



### 저작자표시-비영리-변경금지 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.

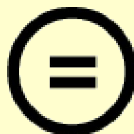
다음과 같은 조건을 따라야 합니다:



저작자표시. 귀하는 원저작자를 표시하여야 합니다.



비영리. 귀하는 이 저작물을 영리 목적으로 이용할 수 없습니다.



변경금지. 귀하는 이 저작물을 개작, 변형 또는 가공할 수 없습니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 [이용허락규약\(Legal Code\)](#)을 이해하기 쉽게 요약한 것입니다.

[Disclaimer](#) 

간호학 석사 학위논문

청소년 우울 관련 소셜데이터 수집과  
분석을 위한 온톨로지 개발 및 평가

2016년 2월

서울대학교 대학원

간호학과 간호학 전공

정혜실

## 국문 초록

본 연구에서는 소셜미디어 상에서 청소년 우울과 관련된 정보를 수집하고 분석하기 위한 분석틀(framework)으로써 청소년 우울 온톨로지를 개발하였다,

Noy & McGuinness(2001)가 고안한 ontology development 101를 근거로, 4개국의 임상실무지침서 내 주요 개념을 추출한 뒤 분류체계를 작성하였고 클래스별 속성과 속성값, 클래스간 관계를 정의하는 청소년 우울 온톨로지를 개발하였다.

개발된 청소년 우울 온톨로지는 452개의 클래스를 포함하여 총 2,073개의 용어와 클래스의 유사어·동의어를 수록하고 있으며 청소년 우울의 위험요인, 증상 및 징후, 측정, 진단결과, 치료 및 관리 도메인으로 구성되어있다.

청소년 우울 온톨로지는 아래와 같은 세 가지 방법으로 평가되었고 모든 평가에서 우수성이 입증되었다.

- 1) 약 16만 건의 청소년 우울 관련 소셜미디어 데이터를 대상으로 온톨로지 개념의 출현여부를 조사한 결과 온톨로지 개념의 72.8%가 소셜미디어 데이터 내에서 출현하였다.
- 2) 청소년 우울 관련 상담기록 66건을 자연어처리과정(NLP)을 거쳐 핵심개념을 추출하고 핵심 개념과 온톨로지 개념을 매핑한 결과 73.1%가 완전히 매핑(completely mapped)되었으며 부분 매핑(partially mapped)의 경우에도 21%로 나타났다.
- 3) 온톨로지 개발 경험이 있는 간호정보학 전문가 4인과 청소년 정신건강 전문가 2인을 대상으로 Kehagias, Papadimitriou, Hois, Tzovaras & Bateman(2008)이 제안한 온톨로지 평가 기준에

따라 온톨로지의 구조 및 표현력을 평가한 결과, 모든 항목에서 평균 4.3이상의 점수를 받았다.

본 연구에서 개발한 청소년 우울 온톨로지는 송태민(2015)의 연구를 통해 소셜빅데이터 내 우리나라 청소년의 우울 현황과 우울 요인을 분석하는 틀로써 활용되었고 매우 의미 있는 결과를 도출하였다.

청소년 우울 온톨로지는 청소년 자살을 비롯해 청소년 비행과 같은 개인, 가족, 학교 및 지역사회 차원의 통합적 관점이 요구되는 사회 현상과 이슈를 정의하는데 일반화 되어 활용 가능할 것으로 예상된다.

**주요어:** 청소년우울, 온톨로지, 소셜미디어데이터,

ontology development 101, 자연어처리기법, 청소년상담기록

**학번:** 2013-20442

# 목 차

국 문 초 록.....	i
I. 서론.....	1
1. 연구의 필요성.....	1
2. 연구 목적.....	3
3. 용어 정의.....	3
II. 문헌 고찰.....	5
1. 청소년 우울증 임상실무지침.....	5
2. 온톨로지 개발 및 평가.....	7
3. 보건의료분야 소셜데이터 수집과 분석.....	9
4. 온톨로지 기반의 빅데이터 분석.....	11
5. 용어체계 매핑 사례를 통한 온톨로지 포함력 평가.....	12
III. 연구 방법.....	14
1. 연구 설계.....	14
2. 연구 도구.....	14
3. 연구 진행 절차.....	17
1) 청소년 우울 온톨로지 개발.....	17
2) 청소년 우울 온톨로지 평가.....	19
4. 자료 수집 절차.....	23
1) 연구의 윤리적 고려.....	23
2) 자료 수집 기간 및 과정.....	23
IV. 연구 결과.....	24
1. 청소년 우울 온톨로지 개발.....	24
2. 청소년 우울 온톨로지 평가.....	34
1) 온톨로지 용어의 소셜데이터 내 출현 빈도.....	34
2) 온톨로지 개념과 청소년상담기록 개념 매핑.....	36
3) 온톨로지 구조 및 표현력 평가.....	45
V. 논의.....	48
VI. 결론 및 제언.....	54

