

표준 의학용어 체계에서의 효과적인 용어 비교 검색 기법

류우석*

Effective Scheme for Comparative Search of Clinical Terms from Standard Clinical Terminology

Wooseok Ryu*

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

요 약

종합적 표준 임상 의학용어 체계인 SNOMED CT는 용어의 방대함 및 구조의 복잡성 때문에 환자 상태를 표현하는 가장 적합한 용어를 짧은 진료 시간 내에 선택하기가 어려운 용어 선택의 모호성 문제를 내포하고 있다. 본 논문에서는 용어 검색 과정에서 발생하는 동일하거나 유사한 용어를 분석하고 이를 효과적으로 구분하기 위한 용어 비교 검색 기법을 제시한다. 제안하는 기법은 용어의 계층 구조 분석을 통해 용어 간 “is-not-a” 관계를 새로 정의함으로써 두 비교 대상 용어의 차이점을 명확하게 하는 특징이 있다. 이를 통해 진료 과정에서 유사한 이름을 가진 용어들 중 환자의 상태를 가장 적합하게 표현하는 용어를 빠르게 선택함으로써 SNOMED CT의 활용성을 개선한다.

ABSTRACT

SNOMED CT, which is a standard clinical terminology, imposes an ambiguity problem of terminology selections caused by its huge expressive power and structural complexity. It is very difficult to distinguish similar terms and to select an appropriate term among them within short consultation hours. This paper analyzes the ambiguity problem and proposes a novel scheme for comparative search of similar terms. The proposed scheme provides a differential view of similar terms by defining a “is-not-a” relationship based on the hierarchical structure of the concepts. The result of this work improves the utilization of SNOMED CT such that medical officers can efficiently select an appropriate term among similar terms which represents patient’s status adequately.

키워드 : SNOMED CT, 의학용어, 모호성, 용어 비교, 관계 정의

Key word : SNOMED CT, Clinical Terms, Ambiguity, Terminology Comparison, Defining Relationship

접수일자 : 2014. 09. 22 심사완료일자 : 2014. 10. 23 게재확정일자 : 2014. 11. 06

* **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.2015.19.3.537>

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 임상 과정에서 발생하는 각종 진료기록을 전산화하기 위한 종합적 임상 의학용어 체계이다. 이는 전자건강기록(electronic health record, EHR)을 위한 표준 용어 체계로서 미국 등 해외에서 활발히 그 활용 범위를 넓혀가고 있다[1]. 국내의 경우에도 의무기록의 체계적 저장, 관리 및 연구에 활용하기 위한 목적으로 전자의무기록(Electronic Medical Record, EMR) 및 EHR의 작성에 SNOMED CT 용어체계를 도입하고자 노력하고 있다[2].

SNOMED CT는 다른 용어체계와는 비교하기 어려울 정도로 용어의 방대함, 용어 활용의 확장성 및 유연성을 가지고 있다. 하지만, 용어의 방대함에 따른 용어 체계의 복잡성이 임상에서의 적극적인 활용 측면에서 문제를 야기하고 있는데, 직접 사용 가능한 의학용어가 40만개에 달하고 있으며 이를 조합하면 사실상 무한대에 가까운 표현이 가능하기 때문이다. SNOMED CT가 가지는 또 다른 특성으로 환자의 상태를 기술하기 위한 한가지 임상적 의미가 SNOMED CT에서 제시하는 다양한 용어들 및 그 조합으로 표현될 수 있으며, 하나 또는 유사한 용어가 여러 가지 임상적 의미에 중복해서 사용되기도 한다.

이러한 특성은 동일한 증상을 여러 방식으로 기술하거나 서로 다른 증상을 동일한 용어로 표현하게 하므로 진료기록의 작성 단계에서 환자 상태를 기술하는 가장 적합한 용어를 짧은 진료 시간 내에 선택하기가 어려운 문제가 발생한다. 작성 단계에서의 용어 선택의 모호성(ambiguity) 문제[3-5]는 진료기록 뿐만 아니라, 이후 분석 단계에서 그 의미가 왜곡되거나 잘못 해석될 여지도 발생한다. 용어체계의 원활한 활용을 위해서는 SNOMED CT 용어체계의 복잡성에 따른 용어의 모호성 문제의 해결이 필요하다. 본 논문은 선행 연구[5]를 확장하여 먼저 SNOMED CT 용어체계의 복잡성에 따른 용어 선택의 모호성 문제를 분석하고, 적합한 의학용어를 검색하고 선택하는 과정에서 발생하는 모호성을 제거 또는 완화하기 위한 효과적인 용어 비교 기법을 제시한다.

