의학용어의 효율적인 검색을 위한 검색 브라우저의 요건 분석

류우석*

Requirement Analysis of Search Browser for Efficient Searching of Clinical Terminology

Wooseok Ryu*

Department of Health Care Management, Catholic University of Pusan, Busan 609-757, Korea

요 약

SNOMED CT는 전자의무기록(EMR) 및 전자건강기록(EHR) 시스템에서 표준화된 용어를 사용하여 진료기록을 작성하고 관리하기 위한 표준 의학용어 체계이다. 이 용어체계는 용어의 방대함 및 설계 구조로 인해 용어 체계가 매우 복잡한 특징이 있다. SNOMED CT에서 제공하는 의학 용어를 검색하기 위해 진료 과정에서 사용하는 기존의 브라우저들은 용어체계의 복잡성을 반영하지 못하여 진료기록의 작성 단계에서 그 효용성이 떨어지는 문제가 있다. 본 연구에서는 SNOMED CT 브라우저에 내재된 문제점을 제시하고 용어체계의 분석을 통해 의학 용어를 빠르고 효율적으로 검색하기 위한 검색 브라우저의 요건을 분석하고 개선안을 제시한다.

ABSTRACT

SNOMED CT is a standard clinical terminology to provide a standardized way to record and manage clinical records for EMR or EHR. The structure of SNOMED CT is very complex because of huge expressive power and its internal design mechanism. Although there are some SNOMED CT browsers for browsing and searching SNOMED CT concepts, they are less effective because they do not consider such complexity of SNOMED CT. This paper depicts problems of current SNOMED CT browsers and analyze SNOMED CT dataset. Then, this paper analyze requirements of SNOMED CT browser which improves searching and selecting of appropriate clinical terms.

키워드 : SNOMED CT, 의학 용어, 검색 브라우저, 요구 분석

Key word : SNOMED CT, Clinical Terms, Search Browser, Requirement Analysis

접수일자 : 2014. 10. 01 심사완료일자 : 2014. 10. 31 게재확정일자 : 2014. 11. 05

* **Corresponding Author** Wooseok Ryu (E-mail:wsryu@cup.ac.kr, Tel:+82-51-510-0611)

Department of Health Care Management, Catholic University of Pusan, Busan, 609-757, Korea

Open Access <http://dx.doi.org/10.6109/jkiice.2014.18.11.2691>

print ISSN: 2234-4772 online ISSN: 2288-4165

©This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License(<http://creativecommons.org/licenses/by-nc/3.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.
Copyright © The Korea Institute of Information and Communication Engineering.

I. 서 론

SNOMED CT(Systematized Nomenclature of Medicine-Clinical Terms)는 진료기록을 표준화된 용어를 이용하여 기록 및 관리하기 위해 사용하는 표준 의학용어 체계이다. 이 용어체계는 진료 기록의 관리, 임상 과정의 분석 및 활용을 목적으로 전 세계적으로 도입을 확대하고 있는데, 미국에서는 전자건강기록(Electronic Health Record)을 위한 표준 용어 체계로 활용하고 있으며 국내에서도 전자의무기록(Electronic Medical Record) 시스템으로 그 활용 범위를 넓혀가고 있다.

SNOMED CT는 40만 여 개의 방대한 의학적 의미(컨셉)들을 포함하고 있으며, 그 방대함과 설계 구조 때문에 컨셉 간에 용어가 서로 동일하거나 하나의 컨셉이 여러 용어를 포함하는 특징이 있다[1,2]. 의사, 의무기록사 등 이 용어 체계를 활용하는 사람들이 컨셉들을 모두 기억하고 있지는 못하므로 복잡한 용어체계를 검색하기 위한 용어 검색 프로그램인 SNOMED CT 브라우저[3]를 활용하여 환자의 진단, 수술 등 임상 과정에서 발생하는 상황을 표현하는 컨셉을 기록하고 있다. 하지만 기존에 사용되는 브라우저들은 문자열 검색 수준의 컨셉 검색의 결과로 컨셉에 대한 용어(description), 계층 구조(Concept Hierarchy), 관계(Relationship)들을 제시하는 것에 머무르고 있음에 따라 용어가 동일한 컨셉들을 빠른 시간 내에 검색/선택하여 기록하기가 어려운 한계가 있다. 기존의 브라우저만으로는 축박한 진료기록의 작성 단계에서 임상적 의미를 정확하게 표현하는 용어를 빠르게 선택하기가 어려운 문제가 발생한다.

