

# 표준 의학용어 체계에서의 효과적인 용어 검색 방안

류우석\*

\*부산가톨릭대학교

Effective Searching of Clinical Terms from Standard Clinical Terminology

Wooseok Ryu\*

\*Catholic University of Pusan

E-mail : wsryu@cup.ac.kr

## 요 약

SNOMED CT는 임상 과정에서의 진료 기록을 전산화하고 이를 효율적으로 관리 및 활용하기 위한 종합적 표준 임상 의학용어 체계이다. 그러나 이는 표현 가능한 용어의 종류가 지나치게 많음에 따라 환자 상태를 표현하는 가장 적합한 용어를 짧은 진료 시간 내에 선택하여 기록하기가 어렵고, 또한 한가지 임상적 의미가 다양한 용어 및 그 조합으로 표현될 수 있음에 따라 의무기록의 작성시 그 의미가 왜곡되거나 잘못 해석될 여지가 발생한다. 본 연구에서는 이러한 문제점들을 분석하고 이를 해결하기 위한 중복 용어 선택 기법을 제시한다.

## ABSTRACT

SNOMED CT is a standard clinical terminology which is to efficiently record, manage and utilize clinical records during clinical processes. However, huge expressive power of SNOMED CT makes it difficult to select appropriate terms during short consultation hours. In addition, true meaning of a written record using the terminology may be misunderstood or even distorted since one clinical meaning could be expressed in a variety of ways. This paper analyzes such known problems in a database point of view, and presents effective selection methods of similar terminologies to mitigate the problem.

## 키워드

SNOMED CT, Clinical Terms, Ambiguity, Terminology Selection

## 1. 서 론

SNOMED CT(Systematized Nomenclature of Medicine-Clinical Terms)는 종합적 임상 의학용어 체계이다. 해외에서는 전자건강기록(electronic health record, EHR)을 위한 표준 용어 체계로 그 활용 범위를 넓혀가고 있으며, 국내에서도 진료기록을 체계적 저장, 분석, 관리하려는 목적에 따라 표준화된 용어를 이용한 전자의무기록(electronic medical record, EMR)을 작성하기 위해 SNOMED CT 용어체계의 도입을 확대하고자 노력하고 있다.

SNOMED CT가 제공하는 여러 장점에도 불구하고

하고 몇가지 문제로 인해 도입의 어려움을 겪고 있다. 이는 첫째, 표현 가능한 용어의 종류가 지나치게 많음에 따라 환자 상태를 표현하는 가장 적합한 용어를 짧은 진료 시간 내에 선택하여 기록하기가 어렵고, 둘째, 한 가지 임상적 의미가 다양한 용어 및 그 조합으로 표현될 수 있음에 따라 의무기록의 작성시 그 의미가 왜곡되거나 잘못 해석될 여지가 발생하는 점이다. 진료기록의 정확성을 확보하기 위해서는 SNOMED CT에 내재된 용어의 모호성(ambiguity) 문제[1]의 해결이 선행되어야 한다.

본 논문에서는 SNOMED CT 용어체계에서의 모호성 문제를 분석하고, 진료 기록을 작성할 때

적합한 의학용어를 선택하는 과정에서 발생하는 모호성을 제거 또는 완화하기 위한 효과적인 용어 검색/선택 기법을 제시한다.

## II. SNOMED CT 개요

SNOMED CT는 약 40만개의 의학적 의미(Concept, 이하 컨셉)와 각 컨셉들의 속성 및 그들간의 관계로 구성되어 있다. 각 컨셉들은 "SNOMED Clinical Terms"라는 루트 컨셉부터 하위 컨셉으로 연결되는 계층구조로 이루어져 있고, 계층구조 간에는 "is a"라는 관계(Relationship)가 정의되어 있다. 그리고, 각 컨셉은 "is a" 관계 이외에도 "finding site", "associated with", "part of", "severity" 등 60가지 이상의 관계에 따라 서로 유기적으로 연결되어 있는 특징을 가진다. 그림 1은 SNOMED CT 컨셉 중 장애를 의미하는 "disorder" 컨셉에 대한 하위 컨셉들의 "is a" 계층구조를 도시하고 있다.