논문의 구성은 다음과 같다. 2장에서는 SNOMED CT 용어를 검색하기 위한 브라우저 기술과 용어의 복

잡성을 해결하기 위한 관련 연구를 제시하고, 3장에서는 SNOMED CT 용어체계의 특성 및 용어의 중복된 표현으로 인한 용어의 복잡성 문제를 분석한다. 4장에서는 이를 해결하기 위한 유사 용어 간 효과적인 비교 검색 기법을 제시한다. 마지막으로 6장에서 결론 및 향후 연구를 기술한다.

II. 관련기술

SNOMED CT 브라우저는 SNOMED CT에 포함된 의학용어들을 검색하기 위한 검색 프로그램이다[6]. 대표적인 브라우저인 CliniClue Xplore[7]는 PC에 설치하여 사용하는 용어 검색 브라우저로서 용어의 이름 또는 ID를 이용한 검색을 지원하며 해당 용어에 대한 설명, 계층구조, 관련 용어와의 관계들을 조회할 수 있다. 그림 1은 NLM(US National Library of Medicine)에서 제공하는 SNOMED CT 브라우저[8]의 실행 화면인데, 이는 웹에서 구동되며 기능적으로는 CliniClue Xplore와 유사한 특징이 있다. 이들 브라우저는 개별 용어의 자세한 설명을 비교적 쉬운 인터페이스를 통해 접근할 수 있으나 동일하거나 유사한 용어들의 직접적인 비교는 지원하지 않고 있다[6-8].

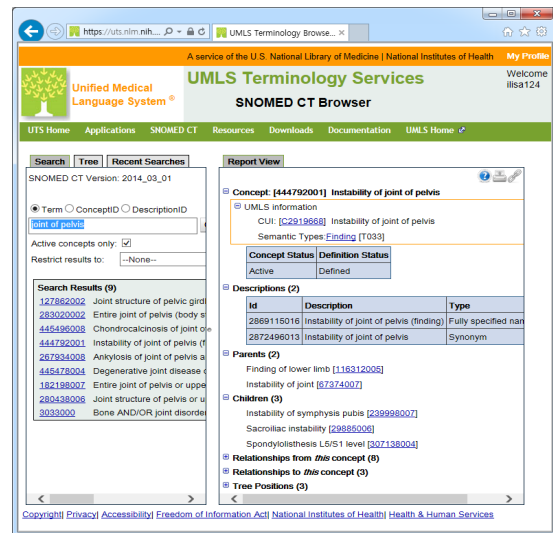


그림 1. NLM SNOMED CT 브라우저
Fig. 1 NLM SNOMED CT Browser

SNOMED CT 용어 체계의 복잡성을 해결하기 위한 연구로서 BLUSNO[9]라는 용어체계의 간략화 기법이 제안되었다. 이 연구는 복잡한 계층구조를 단순화시킨 추상 네트워크를 제안함으로써 이를 통해 복잡한 SNOMED CT 용어체계를 한눈에 파악할 수 있도록 하고 원하는 용어를 빠르게 접근 가능하도록 하는 특징이 있다. 또한, 추상 네트워크에 가변성(scalability)을 부여하여 네트워크 노드 내 컨셉의 수가 많은 경우 이를 더욱 자세하게 조회하는 기법[10]도 제시되었다. 그리고, 어휘 분석을 통해 용어의 비일관성 문제를 도출하고 용어 정의의 비일관성을 개선하는 연구가 진행되었다[11]. 하지만, 이들 연구는 용어 체계의 복잡성을 야기하는 주요 문제인 중복성을 해결하지는 못하고 있으며, 유사한 용어를 구분하는 것은 여전히 사용자에게 맡겨야 하는 한계점을 가진다.