본 논문에서는 효과적인 의학용어 검색을 위해서 SNOMED CT 검색 브라우저가 가져야할 기본적인 요건[4]을 확장하여 효과적인 용어 검색을 위한 검색 브라우저의 개선 방안 및 이를 위한 요구사항을 제시하고자 한다. 이를 위해 본 논문에서는 먼저 SNOMED CT가 가지는 동일한 용어들의 특성을 분석하고, 이 분석 결과를 바탕으로 검색브라우저가 가져야 할 기능적 요건을 제시한다.

본 논문의구성은 다음과 같다. 2장은 문제분석으로서 먼저 SNOMED CT 용어체계를 간단히 기술하고 기존의 검색브라우저가 가지고 있는 문제점을 제시한다.

3장에서는 SNOMED CT의 계층구조간 거리 분석(Distance Analysis)을 통해 용어체계에 내재된 컨셉의 복잡성을 분석하고 4장에서는 분석 결과를 바탕으로 효과적인 의학용어 검색을 위해 검색 브라우저가 추가적으로 제공해야 할 요건을 제시한다. 마지막으로 5장에서 결론 및 향후 연구를 기술한다.

II. 문제정의

2.1. SNOMED CT 용어체계 개요

SNOMED CT의 컨셉은 “SNOMED Clinical Terms”라는 루트 컨셉부터 “is-a”라는 관계(Relationship)를 통해 하위 컨셉으로 계층 구조를 형성하고 있다. 루트 컨셉과 “is-a”관계로 직접적으로 연결된 최상위 컨셉(Top-Level Concept)은 “Body Structure”, “Clinical Finding”, “Procedure” 등 19개로서 의미적으로 명확하게 구분이 되어 있다. 나머지 컨셉들은 하나 이상의 최상위 컨셉의 자손(descendant)으로서 “is-a” 관계를 통해 계층적으로 연결되어 있다.

SNOMED CT 컨셉 계층구조의 예시는 그림 1과 같다. 그림 1에서는 “disorder”라는 컨셉의 하위 컨셉들의 일부를 도시하고 있는데 그림에서 볼 수 있듯이 SNOMED CT 컨셉의 계층구조는 다중 부모 컨셉을 허용하고 있다. 예를 들어 J로 표기된 “Fracture” 컨셉의 부모 컨셉은 E와 F이며 “bone or arm injury”와 “bone or joint injury” 두 컨셉의 특성을 모두 포함한다.

그림 1. SNOMED CT의 컨셉 계층구조 예시[5]

Fig. 1 An example of SNOMED CT concepts hierarchy

개별 컨셉은 숫자로 구성된 ID라는 식별자를 통해서 서로 구분되며 컨셉의 의미는 여러 개의 의학 용어(Description)들을 통해 부여된다. 이때 하나의 컨셉은 여러 개의 용어를 가지는데, 이는 하나의 컨셉이 한가지의 단어 또는 어구로만 표현하기 어렵기 때문이다. 이 용어는 FSN(Fully Specified Name)과 동의어(Synonym)으로 구분하며 한 컨셉 당 하나의 FSN과 여러 개의 동의어를 포함한다.

2.2. SNOMED CT 브라우저

SNOMED CT 브라우저는 ID 또는 의학 용어를 이용하여 SNOMED CT에 포함된 컨셉들을 검색하기 위한 검색 프로그램이다. 이 브라우저는 대부분의 경우 인터넷에서의 검색 사이트 형태로 구현되어 있다[6,7]. 검색 브라우저들의 기능은 비교적 유사한데 용어의 이름 또는 ID를 이용한 검색을 지원하며 용어들을 통한 해당 컨셉에 대한 설명, 계층구조, “is-a”, “site” 등의 관계 정보들을 브라우저를 통해 조회할 수 있다.

그림 2는 US National Library of Medicine (NLM)에서 제공하는 SNOMED CT 브라우저[6]에서 골반 관절을 의미하는 “joint of pelvis”로 검색을 수행하였을 때의 검색 결과 예시이다. NLM 브라우저를 포함한 모든 SNOMED CT 브라우저는 문자열 매칭을 통해 검색 결과를 제시하는데, 그림에서 나타난 바와 같이 그 결과로서 총 9건이 목록으로 표시된다. 이 9건의 검색결과는 “joint of pelvis”라는 어구를 포함하는 컨셉들의 목록인데 이 컨셉들 모두가 해당 어구를 의학 용어(description)으로 포함하는 경우이다.

