

## 바이오 객체 상호작용 네트워크 구축 시스템 설계 및 구현

박종민<sup>o</sup> 최재훈 정재영 박선희  
한국전자통신연구원  
{ jmpark93<sup>o</sup>, jhchoi, jyy72, shp }@etri.re.kr

### Design and Implementation of System for Constructing Bio-Object Interaction Network

JongMin Park<sup>o</sup> JaeHun Choi JaeYoung Jung SeonHee Park  
Electronic Telecommunication Research Institute(ETRI)

#### 요 약

본 논문에서는 생물체의 세포 내에 존재하는 방대한 객체들 사이의 복잡한 관계들로 표현되는 상호작용 네트워크를 효율적으로 구축할 수 있는 시스템을 제안한다. 이 시스템은 바이오 도메인 지식을 사용하여 상호작용 네트워크를 관리 및 활용하기 쉽도록 구축하고, 단백질과 같은 단순 바이오 객체뿐만 아니라 여러 개의 바이오 객체들로 구성된 복합 객체도 관리할 수 있다. 여기서, 사용자가 바이오 객체들과 이들 간의 복잡한 상호작용 관계를 직관적으로 정의할 수 있는 인터페이스를 제공한다. 또한, 정의된 객체 및 상호작용 관계 정보를 이용하여 바이오 네트워크를 개념적으로 단순하게 표현할 수 있으며, 시각적으로도 네트워크를 자동으로 최적화하여 사용자가 복잡한 네트워크를 쉽게 분석할 수 있도록 지원한다.

#### 1. 서 론

생물체의 세포 내에서 바이오 객체들은 다양한 생물학적 역할을 수행하기 위해 여러 다른 바이오 객체들과 상호작용을 한다. 따라서, 하나의 세포 내에는 방대한 바이오 객체들간의 복잡한 상호작용 관계들이 존재하며 네트워크로 표현할 수 있다[1][2]. 여기서, 바이오 객체는 DNA, RNA, 단백질 등을 나타내며, 상호작용 관계는 활성, 억제 등과 같은 바이오 객체들 사이의 관계를 정의한다. 그리고, '해모글로빈'이나 'DNA Polymerase'와 같이 특정 바이오 객체들은 고유한 기능을 수행하기 위하여 여러 개의 바이오 객체들이 하나의 복합체(complex)로 구성되어 있다.

현재, 생물정보학(bioinformatics) 분야에서는 생화학 반응, 유전자 발현, 그리고 신진대사 조절 과정 등을 분석하고 시뮬레이션 하기 위해 상호작용 네트워크가 중요한 문제로 다루어지고 있다. 그리고, 상호작용 관계를 식별하기 위한 고속 스크린 처리(high-throughput screen) 기법들이 개발됨에 따라, protein-protein, protein-DNA, genetic interaction 등과 같은 상호작용 데이터들이 빠르게 축적되고 있다. 하지만, 상호작용 네트워크는 방대한 바이오 객체들 사이의 복잡한 관계들로 표현되어있어 사용자들이 직접 관리하고 분석하기 어렵다. 따라서, 다양한 형태의 바이오 객체와 상호작용 관계를 관리할 수 있고, 방대한 객체들 사이의 복잡한 관계들로 표현된 상호작용 네트워크를 쉽게 구축하고 분석할 수 있도록 지원

하는 체계적인 시스템이 필요하다.

#### 2. 관련 연구

바이오 객체 상호작용 네트워크를 지원하는 시스템으로는 Biolayout, PIMRider, PIVOT, Cytoscape 등이 있다. 하지만, 대부분의 시스템들은 기존의 구축된 상호작용 네트워크를 분석하기 위한 용도로 사용된다. 상호작용 네트워크를 직접 구축하고 관리할 수 있는 시스템으로는 Interviewer, Osprey 가 있다.

