

I. 한국판 PICO Linguist와 BabelMeSH 구축

Construction of Korean version PICO Linguist and BabelMeSH

Boyoung Jung*, Sangmi Lee*, Paul Fontelo†, Fang Liu †., Sukil Kim*

*Department of Preventive Medicine, College of Medicine, The Catholic University of Korea, Seoul 137-701, sikimMD@catholic.ac.kr

†High Performance Computing and Communications, Lister Hill National Center for Biomedical Communications

National Library of Medicine Associate Fellowship Program

National Library of Medicine, Bethesda, MD 20894

Abstract: The PICO Linguist and BabelMeSH are search tools intended for non-English speaking PubMed users from 10 countries. A database of Korean medical terms was added to the search engine using concept equivalents of English terms to Korean ones. There are 2 problems in constructing Korean version. Korean user interface was built to fit Korean version of BabelMeSH. Users can limit the search output according to the language of publications. However, there are some limitations Korean version PICO Linguist and BabelMeSH. The retrieved citations are still in English. To satisfy Korean readers, the citations should be translated in Korean in the future. It would be useful to search both Korean and English journals just using Korean terms. The preliminary evaluation of the Spanish system showed positive responses. It could be useful and convenient tools in finding the medical literature especially for Korean medical terms that may be difficult to express in English.

Keyword: MeSH, Medical Subject Heading, MEDLINE/Pubmed, PICO, UMLS

배 경

선진 복지국가의 궁극적인 목적은 국민의 건강을 증진시키고 삶의 질을 높이는 것이다. 정보화시대에 이를 가능하게 하기 위해서 의료정보의 활용은 필수적이다. 현재 우리나라는 의료정보 인프라구축단계에 있으며, 넘쳐나는 의료정보들 중 유익한 정보를 빠르고 정확하게 검색하는 도구 개발이 시급하다. 최신의료에 관한 유익한 정보는 교육, 연구, 정책에 사용될 수 있다. 또한 의료서비스제공자들에게 진료시점에 적절히 제공되어 근거중심의학(evidence-based medicine)을 뒷받침 할 수 있는 자료가 된다. MEDLINE은 세계 70여 개국 3,363개 저널에서 생성된 생명의학에 관한 정보를 담은 데이터베이스다. MEDLINE에서 free-text 검색도구로써 개발된 askMEDLINE 개발 보고서에 따르면, 사람들이 보통 의학문헌 검색 시 가장 친숙한 단어로 검색하는 경향을 보인다. 특히, 비영어권 국가의 경우 영어로 표현하기 힘든 의학용어를 가진 문헌검색을 필요로 하기 때문에 이런 현상은 더욱 두드러진다. 미국국립의학도서관(National Library of Medicine, NLM)에서는 영어를 모국어로 사용하지 않는 나라 국민의 문헌검색을 용이하게 하기 위해 PICO Linguist와 BabelMeSH를 개발했다.

PICO Linguist는 Patient(환자)/Problem(환자문제), Intervention(치료방법), Comparison(비 교), Outcome(결과)에 따른 구조화된 임상 질문을 이용해서 의학문헌을 찾는 틀이다. Babel MeSH는 사용자에게 친숙한 모국어를 이용해서 영문 의학문헌을 검색가능하게 해주는 틀이다. 본 연구에서는 한글-영어 용어 쌍을 적용한 한글판 PICO Linguist와 BabelMeSH를 소개하고자 한다. II. 방 법

현재 BabelMeSH에는 Arabic, Chinese, French, German, Italian, Japanese, Portuguese, Russian, Spanish 등 9개 언어가 구축되어 있고, 언어별 데이터베이스 간에는 양적 차이가 있다. 여기에 한국어를 추가시킴으로써 우리에게 익숙한 한국어를 사용해서 영문 의학문헌검색을 가능케 했다.