대부분의 검색 과정에서 검색 결과는 두 개 이상의 컨셉으로 나타나는데 문제는 이 목록만으로는 의도하는 컨셉을 바로 선택하기가 매우 어렵다는 점이다. 그림 2과 같은 경우 임상 과정에서 정확한 컨셉을 선택하기 위해서는 9개의 컨셉을 모두 클릭해서 상세 정보를 조회해야 하는데 이 과정을 수행하기에는 기록 시간이 사실상 매우 짧으며 그 의미를 구분하기도 어렵다.

SNOMED CT 용어 체계에서는 동일한 의학 용어가 여러 컨셉에서 사용되고, 이때 용어가 서로 동일한 컨셉들은 용어 검색 시에 항상 같이 검색되므로 사용자에게 혼란을 야기한다. 그러므로 검색 결과를 단순한 목록으로 제시하는 것 대신 효과적으로 검색 결과를 제시하는 다른 기법이 필요하다.

그림 2. NLM 브라우저의 검색 결과 예시

Fig. 2 An example of search results

III. SNOMED CT 용어체계의 분석

이 장에서는 검색 결과 목록을 개선하기 위하여 SNOMED CT 데이터 셋에서 컨셉 간 용어의 동일성을 분석한다. 분석 대상 데이터 셋은 2014년 1월 배포된 SNOMED CT International Release Format 2 (Snapshot)[8]이다. 이 데이터 셋에서의 활성 컨셉(Active Concept)의 수는 약 30만개, 활성 용어(Active Description)의 수는 약 100만개 가량이다. 그 중 FSN, 동의어 등의 용어가 서로 완전히 동일한 컨셉의 쌍은 분석 결과 약 12,714쌍이 존재하는 것으로 확인되었다.

본 논문에서는 동일한 용어를 가지는 두 컨셉을 분석하기 위하여 컨셉 간의 거리(distance)를 이용한다. SNOMED CT 용어체계에서 컨셉은 모두 “is-a” 관계를 통해서 계층구조 상에서 서로 연결되어 있으므로, “is-a” 관계를 이용하여 두 컨셉 간 거리의 산출이 가능하다. 이때 거리는 두 컨셉이 서로 부모-자식(Parent-Child)으로 연결되어 있는 경우 1로 계산하였으며, 두 컨셉이 형제(sibling)인 경우에는 거리를 2로 계산하였다. 그리고 두 컨셉 간 도달하는 계층구조 상에서의 경로(path)가 여러 개인 경우에는 거리가 가장 짧은 경로에서의 거리로 산정하였다.

표 1은 용어가 동일한 컨셉의 쌍에 대해서 계층구조 상에서의 두 컨셉 간 거리를 분석하여 제시한 표이다. 분석결과 거리가 5 이하인 쌍의 비율이 58.1%로 절반을 넘는 것을 확인할 수 있으며 표에는 모두 기술되지 않았으나 거리가 1, 즉 부모-자식 관계인 컨셉의 비율이 전체의 50%인 것으로 확인되었다.

표 1. 동일 용어를 가지는 컨셉의 거리 분석

Table. 1 Distance analysis of concepts with an identical description

거리	컨셉 쌍의 수	비율
1-5	7,387	58.1%
6-10	2,104	16.5%
11-15	2,762	21.7%
16-20	430	3.4%
21-25	31	0.2%
합계	12,714	100.0%

표 2는 표 1에 제시된 컨셉들을 두 가지 경우로 구분하여 분석한 표이다. 이는 용어가 동일한 두 컨셉의 최상위 컨셉(Top-level concepts)을 확인하여 최상위 컨셉이 서로 동일한지 아닌지에 따라 구분한 것이다. 최상위 컨셉 내 동일 용어 컨셉의 경우는 용어가 동일한 두 컨셉의 최상위 컨셉이 서로 동일한 경우이고 최상위 컨셉 간 동일 용어는 두 컨셉의 최상위 컨셉이 서로 다른 경우를 의미한다.

표 2를 보면 동일 용어를 가진 두 컨셉이 같은 최상위 컨셉 내에 있는 경우 사실상 거의 모두가 1-5 이내의 거리를 보이고 있음을 확인할 수 있다. 표 2에 모두 기술되지 않았지만 거리 1-3사이에 위치한 컨셉들이 90% 이상을 차지하고 있음이 분석되었다. 특히, 거리가 1인 경우 즉, 부모-자식 관계로 연결되어 있는 경우가 전체의 83% 가량으로 분석됨에 따라 대부분의 동일한 용어가 부모, 자식 간에 발생한다는 것을 확인할 수 있다.