먼저, Interviewer[3]는 복합 객체를 관리할 수 있으며 네트워크를 단순화 할 수 있는 방법을 지원하는 등 다양한 기능을 가지고 있다. 하지만, 바이오 객체 및 상호작용 관계에 대한 바이오 도메인 지식을 개념적으로 사용하지 않고 물리적인 관점에서만 처리한다. Osprey[4]는 바이오 도메인 지식으로 'GO(gene ontology)' 및 종(species)별 바이오 객체 정보를 가진 데이터베이스와 연동하여 사용한다. 하지만, 복합 객체를 관리할 수 없으며, 복잡한 네트워크를 단순화하여 표현할 수 있는 방법을 지원하지 않고 있다. 따라서, 본 논문에서는 바이오 도메인 지식을 통해 바이오 네트워크를 체계적으로 관리할 수 있으며, 단백질과 같은 단순 바이오 객체뿐만 아니라 여러 개의 바이오 객체들로 구성된 복합 객체도 표현할 수 있는 바이오 객체 상호작용 네트워크 구축 시스템을 제안한다.

### 3. 바이오 객체 상호작용 네트워크 모델

본 논문에서 정의한 바이오 객체 상호작용 네트워크 모델은 네트워크 데이터의 교환 및 통합, 그리고 비교 및 검증 등 다양한 용도에 활용될 수 있고 향후 확장이 쉽도록 유연한 구조로 정의하였다. 상호작용 네트워크 모델은 작성된 네트워크에 대한 상세한 정보를 담고 있는 메타데이터, 네트워크에 참여하는 바이오 객체, 그리고 바이오 객체 사이의 상호작용 관계로 구성된다. 그림 1은 본 논문에서 정의한 바이오 객체 상호작용 네트워크 모델을 도식화한 그림이다.

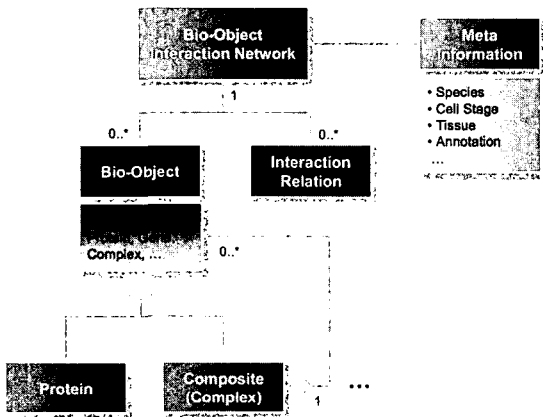


그림 1. 바이오 객체 상호작용 네트워크 모델

#### 3.1 상호작용 네트워크 메타데이터

상호작용 관계를 식별하기 위한 기술이 발달함에 따라 상호작용 네트워크 데이터들이 빠르게 축적되고 있으며, 구축된 상호작용 네트워크는 종(species)에 따라 참여하는 바이오 객체 및 상호작용 관계가 다를 뿐 아니라, 같은 종이라도 세포 분화 단계(cell stage) 및 실험에 사용된 조직(tissue)에 따라 다르다. 즉, 상호작용 관계를 식별하기 위한 실험에 사용된 다양한 조건들에 따라서 서로 상이한 상호작용 네트워크가 만들어진다. 따라서, 메타데이터에는 네트워크에 대한 기본적인 내용뿐 아니라 방대한 네트워크 데이터들로부터 사용자가 원하는 네트워크를 추출하기 위한 데이터들을 정의하고 있다. 즉, 종, 실험 방법, 세포 분화 단계, 조직, 그리고 GO에 정의된 용어로 작성된 생물학적 기능에 대한 설명(annotation) 들로 구성되어 있다.

#### 3.2 바이오 객체

바이오 객체는 세포 내에서 독립적으로 존재할 수 있는 객체들을 말하며 크게 단순(primitive) 객체와 복합(composite) 객체로 나누어 정의하였다. 먼저, 단순 객체는 DNA, RNA, 또는 단백질 등과

같은 객체를 말한다. 복합 객체는 '헤모글로빈', 'DNA Polymerase'와 같이 세포 내에서 여러 단백질이 특정한 기능을 수행하기 위하여 하나의 묶음으로 작용하는 단백질 집합체인 복합체(complex)나 특별한 이유에 따라 몇 개의 바이오 객체를 하나의 단위로 만든 객체를 말한다. 각 바이오 객체는 GO에 정의된 용어로 작성된 생물학적 기능에 대한 설명(annotation) 가지며, 추가적인 상세한 생물학적인 정보를 접근하기 위한 외부 데이터베이스의 식별자를 가진다.