III. SNOMED CT 용어체계

3.1. SNOMED CT 용어 체계 개요

SNOMED CT는 컨셉(Concept)이라고 부르는 약 40만 개의 의학적 의미를 포함하고 있다. 컨셉은 최대 18자리 숫자로 구성된 ID를 통해 서로 구분되며 컨셉의 의미를 표현하는 여러 개의 설명(Description)들을 포함한다. 이는 하나의 컨셉이 한가지의 단어 또는 어구로만 표현하기 어려운 문제를 해결하기 위한 방법으로 그림 2와 같이 하나의 컨셉에 FSN(Fully Specified Name)이라는 대표 설명과 여러 개의 동의어(Synonym)를 포함하는 특성이 있다[12].

이 컨셉들은 “SNOMED Clinical Terms”라는 루트 컨셉부터 “is-a”라는 관계(Relationship)를 통해 하위 컨셉으로 계층 구조를 형성하고 있다. 루트 컨셉과 “is-a”관계로 연결된 최상위 컨셉은 “Body Structure”, “Clinical Finding” 등 19개이다. 이 계층구조는 다중 부모 컨셉을 허용하는 특징이 있는데, 이는 기존의 의학 용어 체계와 다른 SNOMED CT의 고유한 특징으로서 용어 체계의 복잡성을 야기하는 원인이 된다. 그리고, 각 컨셉들은 “is-a”관계 뿐만 아니라 “finding site”, “part of”, “severity” 등 60가지 이상의 관계로 의미에 따라 서로 유기적으로 연결되어 있다.

Concept: [127278005] Injury of upper extremity

UMLS information
 CUI: [C0003794] Arm Injuries
 Semantic Types: Injury or Poisoning [T037]

Concept Status	Definition Status
Active	Defined

Descriptions (4)

Id	Description	Type	Status
731349018	Injury of upper extremity (disorder)	Fully specified name	Active
566013	Injury of upper extremity	Synonym	Active
474090013	Arm injury	Synonym	Active
474089016	Injury of upper limb	Synonym	Active

그림 2. 컨셉의 설명(Description) 예시
 Fig. 2 An example of descriptions of a concept

3.2. 용어체계의 중복성 분석

SNOMED CT 의학용어의 중복성은 하나의 설명(Description)이 서로 다른 컨셉에서 동일하거나 유사하게 사용되는 것에서 초래한다. 그림 3을 예를 들어 보면 해부학적 인체구조 중 골반 관절을 의미하는 “joint of pelvis”는 서로 다른 두 컨셉에 모두 포함되어 있다. 이 경우 “joint of pelvis”를 진료 과정에서 기술하려고 할 때 둘 중 어떤 컨셉을 선택해야 할 지에 대한 선택의 모호성이 발생한다. 본 예시는 설명이 완전히 동일한 예시이나 NLM SNOMED CT Browser에서 “joint of pelvis”를 검색하면 부분 단어 검색(substring match)을 통해 총 9개의 컨셉이 검색된다.

Concept: [127862002] Joint structure of pelvic girdle

Descriptions (3)

Id	Description	Type
2987343016	Joint structure of pelvic girdle (body structure)	Fully specified name
2987390019	Joint structure of pelvic girdle	Synonym
1149013	Joint of pelvis	Synonym

Concept: [283020002] Entire joint of pelvis

Descriptions (3)

Id	Description	Type	Status
676741013	Entire joint of pelvis (body structure)	Fully specified name	Active
421481012	Entire joint of pelvis	Synonym	Active
421482017	Joint of pelvis	Synonym	Active

그림 3. 동일한 설명(Description)을 가진 컨셉의 예시
 Fig. 3 An example of two concepts with the same descriptions

표 1은 19개의 최상위 컨셉 중 주요 컨셉을 대상으로 하위 컨셉의 개수 및 설명의 중복이 발생하는 컨셉의 개수 및 비율을 정리한 표이다. 신체 구조(Body Structure)

의 경우 절반에 가까운 컨셉들이 서로 중복되며, 의약품 및 의약 물질에서도 적지 않은 비율의 컨셉이 서로 중복된다. 유사하거나 중복 표현된 컨셉으로 인해 빠른 시간 내에 적합한 컨셉을 검색하기가 매우 어려운 모호성이 발생하므로 중복된 컨셉의 의미를 명확하게 구분하기 위한 컨셉의 비교 기법이 필요하다.

