

한국인 유전체 역학 정보의 효율적 관리를 위한 정보 관리 시스템

양은주^o, 안윤진, 이지은, 김규찬

국립보건원 유전체연구소 유전체역학정보실

{ejyang^o, carotene, jinney}@ngri.re.kr, k2kimm@nih.go.kr email: ejyang@ngri.re.kr

An System for Efficient Management of Genomic Epidemiological Cohort Study Information from Koreans

Eun Joo Yang^o, Younjhin Ahn, Ji Eun Lee, Kuchan Kimm

Division of Epidemiology and Bioinformatics, National Genome Research Institute, KNIH

요 약

본 논문에서는 경기도 안성과 안산 지역에 설치된 유전체역학센터에서 조사 지역에 거주하는 45세 이상 69세 이하의 성인을 대상으로 고혈압, 당뇨, 골다공증, 천식, 비만 등 총국민의료비용에서 큰 부분을 차지하는 4-5개의 중요 만성질환을 주요 분석 분야로 진행중에 있는 한국인 유전체 역학조사사업으로부터 산출되는 임상검사 및 역학정보를 입력·관리하는 시스템에 대해 소개하고 있다. 검진 대상자를 접수하는 접수자 및 정보 관리자는 본 시스템을 통해 검진 대상자의 개인식별 정보, 생활습관 정보 등과 같은 설문 정보와 임상 검사 정보를 입력 후 이에 대한 조회, 관리, 집계 현황 출력, 데이터 백업 등을 수행할 수 있다.

1. 서론

2000년 미국을 중심으로 하는 선진국들이 인간 유전체 지도의 초안을 발표한 이래 각 민족에 특이한 질병 관련 유전자의 탐색이 본격화되고 있다. 이에 따라, 전 세계적으로 각 민족에게 특징적으로 노출되는 환경, 식이 요소 등과 유전자와의 관련성을 도출하는 연구 또한 활발히 진행되고 있다[4]. 한국의 경우, 서울대-서울 남성 코호트의 암연구, 연세대-강화 프로젝트의 고혈압 연구, 국립보건원 유전체연구소 유전체역학센터의 안성-안산 코호트 연구가 대표적이다. 유전체역학센터에서는 현재 심혈관 질환, 당뇨와 같은 대사 질환 그리고 근골격계 질환, 골다공증 및 치매 질환과 같은 조사대상 질병을 중심으로 질환 유부 및 병인 요소들의 변화를 파악하기 위해 각종 임상검사 및 역학 관련 설문 정보를 수집한다. 이들 수집된 정보 입력을 위해 각 유전체역학센터에서는 이 논문에서 거론하고 있는 C/S 기반의 역학·임상검사 정보 입력 프로그램을 사용중에 있다. 유전체 역학 사업 단계별로 검진자에 대한 조사 대상 질병, 검진일, 설문 항목이 바뀌므로 이들 정보를 관리하는 시스템 역시 계속 보완하고 있는 실정이다. 현재 선진국의 경우, 역학 조사로 미국 심혈관 질환의 risk factor를 규명하여 심혈관 질환 이환율을 감소시킨 것으로 유명한 Framingham Heart Study에서도 50여년 전부터 주민의 임상검사, 역학 정보