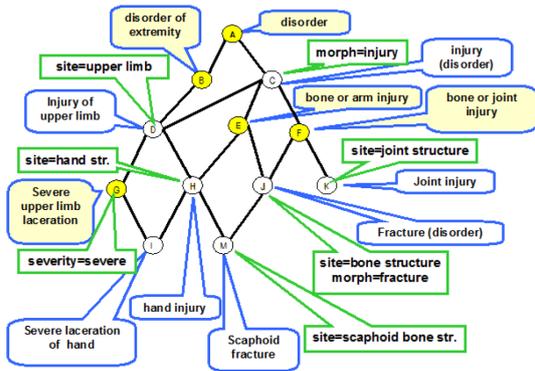


그림 1. SNOMED CT의 컨셉 계층구조 예시[2]

각 컨셉은 최대 18자리 숫자로 구성된 ID로 구분되며 컨셉의 의미를 표현하는 여러 개의 문자열을 포함한다. 이를 Description이라고 하는데, 이는 하나의 의학적 의미가 한가지의 단어(또는 어구)로만 표현되지 못하는 문제점을 해결하기 위한 것으로 각 컨셉당 하나의 FSN(fully specified Name)과 여러 개의 동의어(Synonym)를 Description을 통해 제시하고 있다.

SNOMED CT의 컨셉들을 검색하기 위한 다양한 인터페이스들이 SNOMED CT 브라우저라는 이름으로 제시되어 있다[3][4]. 그림 2은 NLM(US National Library of Medicine)에서 제공하는 SNOMED CT 브라우저[4]의 예시인데, 이를 브라우저를 통해 ID 또는 텍스트로 컨셉을 검색하고 해당 컨셉에 대한 description, 계층 구조(Concept Hierarchy), 관계들을 확인할 수 있다.

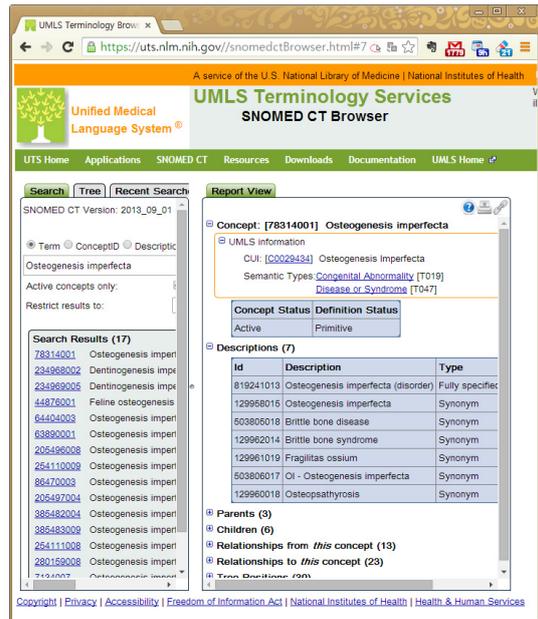


그림 2. UMLS SNOMED CT Browser

## III. 용어의 모호성 분석

SNOMED CT 의학용어의 모호성은 하나의 의학용어가 서로 다른 컨셉에서 동일하게 사용되는 것에서 초래한다. 그림 3을 예를 들어 보면 골형성부전증을 표현하는 "Osteogenesis imperfecta"는 10개의 컨셉을 표현하는데 동일하게 사용되고 있다. 즉, "Osteogenesis imperfecta"라는 용어를 진료 과정에서 사용하고자 할 때 어떤 컨셉을 선택해야 하는지에 대한 모호성이 발생한다. 또한, 동일한 단어가 아니라도 단어 매칭 검색(substring match)을 하면 이보다 훨씬 많은 수의 컨셉의 중복이 발생한다. 골형성부전증의 경우 정확히 일치하는 컨셉은 10건이지만 단어 매칭 검색을 하면 40건 이상의 중복이 발생하는 것이 확인되었다.

conceptid	concept Type	conceptText
78314001	Synonym	Osteogenesis imperfecta (disorder)
157008000	Synonym	(Bowleg) or (Ehlers-Danlos syndrome) or (...)
205492005	Synonym	(Osteogenesis imperfecta) or (Vrolik's dis...)
205493000	Synonym	Osteogenesis imperfecta
254105005	Synonym	Osteogenesis imperfecta
254106006	Synonym	Osteogenesis imperfecta
254107002	Synonym	Osteogenesis imperfecta
254108007	Synonym	Osteogenesis imperfecta
254109004	Synonym	Osteogenesis imperfecta
268352002	Synonym	(Bowleg) or (Ehlers-Danlos syndrome) or (...)

