

# 암 데이터 분석을 위한 통합의료정보시스템의 데이터베이스 설계 및 구축

신동문\*, 허룡\*, 심재민\*\*, 손호진\*, 류근호\*

\*충북대학교 데이터베이스/바이오인포매틱스 연구실

e-mail:{mastershin216, heoly, shon0621, khryu}@dblabb.chungbuk.ac.kr

## Database Design and Implementation of an Integrated Medical Information System for Cancer Data Analysis

Dong Mun Shin, Lyong Heo, Jae-Min Shim, Ho Sun Shon, Keun Ho Ryu

Database/Bioinformatics Laboratory, Chungbuk National University

### 요 약

본 논문에서는 개인 특화된 의료를 위한 진단, 치료선택, 예후 추정을 지원하기 위한 정보를 전문 의료인에게 효과적으로 제공하기 위한 데이터베이스 설계와 구축을 제시한다. 내원 환자들의 유전자 수준의 미시 데이터, 임상학적 거시 데이터, 가족력, 유사 질환군 등의 연관정보 데이터를 통합·연계하여 이력으로 관리하고, 데이터의 점진적 누적이 가능한 통합의료시스템을 위한 데이터베이스 설계의 프레임워크를 구축하였다.

### 1. 서론

지난해 8월 발표된 우리나라 사망 원인 통계에 따르면 악성신생물(암)의 사망률은 인구 10만 명당 140명으로, 뇌혈관질환과 심장질환을 합한 값(10만 명당 100명)의 1.4배다. 따라서 국민건강을 증진하기 위해 암 예방 및 치료는 필수적인 과제다 [1][2]. 암 질환의 현황을 파악하고 적절한 관리대책을 마련하기 위해서는 암의 사망, 발생 수준 및 생존율에 대한 정확한 통계가 필요하다[3]. 따라서 환자의 진단, 치료선택, 예후추정 등을 위해 임상학적 거시 데이터뿐만 아니라 기술의 진보에 따라 가용해진 유전자 수준의 미시 데이터를 활용하기 위한 관심과 기대가 커지고 있으나, 이들 데이터들이 개별적으로 수집·관리·분석되고 있으므로 데이터들이 통합되어 종합적으로 활용되지 못하는 상황이다. 다양한 측정 방법을 통해서 획득된 환자에 대한 거시, 미시 데이터가 일관성 있게 통합되지 않은 상태에서, 환자 중심으로 분석이 이루어지지 않고서는, 개인 특화된 의료를 위한 진단, 치료선택, 예후 추정을 지원하기 위한 정보를 전문 의료인에게 효과적으로 제공하는 데 한계가 있다. 따라서 유전자 수준의 미시 데이터, 임상학적 거시 데이터, 가족력, 유사 질환군 등의 연관정보 데이터를 통합·연계하여 이력으로 관리하고, 데이터의 점진

적 누적이 가능한 통합의료 데이터웨어하우스를 구축하여 다차원적인 분석이 가능한 데이터 뷰(view)를 제공해야 한다[4].

본 연구에서는 환자의 진단·치료를 지원하기 위해 데이터를 점진적으로 누적·통합·연계하는 데이터베이스 구축을 위한 프레임워크를 설계한다. 이러한 자료들은 해당 병원 내 암 연구의 인프라로서, 각종 암 질환 데이터베이스를 제공하는 역할을 수행한다. 본 논문에서는 의과대학 비뇨기과 실험실의 비뇨기계 종양 환자(방광암, 신장암, 전립선암)의 임상학적 거시데이터 및 유전학적 미시데이터를 수집 완료하고 수집된 데이터의 각 종양별 특징을 분석하여 독립적인 데이터베이스 시스템을 구축한다. 또한, 데이터베이스 구축 시 고려해야할 사항을 도출하고 비뇨기계 종양에 관련한 데이터 모델들의 분석을 통해 데이터의 처리와 이력을 효율적으로 관리할 수 있도록 한다. 그리고 의료 데이터베이스에 대한 신속한 수정 및 갱신을 극대화하기 위해 기존 데이터베이스를 활용하여 효율적인 비뇨기계 종양관련 데이터베이스를 설계하고, 의료통합관리시스템의 구축을 위한 프레임워크를 설계한다.