를 수집하고 있다. 그외에 영국(health survey), 일본(국립 순환기연구소의 Suita 연구), 캐나다(hearth health survey)에서는 질환의 발생률, 진행 정도, 사망률을 조사하기 위해 이미 수십년 전부터 대규모의 국가적 역학사업이 진행중이다[4]. 그러나, 시공간적인(spatio-temporal) 특성(attribute)을 지닌 이들 임상검사/역학 정보에 대한 체계적인 data- and process modeling 기법 및 구현(implement) 시스템에 대한 논문을 찾기란 쉽지 않다. 전세계 computer science 관련 논문 검색 엔진인 DBLP(DataBase systems and Logic Programming)를 통해서도 역학 정보 관리 시스템에 대한 논문은 1편도 없었다[1]. 단지, cancer cluster 분석을 위해 고안된 독일의 CARLOS (Cancer Registry Lower Saxony) 프로젝트에서 시행한 역학 정보, 통계 정보 통합 분석 시스템 구축 시 고려한 시간, 공간, 통계적 특성에 기반한 multidimensional data 분석 모델링 기법을 소개한 논문만 찾아볼 수 있었다[8, 9]. 또한, 모든 생물 정보에 대한 논문 검색 시스템인 PubMed에서 조차 역학 정보 관리 시스템에 대한 논문을 찾아볼 수 없었다[7]. 이러한 이유는 유전체 구조, 단백질 발현 정보 등 방대한 생물 정보를 활용하는 대규모 역학 연구가 진행된 적이 없으므로, 따라서 시-공간 개념을 제외하고는 비교적 단순 형태의 역학 정보 DB와 복잡하고 계속 변이할 것으로 예상되는 schema를 지니는 생물 정보 DB가 연동된 전례가 없었기 때문이 아닌가 추정되고 향후 이 분야의 논문이 계속 발표될 것으로 예상된다.

2. 업무 흐름도 및 개발 환경

이 연구는 경기도 안성과 안산 지역에 설치된 유전체역학센터에서 조사 지역에 거주하는 45세 이상 69세 이하의 성인을 대상으로, 고혈압, 당뇨, 골다공증, 천식, 비만 등 총국민의료비용에서 큰 부분을 차지하는 4-5개의 중요 만성질환을 주요 분석 분야로 하여 진행되고 있다. 구현 시스템은 역학 정보 입력/조회/관리, 자료집계 현황 및 통계, 영양소 정보 입력/조회, 데이터 백업, 임상검사 정보 업로드, 사용자 관리, 시스템 관리 업무 구분 모듈 등으로 구성되어 있고, 각 부분별로 세분화된 기능이 수행되도록 설계하였다. 시스템은 클라이언트/서버 구조로서 서버 운영체제(OS)는 Linux사의 Red Hat release 7.1 환경이며, 클라이언트는 모든 windows platform에 호환되도록 개발하였다. 서버에서 운영되는 DBMS는 Oracle 9i를 사용하였으며, 개발툴은 Delphi v.5.0을 사용하였다.

[표 1] 시스템 구현 환경

구분	환경	
기종	client	Pentium급 PC
	server	HP workstation x1100 Base SPU
운영체제	client	Windows 95/98/NT/XP
	server	Linux Red Hat r. 7.1
개발 언어	Delphi v. 5.0	
DBMS	Oracle DBMS Standard Edition 9i	
LAN	TCP/IP	

3. 구현 함수 및 화면

3.1 기능별 화면

1) 로그인 시 1차, 2차 검진 대상별 DB 선택 기능

- DB 서버에 로그인 시 로그인 화면상의 combo-box componen를 통해 1, 2차 검진을 중 하나를 선택
- 1, 2차 server name은 동일하나 각 user name이 1, 2차로 구분되므로 이에 따라 CBOGROUPChange() 함수가 호출되어 기본적인 연결설정이 셋팅됨.
- 각 연차별 (1, 2차) 정보는 서로 다른 tablespace 영역에 저장되어 있음.

2) 정보 조회 기능

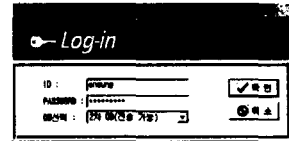
- 사용자가 화면상에서 체크한 조회조건은 쿼리조건이 되어 쿼리문을 담고있는 strSQLADD변수에 계속해서 추가됨. 이로 인해 얻어진 최종 strSQLADD값이 조회할 쿼리가 되어 조회를 수행함.