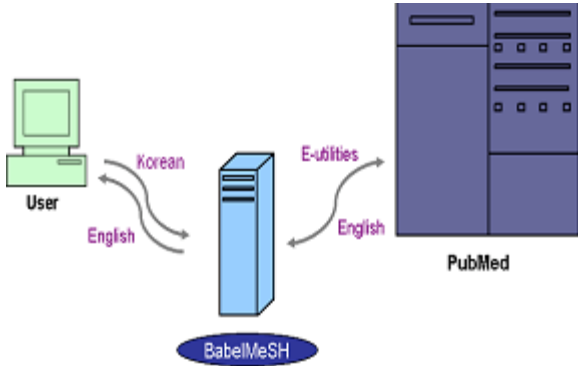
1. 데이터베이스 자료원 및 사용자 인터페이스

데이터베이스의 자료원은 UMLS Metathesaurus다. UMLS Metathesaurus안에는 프랑스어, 독일어, 이탈리아어, 일본어, 포르투갈어, 러시아어, 스페인어 MeSH번역이 있다. Metathesaurus를 바탕으로 각 나라별 소스가 추가 되어 데이터베이스를 구성했다. 프랑스 MeSH 번역판과 UMLF (unified medical lexicon for french)는 Dr. Stephan Darmoni와 Dr. Patrick Ruch로부터 제공되었고, MeSH 아랍어와 통합의료사전은 동양 Mediterranean과 WHO에서 제공했다. 중국어는 웹사이트 (multiple open source web site)에서 추출했다. 한국어는 보건복지부가 보건의료정보표준화구축을 위해 진행했던 보건의료용어표준화연구 결과인 표준통합용어를 이용했다. 이것은 개념기반 213,942개 한글-영어 용어쌍이며, 한글판 PICO Linguist와 BabelMeSH 작업의 기초 데이터베이스가 되었다.

English version PICO Linguist와 BabelMeSH를 기초로 한글 번역작업을 통해 한글판 사용자 인터페이스를 만들었다.

2. 파싱 알고리즘

사용자가 검색창에 한글을 입력하면 파서가 표준통합용어의 한글에 대응되는 영어를 찾는다. 그 후 PubMed내 영어가 속한 문헌을 결과로 보여준다.



복합어가 입력되면, 파서는 띄어쓰기 기준으로 개별용어로 분리하고 개별용어에 해당하는 영어를 찾는다. 만약 표준통합용어 안에 개별언어의 합인 복합용어가 존재할 경우, 복합 용어를 우선 매칭 한다. Table 1은 복합한글용어 [당뇨 신장병증을 동반한 인슐린-비의존 당뇨병, Non-Insulin-dependent diabetic neuropathy]에 대한 개별용어별 매칭을 보여주는 예이다. 파서는 한글 개별용어별 영어가 존재하지만 적합한 번역형태인 복합용어에 해당하는 8번과 11번을 채택한다. 결국, 파서는 이 영어에 해당하는 PubMed내 의학문헌을 찾아 결과로 보여준다. (Table 1)

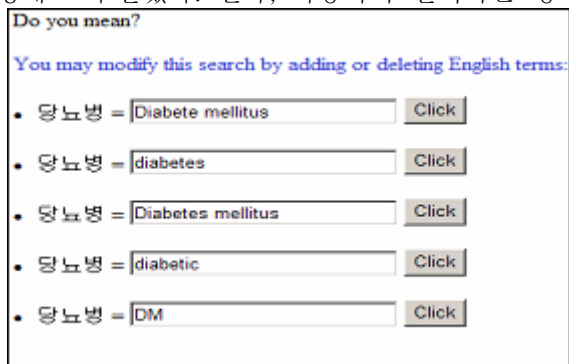
No	English	Korean
1	diabetic	당뇨
2	diabetic	당뇨병
3	diabetes mellitus	당뇨병
4	DM	당뇨병
5	nephropathia	신장병증
6	nephropathy	신장병증
7	renopathy	신장병증
8	<i>diabetic neuropathy</i>	당뇨 신장병증
9	insulin	인슐린
10	non-dependence	비의존
11	<i>non-insulin-dependen</i>	인슐린-비의존

3. 표준통합용어 적용

표준통합용어 적용에 크게 2가지 문제점이 있었다. 첫째, 한글이 BabelMeSH 데이터베이스에 삽입될 때 글자가 깨지는 현상이 나타났다. 이는 상호 다른 문자 체계 때문에 발생하는 문제로, 한글코드를 UTF-8코드로 전환시켜 해결했다. UTF-8코드를 사용하면 전 세계 모든 문자를 동시에 표현가능하고, 다른 언어 체계 간 충돌을 없애 줄 뿐만 아니라 다국어입력 또한 수월하다.