표 1. 최상위 컨셉별 컨셉의 중복 비율 [5]
Table. 1 Ratio of duplicated concepts per top-level concept

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

IV. 유사 용어 비교 검색 기법

3장에서 소개한 바와 같이 SNOMED CT의 컨셉들 중 동일 또는 유사한 명칭의 컨셉의 의미를 명확히 이해하기 위해서는 계층구조상에서 해당 컨셉들 간의 상호관계를 정량화하여 표현하는 것이 필요하다. SNOMED CT의 컨셉을 검색하고 조회하는 역할을 수행하는 SNOMED CT 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 컨셉의 설명, 컨셉의 계층구조, 관계와 같은 기본 정보 모두를 비교하는 것은 용어체계의 복잡성으로 인해 오히려 올바른 컨셉의 선택을 저해한다. 본 논문에서는 위의 정보들을 모두 제시하는 대신 “is-a” 관계를

바탕으로 최소한의 비교 정보를 제시하기 위해 계층 구조를 아래와 같이 분석한다.

- Ancestor (C_A) : 특정 컨셉에서 “is-a”관계로 이어진 부모 및 조상 컨셉들의 집합
- Common Ancestor (C_{CA}) : 두 컨셉 각각의 조상 컨셉 집합들에 대한 교집합
- Distance (D) : 계층구조 상에서 두 컨셉의 거리. 이는 한 컨셉에서 다른 컨셉으로 도달하기 위해 몇 번의 “is-a” 관계를 따라 컨셉들을 방문해야 하는지를 의미함. 이때, 두 컨셉간 도달하는 경로가 두 가지 이상인 경우 짧은 경로의 거리로 계산
- Nearest Common Ancestor (C_{NCA}) : 두 컨셉의 C_{CA} 중 컨셉 각각에 대한 Distance의 합이 가장 작은 컨셉

그림 4는 SNOMED CT 컨셉 중 장애를 의미하는 “disorder” 컨셉에 대한 하위 컨셉들의 “is-a” 계층구조에 대한 예시를 도시하고 있다. “disorder” 컨셉은 최상위 컨셉인 “Clinical Finding”의 자식 컨셉으로 “disorder” 컨셉의 자식 컨셉의 수는 실제로는 55개이나 그림 4에는 그 일부만 도시되어 있다.

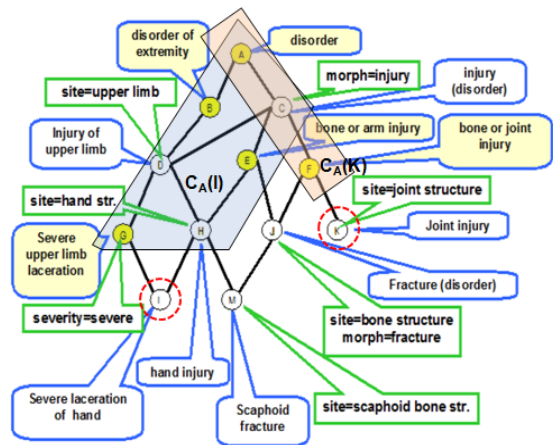


그림 4. 두 컨셉에 대한 Nearest Common Ancestor의 예시
Fig. 4 An example of the NCAs of two concepts

그림 4를 예를 들어 보면 컨셉 I는 “Severe laceration of hand”로서 $C_A(I)$ 는 {A, B, C, D, E, G, H}이다. 그리고 컨셉 K는 “Joint Injury”로서 $C_A(K)$ 는 {A, C, F}이다. 이때 $C_{CA}(I,K)$ 는 $C_A(I)$ 와 $C_A(K)$ 의 교집합인 {A, C}이

다. 이때, $C_{NCA}(I, K)$ 는 컨셉 I, K에 대해 컨셉 A, C의 Distance를 비교해 보면 {C}임을 계산할 수 있다.