그림 3. 골형성부전증(Osteogenesis imperfecta)으로 표현된 컨셉의 목록

표 1은 19개의 최상위 컨셉(루트 컨셉에 직접 연결된 하위 컨셉) 중 주요 컨셉에 대해 하위 컨셉의 개수와 그 중 용어의 중복이 발생하는 컨셉의 수를 분석한 표이다. 표에 제시되어 있는 바와 같이 신체구조(Body Structure)의 경우 절반에 가까운 컨셉이 용어의 중복성 문제를 내포하고 있으며 의약품에서도 적지않은 비율의 중복성 문제를 내포하고 있다. 이에 따라 빠른 시간 내에 적합한 컨셉을 검색하여 선택하기가 매우 어려우며, 작성된 진료기록 또한 그 의도를 명확하게 이해하기 어려운 문제가 발생한다.

표 1. 최상위 컨셉별 컨셉의 중복 분석

최상위 컨셉	컨셉 수	중복 컨셉 수	비율
Body structure	30,588	12,787	41.8%
Clinical finding	99,811	2,171	2.2%
Organism	33,156	600	1.8%
Pharmaceutical / biologic product	16,769	3,305	19.7%
Procedure	53,628	898	1.7%
Qualifier value	8,943	429	4.8%
Special concept	648	5	0.8%
Specimen	1,439	8	0.6%
Substance	23,844	3,766	15.8%

#### IV. 용어 검색 및 선택 방법

2장에서 소개한 바와 같이 SNOMED CT의 컨셉들은 계층구조적이며, 동일 또는 유사한 명칭의 컨셉의 의미를 명확히 이해하기 위해서는 계층구조상에서 해당 컨셉들 간의 상호관계를 정량화하여 표현하는 것이 필요하다. SNOMED CT의 컨셉을 검색하고 조회하기 위한 다양한 브라우저들은 단일 컨셉에 대한 상세 내역들만 검색 가능하며, 여러 개의 컨셉을 동시에 비교하는 방법은 제시되어 있지 않다.

용어 검색 결과 두 개 이상의 컨셉들이 도출된 경우 이들 중 가장 적합한 것을 선택하는 건 진료기록을 작성하는 진료의 또는 의무기록사가 담당하며 이들은 각각 진료과정, 코딩 과정에서 해당 업무를 수행한다. 작성자가 검색을 통해 적절한 용어를 선택하는 방법으로써 각 컨셉들의 description, 컨셉 계층구조, 관계와 같은 기본 정보 모두를 비교하는 것은 용어체계의 복잡성으로 인해 오히려 올바른 컨셉의 선택을 저해한다. 본 논문에서는 위의 기본 정보들을 제시하는 대신 다음과 같은 비교 정보를 제시할 것을 제안한다.

- Number of Hops : 계층구조 상에서 두 컨셉의 거리. 이는 한 컨셉에서 다른 컨셉으로 도달하기 위해 몇 번의 "is a" 관계를 따라 컨셉들을 방문해야 하는지를 의미한다.
- Common Ancestor : 두 컨셉에서 중복되는 상위 컨셉들 중 Number of Hops이 가장 적은 컨셉.
- Differentiated Attributes : Common Ancestor에서 각 컨셉들까지의 하위 컨셉들 및 그들의 속성(관계, Description 등)들

동시에 검색되는 두 개 이상의 컨셉들의 차이점을 명확하게 구분하기 위해서는 위에서 제시된 정보들을 통해 계층구조 내에서 서로 차별되는 요소들만을 도출하고 이를 비교할 수 있도록 한다. 이를 통해 짧은 시간 내에 보다 더 목적에 부합하는 용어를 선택하도록 함으로써 SNOMED CT의 모호성 문제를 해결하고 용어체계의 활용도를 보다 높일 수 있다.

#### V. 결 론

본 논문에서는 SNOMED CT 표준 용어체계의 사용시 발생하는 용어 선택의 모호성 문제를 분석하고 효과적인 용어 검색 및 선택을 위한 유사 컨셉간 비교 방법을 제시하였다. 이를 통해 기존의 free-text 기반의 진료 기록을 표준화된 의학용어 기반의 진료 기록으로 보다 수월하고 효과적으로 변환 또는 작성할 수 있다. 본 연구에서는 모호성 문제를 동일/유사 이름의 컨셉으로 한정하였지만 향후 연구로 이를 더욱 일반화하여 분석하고 더욱 자동화된 선택을 가능하게 하는 검색 및 선택 기법의 제시하는 것이 필요하다.

#### 참고문헌

- [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] IHTSDO, "SNOMED CT Technical Implementation Guide", www.snomed.org, 2014.
- [3] Lee D., Cornet R., Lau F., Keizer N., "A survey of SNOMED CT implementations", Journal of Biomedical Informatics, Vol 46, pp. 87-96, 2013.
- [4] IHTSDO, "UMLS SNOMED CT Browser", <http://uts.nlm.nih.gov/snomedctBrowser.html>