동일 용어를 가진 두 컨셉이 서로 다른 최상위 컨셉인 경우는 거리가 최소 5이며 최대 23인 것으로 나타났으며 그 대부분은 표 2와 같이 거리가 6-15 사이인 것으로 분석되었다. 그러나 이 경우는 두 컨셉의 최상위 컨셉들이 서로 다르므로 용어는 동일하더라도 의미적으로 혼동할 우려는 적은 특징이 있다.

표 2. 최상위 컨셉 내 및 최상위 컨셉 간 동일 용어를 가지는 컨셉의 거리 분석

Table. 2 Distance analysis of concepts with an identical description grouped by intra/inter top-level concepts

거리	최상위 컨셉 내 동일 용어 컨셉 쌍	최상위 컨셉 간 동일 용어 컨셉 쌍
1-5	7,372 (96.6%)	15 (0.3%)
6-10	229 (3.0%)	1,875 (36.9%)
11-15	25 (0.3%)	2,737 (53.8%)
16-20	4 (0.1%)	426 (8.4%)
21-25	0 (0.0%)	31 (0.6%)
합계	7,630 (100%)	5,084 (100%)

IV. 검색 브라우저의 요구사항 분석 및 개선방안

이 장에서는 3장에서 도출된 분석 결과를 바탕으로 효율적인 용어 검색을 위해 브라우저가 제공해야 할 기능적 요건을 개선 방안 형태로 제시한다.

4.1. 검색 인터페이스

검색 브라우저는 기본적으로 ID 또는 텍스트를 통한 컨셉의 검색을 지원하여야 한다. 이때, ID를 이용한 검색은 하나의 컨셉이 검색 결과로 나타나게 되나 용어 텍스트를 통한 검색은 부분 문자열 매칭 검색(substring matching query)으로서 그림 2와 같이 용어가 동일하거나 해당 검색어를 포함하는 컨셉들이 단순 목록의 형태로 나타나게 된다. 본 논문에서는 3장에서 분석한 내용에 따라 검색 결과 목록에 다음과 같은 내용을 추가하도록 제안한다.

1) 컨셉간 관계의 표현

검색 결과로 나타나는 컨셉들은 단순 목록이 아닌 관계에 기반하여 그 관계정보를 함께 표현하여야 한다. 3장 분석 결과에서 도출된 바와 같이 용어가 동일한 대부분의 컨셉 상은 그 거리 값이 작고 특히 거리가 1에서 3인 컨셉이 대부분이다. 이때 두 컨셉들 간의 계층구조 위상(topology)은 부모-자식(Parent-Child), 형제(Sibling), 조상-손자(Anccestor-Descendant) 등이 있는데 이를 간략화 하여 표현하면 컨셉들을 보다 직관적으로 비교할 수 있다. 구현 관점에서는 이를 서브 그래프(sub-graph) 형식으로 간략화 하여 도시할 수 있다.

2) 최상위 컨셉으로 그룹핑

검색 결과에 최상위 컨셉을 함께 표시하고 최상위 컨셉으로 검색 결과를 그룹핑하여 표현하는 것이 필요하다. 이는 3장에서 분석된 바와 같이 용어 중복이 발생하는 두 컨셉의 최상위 컨셉이 다른 경우의 수가 용어 중복이 발생하는 전체 컨셉들 수 대비 40%에 이르기 때문이다. 그러므로 최상위 컨셉을 함께 표현하는 것만으로도 검색 결과를 효율적으로 이해할 수 있는 장점이 발생한다.

3) 매칭 수준의 분리

부분 문자열 매칭 검색은 기본적으로 검색어와 일치되는 컨셉들과 부분 매칭된 컨셉들이 함께 표시된다. 의도한 컨셉을 빠르게 찾기 위해 검색어와 정확히 일치하는 컨셉들과 부분 매칭된 컨셉들을 분리하여 표현하는 것이 필요하다.

4.2. 컨셉 상세정보 조회

기존의 SNOMED CT 검색 브라우저들은 검색 결과에서 하나의 컨셉을 선택하면 그림 3과 같이 컨셉의 상세정보, 용어(Description) 목록, 부모 컨셉, 자식 컨셉, 관계(Relationship), 계층구조상 위치(Tree Position) 등을 표시한다. 이때, SNOMED CT의 컨셉은 그 정의가 명확하게 기술되어 있지 않으므로 용어 목록 및 계층구조 상에서의 상대적 위치를 통해 그 의미를 파악해야 한다. 그러나 컨셉은 부모 노드가 여러 개일 수 있으므로 계층구조의 표현을 그림 3과 같이 Context 1, Context 2등으로 단순 나열하는 경우 그 의미의 파악이 매우 어렵고 불편하다.