### 2. 비뇨기계 종양관련 데이터베이스의 설계

통합 의료정보 데이터베이스의 설계를 위해서는 먼저 요구사항 분석이 이루어져야 하며, 이러한 요구사항 분석을 통해 의료정보 테이블들을 구성하여 ER-diagram 을 설계할 수 있다.

1 이 논문은 2009년 교육과학기술부로부터 지원받아 수행된 연구 "지역거점연구단육성사업 / 충북BIT연구중심대학사업단)와 2009년도 정부(교육과학기술부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(No. 2009-0063257).

2.1 요구사항 분석

통합 의료정보 데이터베이스를 구축하기 위해서는 대상 업무의 특성을 명확하게 파악하고, 관련된 환경적 특성을 잘 이해하는 것이 중요하다.

표 1 은 비뇨기과 의사들의 요구 조건에 따라 통합 암 정보 데이터베이스에서 제공하는 총 6가지의 주요 정보를 분류하여 보여주고 있다.

<표 1> 통합 암 정보 데이터베이스의 요구사항 분석

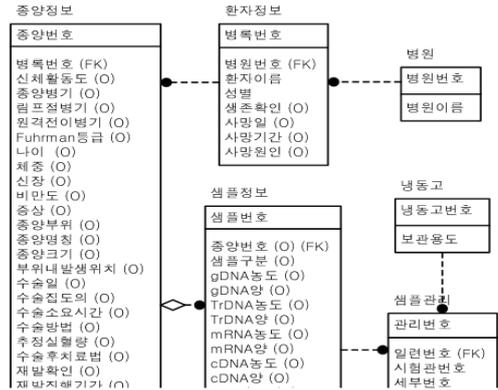
번호	제공 정보	요약
1	환자정보	암 환자의 이름, 성별, 생년월일 등 기본정보를 제공
2	샘플정보	암 환자의 혈액, 혈청, 유린, 조직에서 추출한 유전 정보를 제공
3	임상정보	암 환자의 수술 전, 또는 수술 후의 임상 정보를 제공
4	수술정보	수술이 진행되고 있는 동안에 측정된 정보를 제공
5	실험정보	암 환자에게서 추출한 종양 세포를 이용한 실험정보를 제공
6	보관정보	종양 조직을 냉동상태로 보관하기 위한 저장정보를 제공

환자정보는 병원에 입원한 암 환자들을 대상으로 환자의 이름, 성별, 생년월일 등과 같은 개인정보를 제공하고, 샘플정보는 암 환자들에게서 채취한 혈액, 혈청, 유린, 조직에서 추출 가능한 유전정보를 제공한다. 임상정보에서는 암 환자의 수술 전, 또는 수술 후의 임상상태를 파악하기 위한 정보를 제공하며, 외과적 수술을 통해 제거한 종양의 특성도 이 임상정보에 해당한다. 수술정보는 종양을 제거하기 위해 수술을 진행하는 동안에 측정되는 수술시간, 수술방법 등과 같은 정보를 제공하게 된다. 실험정보는 종양 세포에서 추출한 단백질, RNA 등을 사용하여 획득되어지는 유전자 발현 정보를 제공한다. 보관정보는 암 환자로부터 획득한 종양 조직을 체계적으로 관리하기 위해 초저온 냉동고에 저장되어지는 상태정보를 제공하게 된다.

이러한 사용자 요구 분석을 통해 분류된 환자정보, 샘플정보, 임상정보, 수술정보, 실험정보 그리고 보관정보를 연계하여 데이터베이스를 구축함으로써 다양한 형태의 정보를 제공하는 것이 가능해진다. 따라서 이러한 요구조건이 올바르게 이루어짐으로써 개인 특화된 의료를 위한 진단, 치료선택, 예후 추정을 지원하기 위한 정보를 전문 의료인에게 효과적으로 제공할 수 있다.