#### 3.3 상호작용 관계

바이오 객체 사이에는 억제, 활성화 등과 같은 다양한 상호작용 관계들이 존재한다. 상호작용 관계에는 상호작용에 참여하는 바이오 객체에 대한 식별자를 가지고 있다. 또한, 상호작용 관계는 실험 및 예측 방법 등에 따라 상호작용이 실제로 존재할 정확도가 다를 수 있으므로 정확도의 정도 값을 0에서 1사이의 값으로 표현하여 분석에 활용될 수 있도록 한다.

### 4. 바이오 객체 상호작용 네트워크 구축 시스템

바이오 객체 상호작용 네트워크 구축 시스템은 상호작용 네트워크 모델을 기반으로 네트워크를 관리하는 모듈, 바이오 객체와 상호작용 관계를 최적으로 배열하여 시각화하는 모듈, 그리고 'GO(gene ontology)'와 'Clique'를 이용하여 개념적으로 네트워크를 단순화하는 모듈로 구성되어 있다. 또한, 이러한 기능을 지원하기 위해 지식베이스로 'GO', 바이오 객체에 대한 데이터로 'Swiss-Prot' 데이터베이스가 사용된다. 그림 2는 바이오 객체 상호작용 네트워크 구축 시스템을 도식화한 그림이다.

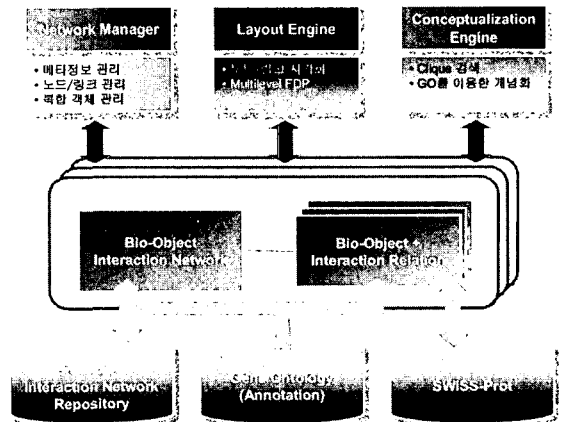


그림 2. 전체 시스템 구성 다이어그램

#### 4.1 네트워크 관리 모듈

바이오 객체 상호작용 네트워크의 구성요소들을 관리하는 모듈이다. 즉, 네트워크 메타데이터, 바이오 객체, 상호작용 관계를 생성, 삭제, 수정하는 일련의 작업들을 수행한다. 또한, 선택된 여러 개의 바이오 객체를 복합 객체로 생성하거나, 복합 객체를 확장하여 분석할 수 있도록 지원한다. 그리고, 해당 네트워크에 참여하는 대량의 바이오 객체의 생물학적 기능에 대한 설명(annotation)들을 취합하여 네트워크에 대한 생물학적 기능을 자동으로 메타데이터에 정의할 수 있다.

#### 4.2 자동 시각화 모듈

바이오 객체와 상호작용 관계를 시각화하는 부분을 담당한다. 여기서, 네트워크를 여러 단계로 레벨화하여 처리하는 개선된 MFDP(Multilevel algorithm for FDP)[5]를 이용하여 자동으로 시각화한다. 즉, 상호작용 관계 네트워크를 여러 단계로 레벨로 압축한 다음, 점차적으로 다시 확장하면서 바이오 객체들에 대한 위치 좌표를 구한다. 이때, 특정 레벨의 위치 데이터는 이전 단계의 위치 데이터를 사용하여 성능을 향상시킨다.