3) 정보 입력 상태 확인 기능

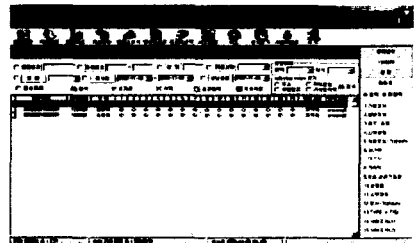
- 품의 상태가 수정모드 (결과항목클릭)인지 입력모드(신규

입력클릭)인지 판별됨.

- 수정모드인 경우 검수모드(입력오류 방지를 위한 cross-validation 체크 기능임)인지 아닌지 판별됨.
- 수정모드이고, 검수모드이면 gumsu 프로시저를 호출, 수정모드이고, 검수모드가 아니면 setting() 함수를 호출, 입력모드이면 setclear() 함수가 호출됨.



[그림 1] 로그인 화면



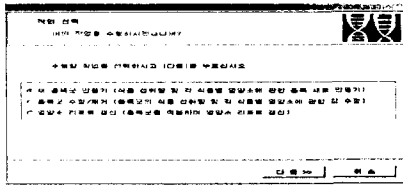
[그림 2] 정보 조회 화면

4) 3차원 통계 분석 디스플레이 및 엑셀 변환 백업 기능

- 해당 읍, 면, 동 count 와 해당 리 count를 구해 grid의 세로줄을 정함.
- '5세 단위', '10세 단위'에 따라 출력될 grid의 영역이 '0'으로 초기화 되고, '소계' 타이틀이 출력됨.
- 연령대별, 남녀별로 구한 카운트값이 해당 grid 셀부분에 출력됨.
- 메모 컴포넌트가 생성되어 기본 설정 세팅됨.
- grid의 셀에 담긴 내용을 한줄 단위로 탭구분자로 tempstr에 담음.
- 메모장에 tempstr값이 저장된 후 엑셀 file로 저장됨.

5) 영양소군 설정 및 분석 기능

- 화면 내의 새가지 항목 중 하나를 선택할 수 있음.
- 새 종목군 만들기 선택시, SpeedButton3Click() 함수 호출로 '새로운 종목'을 추가한 뒤 결과 화면이 출력됨.
- 종목군 수정/제거 선택시, 수정은 SpeedButton8Click() 함수 호출로 수정화면 출력 후, SpeedButton16Click() 함수 호출로 '종목군명'을 수정 가능케함. 삭제는 SpeedButton15Click() 함수 호출로 종목군이 삭제됨.
- 영양소 리포트 갱신(update) 선택시, setcalc() 함수 호출로 '식품섭취량'이 처리되고, setrespd() 함수 호출로 '식품별 영양소 섭취량'이 처리되며, 이때 setrespd() 함수 호출로 '1일 영양소 섭취량'이 처리됨.