“is-a” 관계를 바탕으로 두 컨셉을 비교할 때 C_{CA} 중 C_{NCA} 를 제외한 나머지 조상노드들은 서로 동일하므로 이들을 통해서 두 컨셉을 비교하는 것은 무의미하다. 그러므로, C_{NCA} 의 하위 노드들을 이용하여 두 컨셉을 비교하는 것이 필요하다. 이를 위해서 아래와 같이 Representative Ancestor를 정의한다.

- Representative Ancestor (C_{RA}) : 두 컨셉의 C_{NCA} 의 자식 컨셉들 중 한 컨셉의 C_A 에만 포함된 컨셉의 집합.

그림 5를 예를 들어 보면 두 컨셉 I, K의 C_{NCA} 인 $C_{NCA}(I, K)$ 는 {C}이고 이때 $C_{RA}(I, K)$ 는 $C_A(I)$ 중 컨셉 C의 하위 컨셉인 {D, E}가 된다. 그리고 그 반대의 경우인 $C_{RA}(K, I)$ 는 $C_A(K)$ 중 컨셉 C의 하위 컨셉인 {F}가 된다.

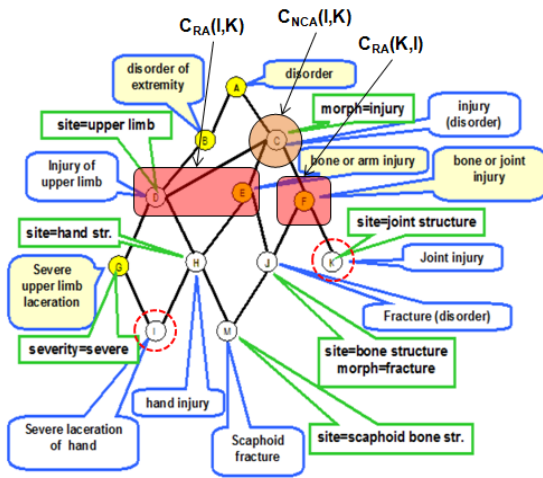


그림 5. Representative Ancestor를 통한 두 컨셉의 비교
Fig. 5 Comparison of two concepts using the representative ancestor

C_{RA} 를 이용하면 I, K, $C_{RA}(I, K) = \{D, E\}$, $C_{RA}(K, I) = \{F\}$ 간에 아래와 같은 관계가 성립한다.

1) 컨셉 D, E는 컨셉 I의 조상이지만 컨셉 K의 조상은 아니다. 즉, 컨셉 K는 컨셉 D와 E의 특성을 갖고 있지 않다.

2) 컨셉 F는 컨셉 K의 조상이지만 컨셉 I의 조상은 아니다. 즉, 컨셉 I는 컨셉 F의 특성을 갖고 있지 않다.

즉, 컨셉 K와 컨셉 D, E는 서로 연관관계가 없으며, 컨셉 I와 컨셉 F와도 아무런 연관관계를 가지지 않는다. 본 논문에서는 이 관계를 “is-not-a”관계로 정의하여 이를 두 컨셉의 차이점을 비교하기 위한 최소한의 정보로 활용한다. “is-not-a” 관계를 보다 명확하게 정의하면 다음과 같다.

- $C_{INA}(A, B)$: $C_{NCA}(A, B)$ 의 자식 컨셉들 중 A의 조상 컨셉이 아닌 컨셉의 집합, $C_{INA}(A, B) = C_{RA}(B, A)$.

위 예시에서 보면 $C_{INA}(I, K) = \{F\}$ 이고 $C_{INA}(K, I) = \{D, E\}$ 인 것을 확인할 수 있다. 즉, “Severe laceration of hand”는 “bone or joint injury”가 아니라는 것을 알 수 있다. 동일하거나 유사한 두 컨셉들의 차이점을 명확하게 구분하기 위해서는 “is-a” 관계로 연결된 모든 조상노드들을 제시하여 비교하도록 하는 대신에 C_{RA} 및 C_{INA} 만을 이용하여 계층구조 내에서 서로 차별되는 요소들만을 도출하고 제시함으로써 두 컨셉의 차이를 보다 효과적으로 비교 구분할 수 있다. 이를 통해 짧은 시간 내에 보다 더 목적에 부합하는 용어를 선택할 수 있다. 이 비교 정보의 제시를 통해 SNOMED CT의 모호성 문제를 완화하고 용어체계의 활용도를 보다 높일 수 있다.