2.2 스키마 설계

그림 1 은 사용자의 요구사항 분석을 통해 도출된 암 관련 정보 테이블에 대한 ER-Diagram을 나타낸다.



(그림 1) 암 정보 테이블 ERD

미리 정의된 요구 분석 내용에 따라 병원 테이블, 환자정보 테이블, 종양정보 테이블, 샘플정보 테이블, 샘플관리 테이블, 냉동고 테이블, 실험정보 테이블로 구성되어 있다. 병원 테이블은 수술을 통해 암을 제거하고 조직 검사를 시행하는 병원 정보를 저장할 수 있도록 병원 식별번호, 병원이름과 같은 속성을 가지도록 설계하였다. 환자정보 테이블은 암 확진 판정을 받고 종양을 제거하는 수술을 시행 받는 환자들의 기본 정보가 저장되며 기본 키로써 병록번호라는 속성을 이용하여 환자를 구분하게 된다. 종양 정보 테이블은 수술을 통해 제거한 종양정보, 암 환자의 수술 전과 후의 임상정보, 그리고 종양을 제거하기 위해 진행했던 수술과정에서 측정된 정보를 제공하기 위한 속성들로 구성되어 있다. 표 2 의 종양번호 필드를 이용하여 각 종양을 구분한다. 종양 분류번호 체계는 문자열 10 자리로 구성되어 있다. 병명은 해당 암의 종류를 나타내고, 임의 번호는 해당 암을 제거하기 위해 수술을 받은 환자들의 수를 표시한다. 마지막 필드에는 암 제거 수술을 받은 암 환자의 재발 횟수를 나타낸다.

<표 2> 종양 분류번호 체계

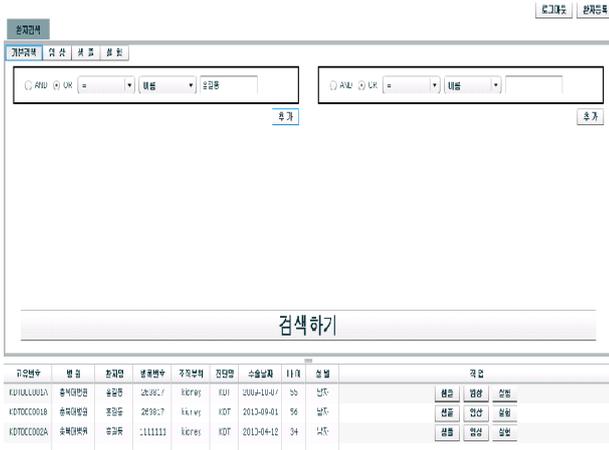
1	2	3	4	5	6	7	8	9	10
병명 (3자리 문자)			임의 번호 (6자리 숫자)						재발 횟수

샘플정보 테이블은 실험을 통해 얻은 혈액, 혈청, 유린, 그리고 조직으로부터 추출한 DNA, RNA, 그리고 단백질의 수집 정보를 나타내기 위한 여러 개의 속성들을 가진다. 샘플관리 테이블은 암 환자로부터 획득한 종양 조직을 체계적으로 관리하기 위하여 관리번호, 냉동고 번호, 시험관 번호, 시험관 번호와 같은 속성을 가지며 샘플관리 번호를 기본 키로 지정하여 보관하고 있는 샘플을 구분한다. 냉동고 테이블은 샘플을 보관하고 있는 초저온 냉동고에 대한 정보를 저장하며, 제품 일련번호를 기본 키로 지정하여 각각의 냉동고를 구분하게 된다. 그리고 각 냉동고에 대한 보관 용도를 파악하기 위한 속성을 가진다. 실험정보

테이블은 암 환자로부터 추출한 종양 세포를 이용하여 실험한 유전자 정보를 저장하며 그 관련 속성들을 포함하고 있다.