둘째, 표준통합용어는 기초 자료로 사전용어 및 임상용어를 다수 포함하고 있다. 이런 이유로 표준통합용어에는 한 개념을 표현하는 한글형태가 다양하다. 즉, 영어 한 개념 혹은 한글 한 개념에 해당하는 한글과 영어가 일대다 형태를

이론다. 이를 고려하여 대표 용어(Preferred Term, PT)를 기준으로 영어 하나에 해당하는 한글을 모두 보여주는 형태로 구현했다. 결국, 사용자가 선택하는 영어에 따라 문헌 결과가 달라진다.(Fig. 2)



4. 사용자 만족도 조사 (스페인어)

스페인어권 사용자 40명을 대상으로 다음 두 가지 질문에 대한 5점 척도 평가를 하였다.

- (1) Is searching MEDLINE/PubMed in your native language useful?
- (2) How would you rate the results obtained?

III. 결 과

1. 사용자 인터페이스



Figure 3). BabelMeSH with autosuggestion

BabelMeSH 는 XML Http object request Java script를 사용해서 사용자가 찾을 단어를 입력하면 자동으로 제시어를 제공해준다. 사용자는 드롭다운리스트로 생긴 제시어 중 찾을 단어를 선택할 수 있다.(Fig. 3)

PICO Linguist는 Java script code로 인해 사용자가 선택하는 언어에 따라 웹 인터페이스(Web interface)가 자동으로 변화되도록 구현되었다(Fig. 4, 5).

2. 결과 필터

사용자는 문헌 검색 요건에 포함할 언어를 선택할 수 있다. 외국 문헌의 경우, MEDLINE에 인덱스 되기 때문에 검색된 문헌의 결과는 영어로만 제공된다. 결과 페이지는 사용자가 쉽게 볼 수 있도록 [Abstract] [Full Text] [Related] 강조 표시가 함께 제공된다.(Fig. 6)

FICO Linguist [Help](#)
 (PICO=Patient, Intervention, Comparison, Outcome)
 Search MEDLINE/PubMed

Select input language: [Arabic](#) [Chinese](#) [French](#) [German](#) [Italian](#)
[Japanese](#) [Korean](#) [Portuguese](#) [Russian](#) [Spanish](#) [English](#)

Patient/Problem:

Age Group: Gender:

Intervention:

Comparison (leave blank if none):

Outcome (optional):

Select a publication type:

Show journal articles published in:
[Arabic](#) [Chinese](#) [French](#) [German](#) [Italian](#) [Japanese](#) [Korean](#) [Portuguese](#)

FICO 도우미 [Help](#)
 (PICO=Patient(환자), Intervention(치료방법), Comparison(비교), Outcome(결과))

언어를 선택하세요: [영어](#) [중국어](#) [프랑스어](#) [독일어](#) [이탈리아어](#)
[일본어](#) [한국어](#) [포르투갈어](#) [러시아어](#) [스페인어](#) [영어](#)

Patient/Problem:

연령군: 성별:

Intervention:

Comparison (없으면 비워두세요):

Outcome (선택):

출판 종류:

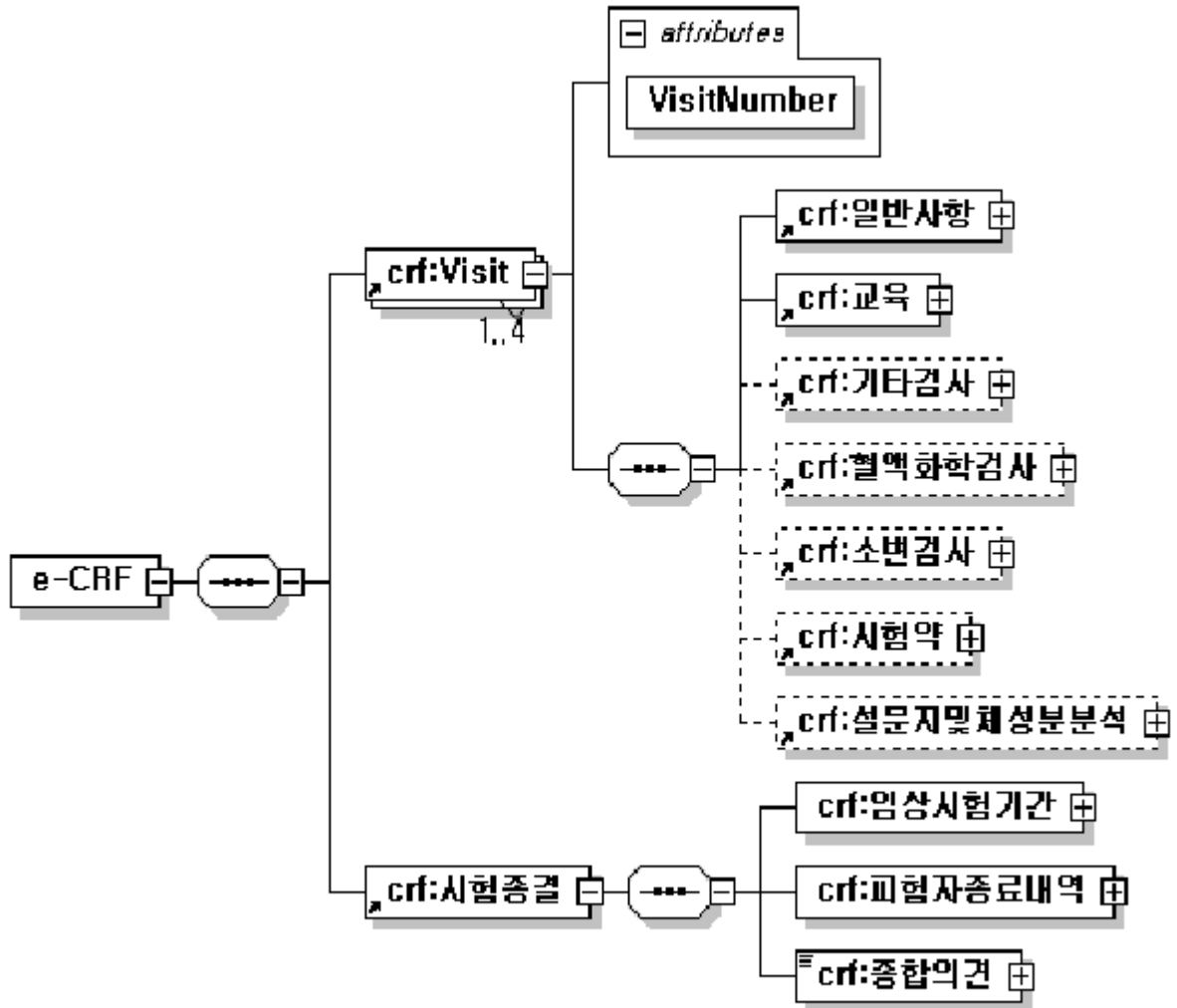
검색에 포함할 논문의 언어:
[이탈리아어](#) [중국어](#) [프랑스어](#) [독일어](#) [이탈리아어](#) [일본어](#) [한국어](#) [포르투갈어](#) [러시아어](#) [스페인어](#) [영어](#)

