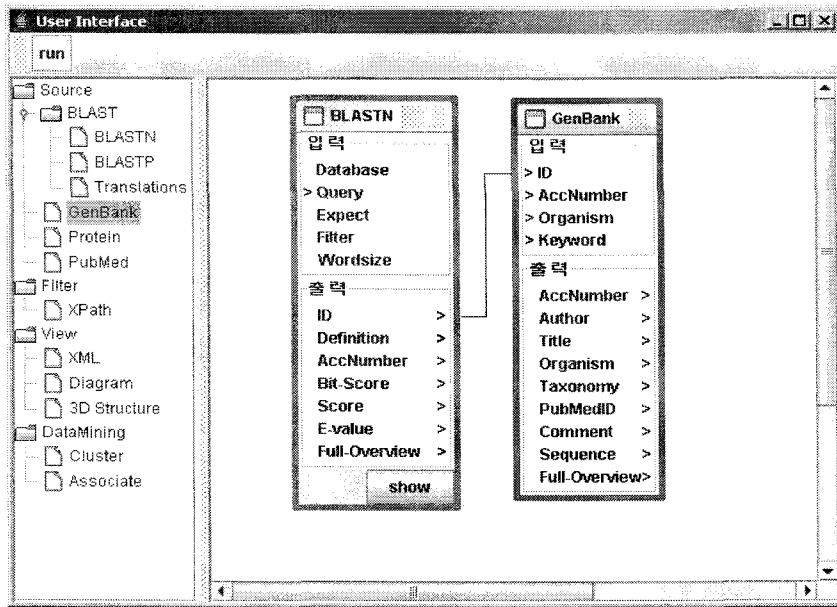


그림 13. BLASTN 수행결과 중 ID를 GenBank의 입력으로 설정

한다. 만약 BLASTN 수행결과 중 ID를 GenBank 질의를 만들기를 원한다고 하자.

이전 정보원의 [출력] 부분에서 새로운 정보원의 [입력]으로 들어갈 필드와 일치하는 것을 찾아서 연결해 주면, 이전 정보원의 결과가 새로운 정보원의 질의 입력으로 설정된다. 연결 방법은 정보원의 [출력]의 '>'를 클릭하고 새로운 정보원의 [입력]의 '>'를 클릭하면 그림 13과 같이 서로 선으로 연결된다. 이로서 하나의 워크플로우가 완성된다. 그 결과 BLASTN의 검색 결과에서 ID 부분을 추출하여 그것을 GenBank의 ID로 활용하여 계속해서 질의를 자동으로 만들어서 검색을 하게 된다. 마찬가지로 새로운 정보원의 출력을 또 다른 정보원의 입력으로도 설정할 수 있다.

한편, 이렇게 하나의 정보원으로부터 입력이 정해지는 경우 이외에, 복수개의 다른 정보원의 결과를 입력으로 받는 것도 가능하다. 이 경우 역시 마찬가지로의 방식으로 처리할 수 있다.

자신이 수행하고자 하는 워크플로우가 완성되면 프로그램의 상단 툴바에 있는 'Run' 버튼을 클릭하면 작성한 워크플로우에 따라 각 생물 정보원에 대한 질의가 차례대로 수행되고 최종 결과를 확인할 수 있다. 질의 결과를 받아 볼 수 있는 툴은 현재 XML만 구현이 되어 있다. XML을 작업창으로 끌어 두고 최

종 정보원의 '출력'부에 'Full-Overview'를 연결한다. 작성한 워크플로우에 따른 최종 결과를 그림 14에서 제시한 것처럼 XML 형태로 확인이 가능하다.

6. 결론 및 향후과제

이질의 정보원을 통합하는 방법론으로서 XML 기반의 미디어이터 시스템은 모든 정보원의 스키마를 XML 스키마로 변환하여 XML 질의어으로써 질의한다. 따라서 XML 질의어를 효과적으로 생성할 수 있는 사용자 인터페이스가 필요하다. 본 논문에서는 범용 통합 시스템 기반의 질의 인터페이스를 먼저 제시하고, 특히 생물학자가 생물 정보원을 사용하는 작업의 흐름에 따라서 질의를 생성할 수 있도록 도와주는 워크플로우 기반의 사용자 인터페이스를 제시하였다.