#### 4.3 네트워크 개편화 모듈

일반적으로, 'Clique'은 복합체(complex)가 될 수 있는 가능성이 높은 객체로 가정한다. 하지만, 물리적인 바이오 객체간의 상호작용 관계만을 고려하여 식별된 'Clique'의 각 구성요소들은 서로 전혀 다른 생물학적 기능을 수행할 가능성이 있다. 따라서, GO를 이용하여 생물학적 기능이 서로 유사한 바이오 객체만을 복합체로 예측할 수 있다. 따라서, 동일한 생물학적 기능을 수행하는 객체들을 하나의 복합 객체로 표현하여 좀 더 개념적으로 바이오 객체 상호작용 네트워크를 단순화 시킨다.

그림 3은 바이오 객체 상호작용 네트워크 구축 시스템의 인터페이스 화면을 나타낸다.

#### 5. 결론 및 향후 연구 과제

본 논문에서는 생물체의 세포 내에 존재하는 방대한 객체들 사이의 복잡한 관계들로 표현되는 상호작용 네트워크를 체계적으로 관리할 수 있는 시스템을 제안하였다. 이 시스템은 바이오 도메인 지식을 사용하여 네트워크의 구성요소인 네트워크 메타데이터, 바이오 객체, 상호작용 관계를 관리하며, 복합 객체를 생성 및 확장하여 분석할 수 있다. 또한, 네트워크를 여러 단계로 레벨화하여 처리하는 개선된 MFDP를 이용하여 자동으로 시각화하며, GO를 이용하여 검색된 'Clique'의 각 구성요소들의 생물학적 기능이 서로 유사한 바이오 객체만을 복합체로 예측하고 단순화하여 상호작용 네트워크를 쉽게 분석할 수 있도록 지원한다.

현재, 바이오 객체 상호작용 네트워크 데이터의 교환 및 통합, 그리고 비교 및 검증 등을 위해 XML 기반의 네트워크 모델의 표준화가 시도되고 있다. 하지만, 바이오 도메인 지식과의 연관성 및 복합체를 표현하기 위한 표준이 부족하다. 따라서, 본 논문에서 제시한 네트워크 데이터 모델을 표준화된 모델에 활용하거나 표준화된 모델을 본 논문에서 제안한 시스템에서 사용하기 위해 XML 기반으로 작성할 필요가 있다. 또한, 상호작용 네트워크 교환 및 통합, 그리고 비교 및 검증 등을 시스템에서 지원하기 위해서는 네트워크 사이의 다양한 연산을 수행할 수 있어야 한다.

#### 참고문헌

- [1] C. L. Tucker, J. F. Gera, and P. Uetz, "Towards an Understanding of Complex Protein Interaction Maps," Trends in Cell Biology, Vol. 11, No. 23, 2001.
- [2] P. Uetz, T. Ideker and B. Schwikowski, "Visualization and Integration of Protein-Protein Interactions," Cold Spring Harbor Laboratory Press, 2002.
- [3] InterViewer, <http://interviewer.inha.ac.kr/>
- [4] Osprey, <http://biodata.mshri.on.ca/osprey>
- [5] C. Walshaw, "A Multilevel Algorithm for Force-Directed Graph Drawing," Graph Drawing 8th Intl. Symp, Berlin, 2001.

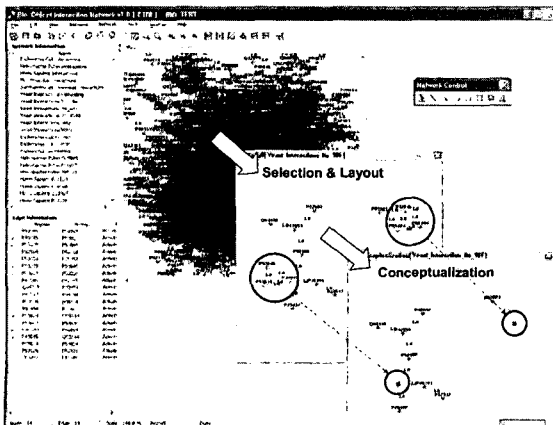


그림 3. 바이오 객체 상호작용 네트워크 모델