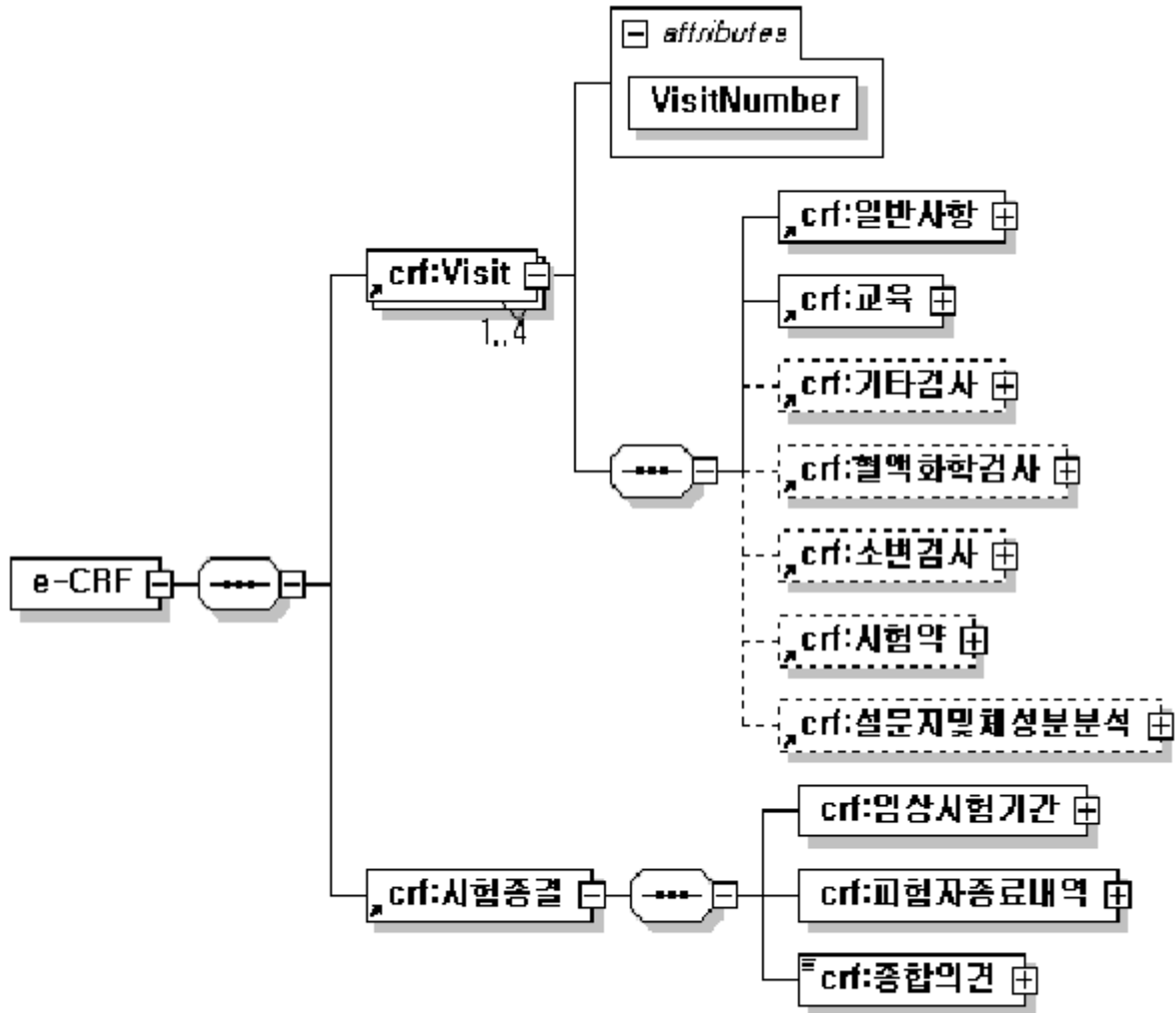
US National Library of Medicine
 Terms: 당뇨병 = **Diabetes** [View](#)
 230 results:

1. [Diabetes type 2 in pediatrics: diagnosis and management]
 Girardin CM, Schwitzgebel VM
 Rev Med Suisse. 2007 Apr; 3(107):1001-5. PubMed ID: 17526374
[Abstract](#) [Full Text](#) [Related](#)

2. [Type 1 diabetes of the child and adult]
 Boffet S, Thérolet C
 Rev Prat. 2007 Jan; 57(1):71-8. PubMed ID: 17432005
[No Abstract](#) [Full Text](#) [Related](#)

3. Recommendations for nutrition best practice in the management of gestational diabetes mellitus. Executive summary (1).
 ...
 Can J Diet Pract Res. 2006; 67(4):206-8. PubMed ID: 17150144
[Abstract](#) [Full Text](#) [Related](#)