[그림 3] 영양소군 설정 화면

영양소군	영양소	단위	기준량	설정량
단백질	질소 (N)	g/kg	14.0	14.0
지방	지방산 (FA)	g/kg	20.0	20.0
탄수화물	탄수화물 (C)	g/kg	20.0	20.0
비타민	비타민 (V)	g/kg	1.0	1.0
미네랄	미네랄 (M)	g/kg	1.0	1.0
수용성 섬유소	수용성 섬유소 (S)	g/kg	1.0	1.0
비타민 B1	비타민 B1 (B1)	g/kg	2.0	2.0
비타민 B2	비타민 B2 (B2)	g/kg	1.0	1.0
비타민 B6	비타민 B6 (B6)	g/kg	1.0	1.0
비타민 C	비타민 C (C)	g/kg	1.0	1.0
비타민 E	비타민 E (E)	g/kg	1.0	1.0
비타민 K	비타민 K (K)	g/kg	1.0	1.0
비타민 B12	비타민 B12 (B12)	g/kg	1.0	1.0
비타민 D	비타민 D (D)	g/kg	1.0	1.0
비타민 A	비타민 A (A)	g/kg	1.0	1.0
비타민 H	비타민 H (H)	g/kg	1.0	1.0
비타민 P	비타민 P (P)	g/kg	1.0	1.0
비타민 Q	비타민 Q (Q)	g/kg	1.0	1.0
비타민 U	비타민 U (U)	g/kg	1.0	1.0
비타민 V	비타민 V (V)	g/kg	1.0	1.0
비타민 W	비타민 W (W)	g/kg	1.0	1.0
비타민 X	비타민 X (X)	g/kg	1.0	1.0
비타민 Y	비타민 Y (Y)	g/kg	1.0	1.0
비타민 Z	비타민 Z (Z)	g/kg	1.0	1.0
비타민 AA	비타민 AA (AA)	g/kg	1.0	1.0
비타민 BB	비타민 BB (BB)	g/kg	1.0	1.0
비타민 CC	비타민 CC (CC)	g/kg	1.0	1.0
비타민 DD	비타민 DD (DD)	g/kg	1.0	1.0
비타민 EE	비타민 EE (EE)	g/kg	1.0	1.0
비타민 FF	비타민 FF (FF)	g/kg	1.0	1.0
비타민 GG	비타민 GG (GG)	g/kg	1.0	1.0
비타민 HH	비타민 HH (HH)	g/kg	1.0	1.0
비타민 II	비타민 II (II)	g/kg	1.0	1.0
비타민 JJ	비타민 JJ (JJ)	g/kg	1.0	1.0
비타민 KK	비타민 KK (KK)	g/kg	1.0	1.0
비타민 LL	비타민 LL (LL)	g/kg	1.0	1.0
비타민 MM	비타민 MM (MM)	g/kg	1.0	1.0
비타민 NN	비타민 NN (NN)	g/kg	1.0	1.0
비타민 OO	비타민 OO (OO)	g/kg	1.0	1.0
비타민 PP	비타민 PP (PP)	g/kg	1.0	1.0
비타민 QQ	비타민 QQ (QQ)	g/kg	1.0	1.0
비타민 RR	비타민 RR (RR)	g/kg	1.0	1.0
비타민 SS	비타민 SS (SS)	g/kg	1.0	1.0
비타민 TT	비타민 TT (TT)	g/kg	1.0	1.0
비타민 UU	비타민 UU (UU)	g/kg	1.0	1.0
비타민 VV	비타민 VV (VV)	g/kg	1.0	1.0
비타민 WW	비타민 WW (WW)	g/kg	1.0	1.0
비타민 XX	비타민 XX (XX)	g/kg	1.0	1.0
비타민 YY	비타민 YY (YY)	g/kg	1.0	1.0
비타민 ZZ	비타민 ZZ (ZZ)	g/kg	1.0	1.0
비타민 AA	비타민 AA (AA)	g/kg	1.0	1.0
비타민 BB	비타민 BB (BB)	g/kg	1.0	1.0
비타민 CC	비타민 CC (CC)	g/kg	1.0	1.0
비타민 DD	비타민 DD (DD)	g/kg	1.0	1.0
비타민 EE	비타민 EE (EE)	g/kg	1.0	1.0
비타민 FF	비타민 FF (FF)	g/kg	1.0	1.0
비타민 GG	비타민 GG (GG)	g/kg	1.0	1.0
비타민 HH	비타민 HH (HH)	g/kg	1.0	1.0
비타민 II	비타민 II (II)	g/kg	1.0	1.0
비타민 JJ	비타민 JJ (JJ)	g/kg	1.0	1.0
비타민 KK	비타민 KK (KK)	g/kg	1.0	1.0
비타민 LL	비타민 LL (LL)	g/kg	1.0	1.0
비타민 MM	비타민 MM (MM)	g/kg	1.0	1.0
비타민 NN	비타민 NN (NN)	g/kg	1.0	1.0
비타민 OO	비타민 OO (OO)	g/kg	1.0	1.0
비타민 PP	비타민 PP (PP)	g/kg	1.0	1.0
비타민 QQ	비타민 QQ (QQ)	g/kg	1.0	1.0
비타민 RR	비타민 RR (RR)	g/kg	1.0	1.0
비타민 SS	비타민 SS (SS)	g/kg	1.0	1.0
비타민 TT	비타민 TT (TT)	g/kg	1.0	1.0
비타민 UU	비타민 UU (UU)	g/kg	1.0	1.0
비타민 VV	비타민 VV (VV)	g/kg	1.0	1.0
비타민 WW	비타민 WW (WW)	g/kg	1.0	1.0
비타민 XX	비타민 XX (XX)	g/kg	1.0	1.0
비타민 YY	비타민 YY (YY)	g/kg	1.0	1.0
비타민 ZZ	비타민 ZZ (ZZ)	g/kg	1.0	1.0