본 논문에서 제안하는 개선 방안은 그림 3과 같은 단순 나열식에서 벗어나 계층구조들을 모두 포함하는 하나의 계층구조만을 도시하는 것이다. 이때, 단일화된 계층구조의 표현은 4.1에서 제시된 개선 방안을 응용하여 서브 그래프 형태로 표시하되 복수개의 부모 노드를 한 번에 표현함으로써 사용자의 빠른 이해를 유도한다. 구현 관점에서는 서브 그래프가 복잡해지면 텍스트 형태로 표현하기에는 부적합해지므로 그래픽 형태로 처리하여 도시하는 것이 추가적으로 필요하다.

그림 3. NLM 브라우저의 컨셉 조회 화면

Fig. 3 A screenshot of concept view of NLM browser

그리고 계층구조상 위치를 표시할 때 동일한 용어 및 유사한 용어를 가지는 다른 컨셉을 함께 표시하는 것도 필요하다. 즉, 계층구조 서브 그래프에서 용어가 동일한 컨셉을 노출시키고 해당 컨셉에 대한 링크를 제시함으로써 여러 컨셉들을 동시에 비교하게 하고 이를 통해서 의도하는 컨셉을 보다 빠르게 선택 가능하도록 한다.

V. 결 론

본 논문에서는 SNOMED CT 표준 용어체계를 활용할 때 사용하는 SNOMED CT 검색 브라우저의 문제점을 분석하고, 용어가 동일한 컨셉들 중 원하는 컨셉을 빠르고 효과적으로 선택하기 위한 검색 브라우저의 요건을 분석하고 개선 방안을 제시하였다. 제안된 개선방안을 적용하면 브라우저 사용과정에서 많은 시간을 소모하는 유사 용어 비교 시간을 획기적으로 줄임으로써 진료기록 과정에서 환자의 상태를 가장 정확하게 표현

하는 컨셉을 빠르게 선택할 수 있으며 결과적으로 진료 기록의 효율성을 제고할 수 있다.

향후 연구로는 본 논문에서 새롭게 제시된 동일 용어 문제를 다각도로 분석하고, 제시한 요구조건 및 개선방안을 바탕으로 새로운 SNOMED CT 검색 브라우저를 설계 구현하는 것이다.

REFERENCES

- [1] Alecu I., Bousquet C., Jaulent M.C., "A case report: using SNOMED CT for grouping Adverse Drug Reactions Terms", BMC Medical Informatics and Decision Making, 2008.
- [2] S. Lusignan, T. Chan, S. Jones, "Large complex terminologies: more coding choice, but harder to find data - reflections on introduction of SNOMED CT (Systematized Nomenclature of Medicine - Clinical Terms) as an NHS standard," *Informatics in primary care*, vol. 19, no. 3, pp. 3-5, Mar. 2011.
- [3] D. Lee, R. Cornet, F. Lau, N. Keizer, "A survey of SNOMED CT implementations", *Journal of Biomedical Informatics*, Vol. 46, pp. 87-96, 2013.
- [4] W. Ryu, "A Requirement of a Search Browser for Effective Searching of Clinical Terminology," in *Proceedings of the Korean Institute of Information and Communication Sciences*, Vol. 18, pp. 416-417, 2014
- [5] IHTSDO, "SNOMED CT Technical Implementation Guide", www.snomed.org, 2014.
- [6] U.S. National Library of Medicine, "UMLS SNOMED CT Browser", <http://uts.nlm.nih.gov/snomedctBrowser.html>
- [7] International Health Terminology Standards Development Organisation, "The IHTSDO SNOMED CT Browser," <http://browser.ihtsdotools.org>
- [8] U.S. National Library of Medicine, "SNOMED CT Release Files", <http://www.nlm.nih.gov/research/umls>

류우석(Wooseok Ryu)

2013년 3월-현재 부산가톨릭대학교 병원경영학과 조교수
2012년 3월-2013년 2월 부산대학교 U-port정보기술산학공동사업단 박사후연구원
2012년 2월 부산대학교 컴퓨터공학과 공학박사
※ 관심분야: 의학용어, U-Health, 의료정보, 빅 데이터