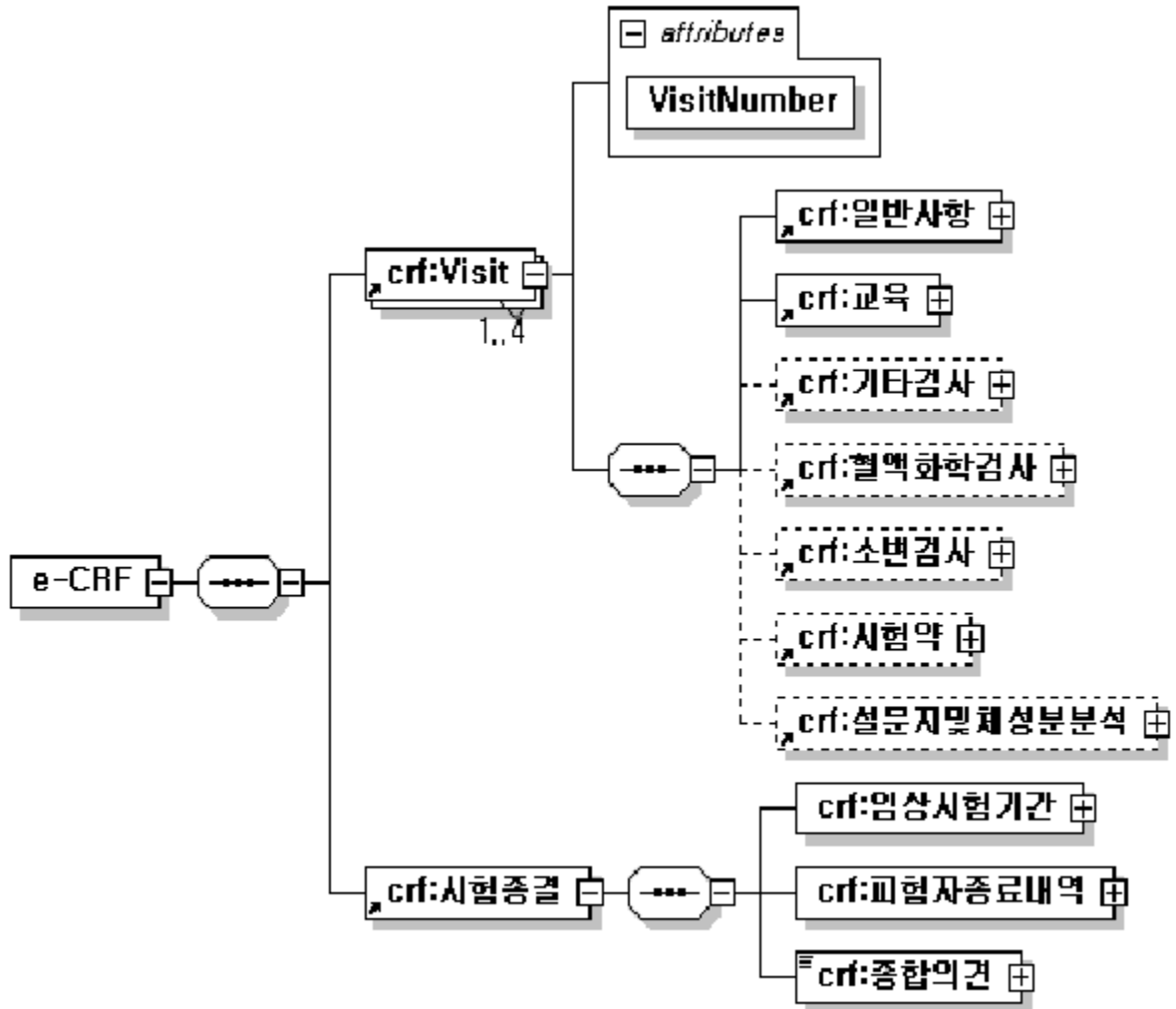


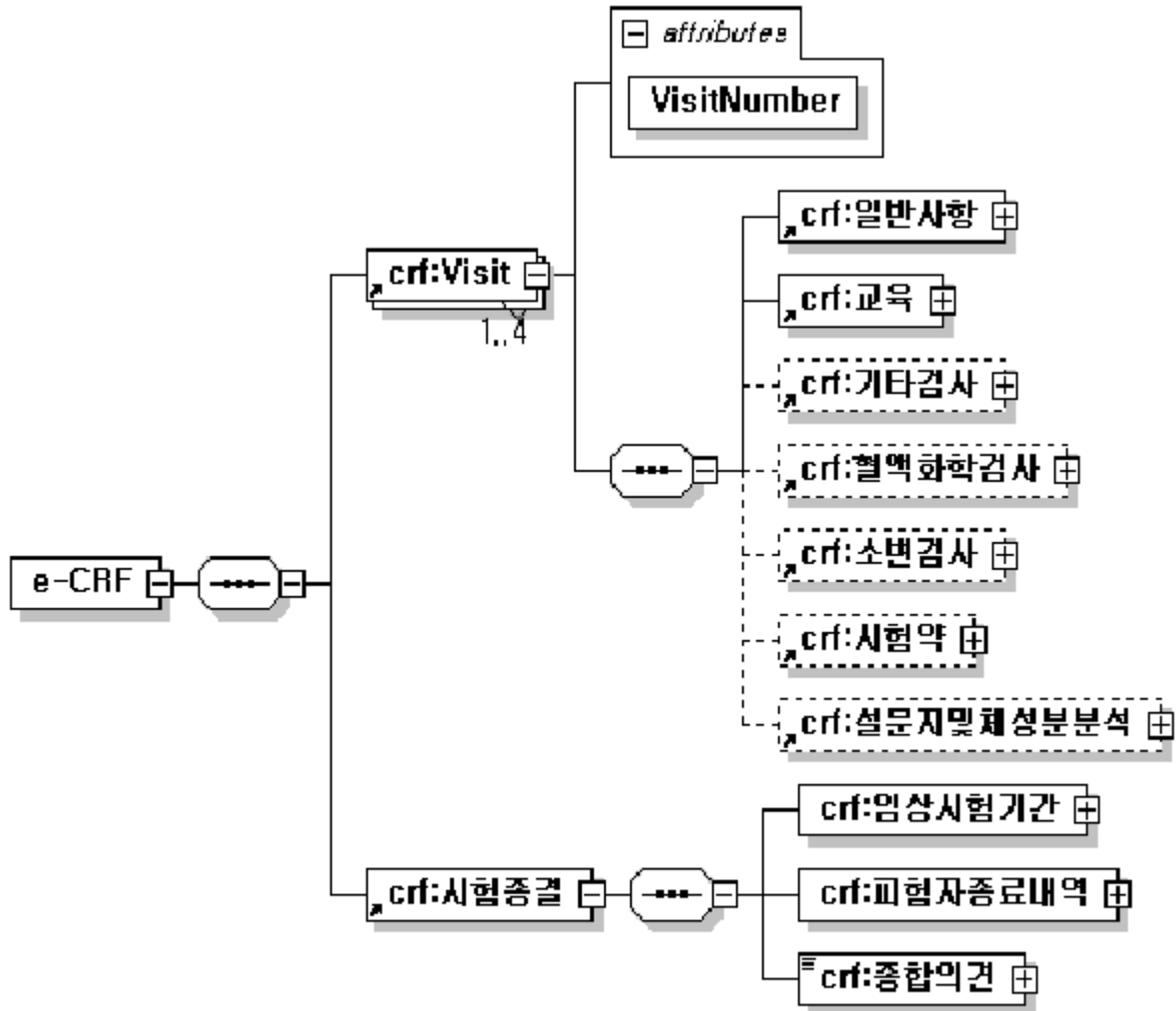


Generated with XMLSpy Schema Editor www.altova.com

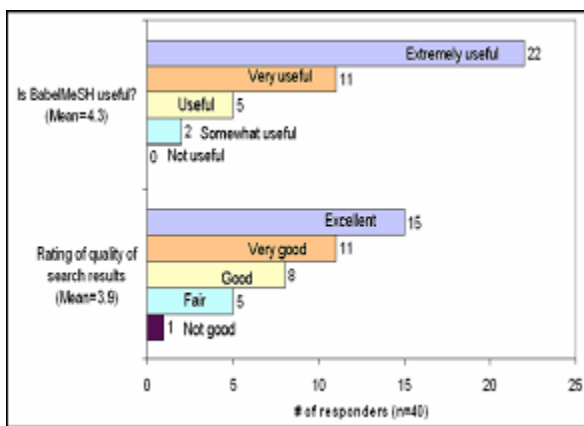
3. 사용자 만족도 조사 결과 (스페인어)

Figure 7은 스페인어판 사용자를 대상으로 한 두 가지 질문의 응답결과를 나타내주는 그래프다. BabelMeSH의 유용성과 검색결과의 질을 평가하는 두 질문에 각각 5점 기준 평균이 4.3 과 3.9의 응답률로 대체적으로 긍정적인 반응을 보였다.(Fig. 7).





Generated with XMLSpy Schema Editor www.altova.com



IV. 결 론

한글판 PICO Linguist와 BabelMeSH는 우리에게 친숙한 한글을 이용해서 영문의학문헌 검색을 할 수 있도록 해주는 툴이다. 현재 BabelMeSH는 영어 외에 먼저 구축된 9개 언어에 본 연구를 통해 한글판 BabelMeSH를 포함하면서 10개 언어로 의학문헌검색이 가능해졌다.

