

바이오 네트워크 구축을 위한 바이오인포매틱스 기반 기술

Bioinformatics for constructing bio-pathways

박선희

한국전자통신연구원 바이오정보연구팀

shp@etri.re.kr

바이오인포매틱스 기술이란 정보통신 인프라와 IT기술을 기반으로 분산되어 있는 대용량의 바이오 데이터 및 문헌 정보를 통합적으로 관리, 분석 및 처리하여 각종 바이오 정보 활용 분야(신약개발, 의료진단, 농산물 개량 등)에 효율적으로 응용 서비스하기 위한 IT-BT융합 핵심 기반 요소 기술을 의미한다.

2000. 6. 인간 유전체 초안 완성 후 대량 획득되는 바이오 데이터의 최대 부가가치 창출을 위해 대용량 바이오 데이터로부터 유용한 지식정보를 자동적으로 추출해내기 위한 효율적인 통합분석 기술이 필요하고 이는 바이오인포매틱스 기본 개념이 된다. 특히 이제까지 생성되고 구축된 바이오 데이터를 가공하여 생체 기능을 예측할 수 있는 컴퓨터 내 연구 방법이 향후 바이오 분야의 연구나 산업에 핵심 기술로 자리매김하고 있다.

컴퓨터 내에서 예측할 수 있는 생체 기능이란 세포 막까지 전달된 신호가 단백질 상호작용 등을 통하여 다시 세포의 핵까지 전달되어 핵 안의 유전자를 발현시키고 이에따른 단백질이 합성되어 작용을 하는 총체적인 네트워크 예측을 의미한다. 이와같은 컴퓨터 내 연구는 엄청난 시간과 비용이 드는 종전의 실험적 방법을 보완해 줄 수 있는 매우 유효한 방법으로서 선진국에서는 정부 차원의 지원은 물론 제약 회사 등에서도 in house로서 연구되고 외부에 공개하지 않은 채 자신들만의 기술을 보유하고 있다. 우리나라에서는 바이오인포매틱스에 기반한 바이오 네트워크 예측 기술 연구는 아직 시작되고 있지 않고 바이오인포매틱스 벤처업체에서 산발적이고 소규모적으로 연구가 진행되고 있는 실정이다.

생체 기능 예측을 위한 바이오인포매틱스는 매우 광범위한 분야를 포함하고 있다. 여기에는 (1) 다양한 형태의 바이오 데이터를 처리하여 분석이 용이한 형태로 변환 및 바이오 데이터의 시각화를 위한 바이오 데이터 처리 기술, (2) 바이오 데이터를 분석하여 내재된 패턴을 찾아내어 유용한 지식을 생성하기 위한 바이오 데이터 마이닝 기술, (3) 분산되어 있는 다양한 형태의 대용량 바이오 데이터를 저장, 관리하고 통합 검색하기 위한 바이오 데이터 통합 관리 기술, (4) 통합 관리된 바이오 데이터를 각종 응용분야에 활용하기 위한 바이오인포매틱스 응용서비스 지원 기술로 나눌 수 있다.

바이오 데이터 처리 기술에는 바이오 영상(Micro-array, 2D PAGE)을 고정밀 영상 정보 처리 기술을 이용하여 객체 정보를 추출하고 패턴 분석을 통해 객체의 상동성, 차별성 및 식별(identification)하는 기술, 대용량의 바이오 데이터를 컴퓨터그래픽스, 가상현실, 과학적 가시화 기법을 적용하여 시각 및 촉각 기반으로 바이오 데이터를 처리하는 기술이 있다.