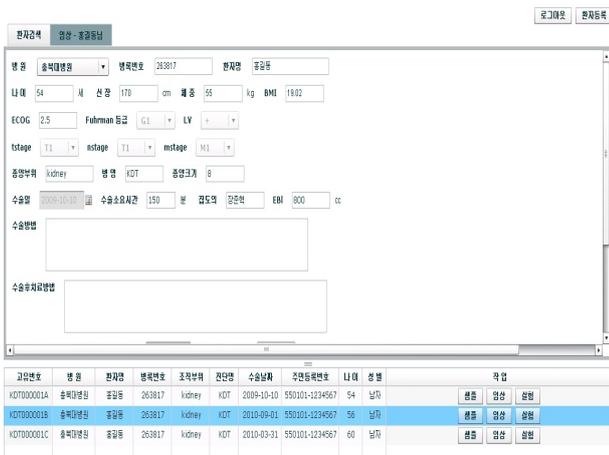
### 3. 암 정보 통합 관리 서비스 구현

수집한 데이터의 암 정보 통합 관리 서비스를 위해 각종 종양에 대한 정보를 제공하는 인터페이스를 구축하였다.



(그림 2) 통합 의료정보 데이터 검색

그림 2 는 특정 암 정보 데이터 조회를 위한 검색 조건을 입력하는 인터페이스를 보여준다. 사용자는 개인, 샘플, 임상 그리고 실험 정보 중 하나를 선택하여 검색범위를 결정한다. 그 다음, 선택된 범위 내에 해당하는 속성들을 이용하여 하나 이상의 검색 조건을 설정한다. 그리고 검색 버튼을 클릭하면 설정한 조건에 맞는 데이터를 검색할 수 있다.



(그림 3) 환자별 임상정보 조회 결과

그림 3 은 암 정보 데이터를 검색하여 조회한 결과를 보여준다. 사용자는 조건검색을 통해 조회 요청을 하게 되

면 검색 조건에 만족하는 해당 종양 정보를 확인할 수 있다. 그리고 샘플, 임상, 실험 버튼 중 하나를 선택하여 특정 암 환자의 샘플정보, 임상정보 그리고 실험정보를 제공받는 것이 가능하다. 이를 통해 비뇨기과 전문의는 정확한 데이터를 더욱 신속하고 간편하게 등록, 조회, 관리할 수 있으며 질병의 진단, 경과, 치료효과 및 예후 등을 판단하는데 중요한 역할을 한다.

### 4. 결론

본 연구에서는 비뇨기계의 암 정보를 체계적으로 수집, 관리, 분석하기 위한 통합의료정보시스템의 데이터베이스 설계 및 구축을 위해 전체적인 프레임워크를 제시하였다. 기존 병원시스템에서는 비뇨기과 환자들의 데이터들이 개별적으로 수집·관리·분석되어 종합적으로 활용되지 못하고 있다. 다양한 측정 방법을 통해 획득된 환자에 대한 거시, 미시 데이터가 일관성 있게 활용되지 못하므로 환자 중심으로 분석이 이루어지지 않고 있다. 통합 의료정보 데이터베이스를 구축함으로써 다른 암과의 비교분석이 용이하여 연관성을 확인할 수 있다. 또한, 암 정보와 관련하여 신뢰도 높은 데이터 수집이 가능함으로써 안전하고 효율적인 정보 분석을 수행할 수 있도록 기반 환경을 제공할 수 있다. 향후 연구로, 사용자가 자주 사용하는 질의에 대해서 모니터링 함으로써 데이터베이스의 처리 능력을 개선할 수 있는 방안에 대한 연구를 진행할 예정이다.

### 참고문헌

[1] <http://www.mkhealth.co.kr/>  
 [2] 강대희, 이경모, “유전체 역학연구의 동향”, 예방의학 회지, 제 36권 3호, pp 213-222, 2003  
 [3] 보건복지부, “한국중양암 등록 사업 연례보고서”, 2002  
 [4] 김유준, 권훈, 곽호영, “웹 서비스를 활용한 의료정보 통합 시스템 설계”, 한국콘텐츠학회, 제5권 제2호, pp 857-860, 2007