한글판 BabelMeSH는 데이터베이스 특징상 한글-영어가 일대다 형태를 보여 사용자가 원하는 언어를 선택하게끔

구현했었다. 이 결과 생긴 사용자 선호도 분석을 통해 사람들이 자주 사용하는 용어가 무엇인지 파악가능하다. 이 정보를 이용하면 제안단어리스트 선후배열이 가능하다. 또한 분석을 통해 추려진 다 빈도용어는 표준통합용어의 대표용어 선정에 활용될 수 있다.

본 연구에는 3 가지 제한사항이 있다.

첫째, 한글판 BabelMeSH는 표준통합용어의 문자열매핑에 기초했다. 표준통합용어 영어 문자열에 대응되는 MEDLINE/PubMed 의학문헌을 검색한다. 이런 이유로 개념은 같으나, 다른 문자열을 가진 단어에 해당하는 문헌 검색수가 다르다. 그러나 이 문제는 문자열에 대응하는 표준통합용어의 SUI가 CUI와 연결되고, CUI가 UMLS에 있는 MeSH 코드로 연결됨으로써, 단기간에 해결이 가능하다. 즉, UMLS의 CUI를 활용하면 개념별 문헌검색이 가능하다

Visit 1
(골수채취수입일)
 피험자 번호
 피험자 이니셜

Visit 1 골수,주입일 Master 화면 입니다.

골수 채취 기록

마취 및 수술일	<input type="text"/>
마취 및 수술시간	<input type="text"/> ~ <input type="text"/> (총 <input type="text"/> 분)
마취 종류	<input type="text"/>
마취시 투여약물	<input type="text"/> <div style="border: 1px solid black; padding: 2px; display: inline-block;"> 정신 착취 국소 기타 </div>
골수 채취 부위	<input type="text"/>
채취량(단위:cc) 및 채취 상태	<input type="text"/>
골수 채취시 특별히 관찰된 사항	<input type="text"/>

둘째, 한글로 검색하나 검색결과 대상이 영어문헌에 국한된다는 점이다. 보통 문헌검색 시 국내외문헌이 필요한 경우가 대부분이다. 국내에는 PubMed와 유사한 검색 기능을 구현하고 있는 사이트들이 있다. 이러한 사이트와 PubMed를 적절히 연계시킨다면, 한글단어로 검색해서 국내외의학문헌 검색이 가능할 것으로 예상된다.

셋째, 검색결과가 한글을 지원하지 않는 상황에서 BabelMeSH의 필요성에 대해 의문을 제기할 수 있다. BabelMeSH의 이상적인 형태는 한글 검색에 한글로 검색결과가 나오는 것이다. 적어도 초록이라도 한글로 제공된다면 사용하기에

한결 편리해 지리라 예상된다. 현재 기술로는 영어를 매끄럽게 모국어처럼 번역이 가능하려면 시간이 걸리겠지만, 향후 지속적인 관심과 체계적인 접근을 시도한다면 가능할 것이다.

우리는 영어로 표현하기 힘든 단어에 대한 문헌검색에 한글판 PICO Linguist 와BabelMeSH를 사용할 수 있다. 물론 현재 기능만으로 모든 사용자의 요구를 충족시킬 수는 없지만, 이미 구축된 베타버전을 지속적으로 업그레이드시킴으로써, 사용자가 만족할 수 있는 양질의 검색도구로 발전하리라 예상된다. 사용자들의 꾸준한 관심과 의견 개진이 필요하다.

참고문헌

- 1.<http://askmedline.nlm.nih.gov>
- 2.<http://babelmesh.nlm.nih.gov>
- 3.<http://babelmesh.nlm.nih.gov/pico.php>
4. Nelson SJ, Schopen M, Savage AG, et al. The MeSH translation maintenance system: structure, interface design, and implementation. Medinfo. 2004;11(Pt 1):67-9.
5. Hersh WR, Donohoe LC. SAPPHIRE International: a tool for cross-language information retrieval. Proc AMIA Symp 1998;673-7
6. Richardson WS, Wilson MC, Nishikawa J, Hayward RS. The well-built clinical question: a key to evidence based decisions. ACP J Club 1995;123(3):A12.3.

Acknowledgements

본 연구는 보건복지부 보건의료기술진흥사업 지원에 의하여 이루어진 것임.