V. 결론

SNOMED CT 표준 용어체계를 이용한 진료기록의 작성시 용어 체계의 복잡성으로 인해 의도하는 용어를 빠르게 찾기 어려운 문제가 있다. 본 논문에서는 SNOMED CT 표준 용어체계의 복잡성으로 인한 용어 선택의 모호성 문제를 분석하고, 유사하거나 동일한 이름을 가진 두 컨셉들 간의 효과적인 비교 검색을 위해 “is-not-a”관계를 정의하였다. 그리고, “is-not-a”관계에 기반한 대표 조상(C_{RA}) 및 차이 조상(C_{INA})을 제시함으로써, 유사한 이름을 가진 컨셉들을 효과적으로 비교하고 의도하는 컨셉을 보다 빠르게 선택하기 위한 기법을 제시하였다. 진료 과정에 본 연구 결과를 도입함으로써

기존에 자연어(free-text)로 작성된 진료기록을 보다 효과적으로 표준 의학용어로 변환할 수 있으며, 또한 진료 기록 과정에서 용어 선택의 모호성에 따른 불필요한 시간 지연을 효과적으로 줄일 수 있다.

향후 연구로서 본 논문에서 제안한 “is-not-a”관계를 지원하는 SNOMED CT 브라우저를 설계 구현하고 이를 진료기록 작성단계에 적용함으로써 임상 현장에서의 유용성을 검증하는 것이 필요하다.

감사의 글

이 논문은 2014년도 부산가톨릭대학교 교내연구비에 의하여 연구되었음

REFERENCES

- [1] K. Nonnelly, “SNOMED-CT: The Advanced Terminology and Coding System for eHealth,” *Stud Health Technol Inform*, Vol. 212, pp.279-290, 2006.
- [2] R. W. Park, “A clinical research strategy using longitudinal observational data in the post-electronic health records era,” *J Korean Med Assoc*, Vol. 55, No. 8, pp.711-719, 2012.
- [3] I. Alecu, C. Bousquet, and M. C. Jaulent, “A case report: using SNOMED CT for grouping Adverse Drug Reactions Terms,” *BMC Medical Informatics and Decision Making*, 8(Suppl 1), 2008.
- [4] S. Lusignan, T. Chan, and 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.
- [5] W. Ryu, “Effective Searching of Clinical Terms from Standard Clinical Terminology,” in *Proceedings of the Korean Institute of Information and Communication Sciences Conference*, vol. 18, pp. 323-325, 2014.
- [6] D. Lee, R. Comet, F. Lau, and N. Keizer, “A survey of SNOMED CT implementations,” *Journal of Biomedical Informatics*, Vol. 46, pp. 87-96, 2013.
- [7] The Clinical Information Consultancy Ltd.. CliniClue Xplore [Internet]. Available: <http://www.cliniclue.com>.
- [8] IHTSDO. UMLS SNOMED CT Browser [Internet]. Available: <http://uts.nlm.nih.gov/snomedctBrowser.html>.
- [9] J. Geller, C. Ochs, Y. Perl, and J. Xu, “New abstraction networks and a new visualization tool in support of auditing the SNOMED CT content,” in *AMIA Annual Symposium Proceedings*, pp. 237-246, 2012.
- [10] C. Ochs et al, “Scalability of abstraction-network-based quality assurance to large SNOMED hierarchies,” in *AMIA Annual Symposium Proceedings*, pp. 1071-1080, 2013.
- [11] A. Agrawal and G. Elhanan, “Contrasting lexical similarity and formal definitions in SNOMED CT: Consistency and implications,” *Journal of Biomedical Informatics*, Vol. 47, pp. 192-198, 2013.
- [12] IHTSDO. SNOMED CT Technical Implementation Guide [Internet] Available: <http://www.snomed.org>.



류우석(Wooseok Ryu)

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