통합 스키마에 대한 질의 인터페이스는 질의 뿐만 아니라 사용자 정의 XML 뷰를 쉽게 생성할 수 있는 기능을 제공하는데, XQuery가 제공하는 다양한 함수를 지원하는 기능을 추가하여 그 기능을 더 향상시킬 필요가 있다. 그리고 워크플로우 기반의 인터페이스는 직관적인 방법으로 사용자가 워크플로우를 작성하여 질의를 수행할 수 있는 기능을 제공하는데, 새로운 정보원 추가하고자할 때 정보원의 스키마를

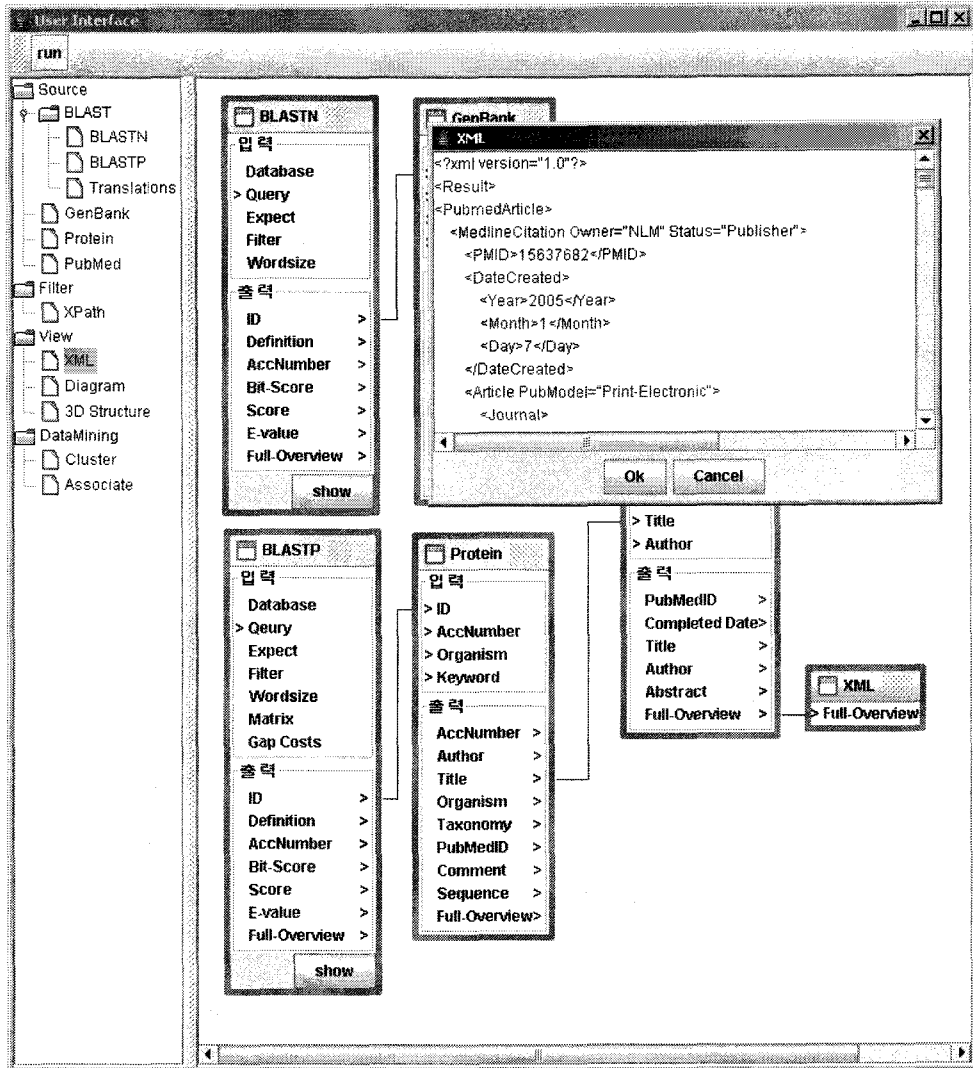


그림 14. 질의 결과 확인

이용하여 자동적으로 클래스를 생성하여 이를 추가 할 수 있는 방법에 대한 연구가 필요하다. 현재는 이 두 개의 인터페이스가 별도로 존재하나 향후 이를 하나의 인터페이스로 통합할 예정이다.

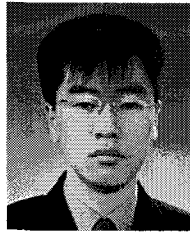
참 고 문 헌

[1] T. Hernandez and S. Kambhampati, "Integration of Biological Sources :Current Systems and Challenges Ahead," *ACM SIGMOD Record* Vol. 33, No. 3, pp. 51-60,

2004.
 [2] B. Eckman, Z. Lacroix, and L. Raschid, "Optimized Seamless Integration of Biomolecular Data," *IEEE International Symposium on Bio-Informatics and Biomedical Engineering (BIBE)*, pp. 1-3, 2001.
 [3] P.G. Baker, A. Brass, S. Bechhofer, C. Goble, N. Paton, and R. Stevens, "TAMBIS - Transparent Access to Multiple Bioinformatics Information Sources," *Intelligent Systems for Molecular Biology 1998 (ISMB98)*, Vol. 6, pp.