[그림 4] 영양소 섭취량 출력 화면

3.2 세부 기능

역학·임상검사 정보 입력/관리 프로그램의 각 세부 모듈에 대한 설명은 아래 표 2와 같다.

[표 2] 세부 기능 설명

구분	기능
사용자 로그인	1차, 2차 검진 대상별 DB 선택
	ID, PW 입력
역학 정보 관리	데이터 입력, 조회 기능
	역학 정보 입력상태 확인 기능
임상검사 정보 관리	데이터 입력, 조회 기능
	임상검사 정보 입력상태 확인 기능
영양소 관리	영양소 설정
	영양소 분석
성별/연령별/지역별통계 분석	3차원 분석 디스플레이 및 엑셀로의 변환 백업
검사 결과 출력	역학, 임상 검사 정보를 토대로 결과 출력
데이터 download	오라클 DB를 access file로 변환 및 백업
임상검사 정보 upload	DB 서버로의 upload 기능
사용자 관리	사용자 id 추가, 수정, 삭제

4. 결론 및 고찰

대부분의 대규모 역학 사업에서는 일정 term을 기준으로 하는 환자의 반복적 검진정보와 같은 시간적 속성뿐만 아니라 지리적 온도, 습도, 생활 풍습, 경위도 좌표 등과 같은 환경요소도 고려하여 이를 엔티티 타입으로 사상한다[4, 8, 9]. 따라서, 역학조사 사업으로 파생되는 정보에 대한 DB 구축시 데이터에 대한 시공간 속성을 모두 고려해야 하는 것은 의심할 여지가 없다. 그러나, 현 시스템은 시간적, 지리적 속성을 그리 많이 고려하지 않았다. 따라서, 이에 대한 속성을 반영토록 계속 보완해 나갈 예정이다.

참고문헌

- [1] DBLP(DataBase systems and Logic Programming), <http://dblp.uni-trier.de/>, 2003.
- [2] Genome Frontier Project of Japan <http://www.genome.ad.jp/documents/gf/frontier.html>, 2003.
- [3] Introduction to the UK Biobank; <http://www.ukbiobank.ac.uk/>, 2003.
- [4] Korea National Institute of Health. Central Genome Research Center plan (Health and Medical Technology Development Project), 2000
- [5] Medical Research Council. Public news; http://www.mrc.ac.uk/index/public-interest/public-news/public-news_archive/publicnews_archive_1_2002/public-biobank_uk.htm, 2003.
- [6] National Human Genome Research Institute. Haplotype map development report. <http://www.genome.gov/>, 2002.
- [7] PubMed, <http://www.ncbi.nih.gov/entrez/query.fcgi>, 2003.
- [8] V. Kamp, Frank Wietek: Database System Support for Multidimensional Data Analysis in Environmental Epidemiology. IDEAS 1997: 180-190.
- [9] V. Kamp and F. Wietek. Intelligent Support for Multidimensional Data Analysis in Environmental Epidemiology. In Advances in Intelligent Data Analysis, Second International Symposium, IDA'97. LNCS 1280, Springer, 1997.