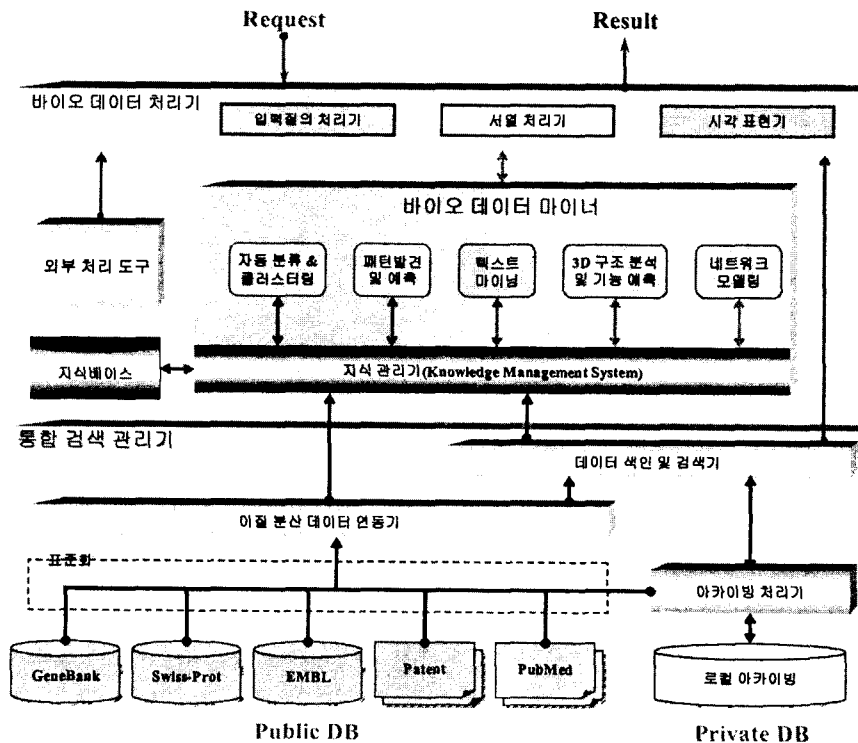
바이오 데이터 마이닝 기술에는 서열, 칩, 문헌 등의 원시 데이터를 가공하여 유용한 정보로 만드는 기술로서 Rule-based 필터링 기술에 의한 바이오 분자 모델을 생성하고 생성한 분자 모델간의 결합 부위 탐색 기술을 이용하여 바이오 데이터의 기능을 예측하는 기술, 바이오 문헌을 대상으로 자연어 처리

및 정보 추출 기술을 적용하여 생물학적 구성 요소들 간의 관계를 추출함으로써, 이들의 기능과 상호관계를 유추해내고 이를 기반으로 새로운 생물학적 지식을 발견하기 위한 텍스트 마이닝 기술, 바이오 데이터를 분류적 또는 객체별 특성에 기반하여 그룹화하고 이를 기반으로 유전자, 단백질 등의 기능 예측이나 마이크로어레이 데이터 분석 등에 활용할 수 있는 바이오 데이터 자동 분류 및 클러스터링 기술, 유전자 서열, 단백질 서열 등의 바이오 데이터에서 바이오 데이터의 특징에 적합한 패턴 발견 기술을 개발하여 유전자, 단백질의 기능을 예측하는 기술, 기존에 밝혀진 바이오 객체들에 대한 유사한 상호작용 네트워크들을 검색하여 기본 템플릿을 생성하고 문헌이나 서열 패턴 정보를 분석함으로써 새로운 객체의 상호작용 네트워크를 구축하는 기술, 대용량의 바이오 데이터를 가공하여 생성 유추된 지식 정보를 효율적으로 저장 및 관리하기 위한 바이오 지식 관리 기술이 있다.

바이오 데이터 통합 관리 기술에는 바이오 온톨로지의 일관성을 유지하고 대용량으로의 확장을 가능하게 하기 위한 바이오 온톨로지 구축 관리 및 표준화 기술, 여러 곳에 다양한 형태로 저장되어 있는 바이오 데이터를 사용자들이 하나의 데이터베이스처럼 사용할 수 있도록 통합된 자료 모델과 검색 인터페이스를 제공하기 위한 분산 이질 바이오 데이터 통합 기술, 페타바이트급의 대규모 바이오 데이터를 바이오 데이터의 특성을 고려하여 효율적으로 저장, 관리하기 위한 대용량 바이오 데이터 아카이빙 기술이 있다.

카이오인포매틱스 응용 서비스 지원 기술에는 의료진단 지원 기술, 신약개발 지원 기술, 실험실 정보 관리 기술(LIMS) 등 광범위한 분야가 포함되어 있다.

이러한 기술의 개념도는 아래와 같다.



바이오인포매틱스는 정보통신 기술과 바이오 기술이 융합되어 서로 시너지효과를 나타내며 발전하고 있다. 생물학, 의학, 전산학, 통계학, 물리학, 광학, 반도체 기술 등 모든 과학기술의 융합체라고 할 수 있는 카이오인포매틱스의 발전으로 향후 새로운 학문 분야나 산업 분야가 탄생될 것으로 기대된다.