25-34, 1998.

- [4] C.A. Goble, R. Stevens, G. Ng, S. Bechhofer, N.W. Paton, P.G. Baker, M. Peim, and A. Brass, "Transparent Access to Multiple Bioinformatics Information Sources," *IBM Systems Journal Special issue on deep computing for the life sciences*, Vol. 40, No. 2, pp. 532-552, 2001.
- [5] C. Baru, A. Gupta, B. Ludascher, R. MarcianoYannis, Y. Papakonstantinou, P. Velikhov, and V. Chu, "XML-Based Information Mediation with MIX," *ACM-SIGMOD 99 Conference*, pp. 597-599, 1999.
- [6] K.D. Munroe, B. Ludäscher, and Y. Papakonstantinou, "Blended Browsing and Querying of XML in a Lazy Mediator System," *Extending Database Technology (EDBT)*, 2000.
- [7] K. Munroe and Y. Papakonstantinou, "BBQ: A Visual Interface for Browsing and Querying XML," *Visual Database Systems (VDB)*, 2000.
- [8] I. Altintas, S. Bhagwanani, D. Buttler, S. Chandra, Z. Cheng, M. Coleman, T. Critchlow, A. Gupta, W. Han, L. Liu, B. Ludaescher, C. Pu, R. Moore, A. Shoshani, and M. Vouk, "A Modeling and Execution Environment for Distributed Scientific Workflows," *Scientific and Statistical Database Management (SSDBM 2003)*, pp. 247-250, 2003.
- [9] B. Ludaescher, I. Altintas, and A. Gupta, "Compiling Abstract Scientific Workflows into Web Service Workflows," *Scientific and Statistical Database Management (SSDBM 2003)*, 2003.
- [10] L. Liu, C. Pu, and W. Han, "XWRAP: An XML-enabled Wrapper Construction System for Web Information Sources," *International Conference on Data Engineering (ICDE)*, 2000.



정 채 영

1997년 경상대학교 전자계산학과 졸업(이학사)
 2001년 경상대학교 컴퓨터과학과 졸업(공학석사)
 2004년 경상대학교 컴퓨터과학과 졸업(공학박사)
 2004년~현재 경상대학교 컴퓨터

과학과 시간강사

관심분야 : XML, 웹 프로그래밍, 정보 통합, 데이터베이스



조 찬 제

2006년 경상대학교 컴퓨터과학과 졸업(이학사)
 2006년~현재 썬크프리 연구원
 관심분야 : 데이터베이스, XML



박 흥 원

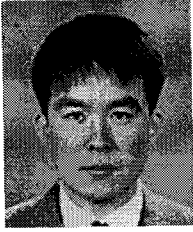
2006년 경상대학교 컴퓨터과학과 졸업(이학사)
 2006년~현재 현대정보기술 연구원
 관심분야 : 데이터베이스, XML



박 은 경

1999년 경상대학교 컴퓨터과학과 졸업(이학사)
 2002년 경상대학교 컴퓨터과학과 졸업(공학석사)
 2007년 경상대학교 컴퓨터과학과 졸업예정(공학박사)
 관심분야 : 데이터베이스 통합, 웹

정보 기술, XML



김 현 주

1988년 경상대학교 전자통계학과 졸업(이학사)
1990년 숭실대학교 전자계산학과 졸업(공학석사)
2000년 경상대학교 전자계산학과 졸업(공학박사)
1994년~1997년 제일정밀공업

(주) 연구원

2000년~2002년 경남정보대학 컴퓨터정보계열 전임강사
2002년~현재 진주산업대학교 컴퓨터공학과 전임강사, 조교수
관심분야 : 정보 검색, 디지털 도서관, 웹 프로그래밍, XML



배 종 민

1980년 서울대학교 수학교육과 졸업(학사)
1983년 서울대학교 대학원 계산통계학과 졸업(석사)
1995년 서울대학교 대학원 계산통계학과 졸업(박사)
1982년~1984년 한국전자통신연

구소 연구원

1997년~1998년 Virginia Tech. 객원연구원
1984년~현재 경상대학교 전임강사, 조교수, 부교수, 교수
관심분야 : XML, 데이터베이스 통합, 웹 정보 기술