<b>참고문헌</b> .....	57
부록1. 생명윤리위원회 (IRB) 심사결과 통지서 .....	62
부록2. 온톨로지 평가용 설문지 .....	63
<b>Abstract</b> .....	68

## List of Tables

Table 1. Evaluation Criterion of structural and representational layer .....	16
Table 2. Youth depression guidelines.....	25
Table 3. Example of the sign and symptoms hierarchy.....	26
Table 4. Example of synonyms in hospital management care .....	27
Table 5. Example of properties of behavioral change class .....	28
Table 6. Example of the facets of the properties.....	29
Table 7. Relationship-linking classes in the ontology .....	33
Table 8. Information of social data collected.....	34
Table 9. The frequency of terms and concepts appeared in the social data .....	35
Table 10. Mapping the concept of youth records to ontology concepts .....	37
Table 11. Mapping the concept of parents' records to ontology concepts.....	39
Table 12. Mapping the concept of teachers' records to ontology concepts .....	41
Table 13. Mapping the concept of counselors 'records to ontology concepts.....	43
Table 14. Mapping the concept of counseling records to ontology concepts.....	44
Table 15. Evaluation Scores of the structural layer (N=4) .....	45
Table 16. Evaluation scores of the representational layer (N=6) .....	47

## List of Figures

Figure 1. Role of an ontology in text mining.....	11
Figure 2. Mapping process.....	21
Figure 3 Youth depression ontology.....	31
Figure 4. Example of an property relationship : Depression.....	32
Figure 5. Terms extracted from youths' records.....	36
Figure 6. Terms extracted from parents' records.....	38
Figure 7. . Terms extracted from teachers' records.....	40
Figure 8. Terms extracted from counselors' records(freq≥5).....	42



# I. 서론

## 1. 연구의 필요성

2012년 국내 청소년 사망원인 1위는 '고의적 자해' 이른바, 자살이었다(통계청, 2014). 정부의 자살예방 5개년 종합대책(보건복지부, 2005)에 따르면 자살을 선택한 사람의 80%는 자살을 시도한 당시 우울증과 깊은 관련이 있었다고 한다. 청소년 자살 문제 역시 청소년 우울증과의 연관성을 배제할 수 없는데, 2014년 통계청 조사 결과 청소년의 자살충동을 유발하는 주요 원인은 '성적 및 진학 문제', '가정불화', '경제적 어려움'이었다. 위 원인이 청소년 우울증의 주요 위험요인(최규련, 2010)이라는 점에서 청소년 자살과 우울증 간 관련이 있음을 알 수 있으며 이미 많은 연구에서 청소년의 자살 생각과 우울증간 높은 관련성이 있음을 밝혀왔다(전영주, 2001; 조성연, 김혜원, & 김민, 2010). 청소년 우울증은 만성적이고 재발적인 경과를 가지며 정신 장애를 유발할 가능성을 20배 이상 증가시키는 등 청소년의 지속적인 발달과 성장을 저해하는 사회문제로 인식되고 있다(권호인, 2012). 또한 청소년 우울은 성인 우울에 비해 감정기복이 심하고 자살과 같은 충동적인 행동을 유발(최인재, 모상현, 강지현, & 김윤희, 2011)한다는 점에서 청소년 우울 위험을 조기에 발견하고 예방할 수 있는 연구가 필요하다.

2013년 통계청 조사에 따르면 국내 청소년의 77.1%가 소셜네트워크 서비스(이하 SNS)계정을 가지고 있으며 이 중 53% 이상이 100명 이상의 팔로워를 맺고 있는 것으로 나타나는 등 SNS는 우리나라 청소년들에게 자유로운 교류와 소통의 장으로써 활용되고 있다. 청소년들은 SNS를 통해 일상생활에서 경험하는 다양한 감정과 스트레스, 고민을 꾸

밌없이 표현하고 이를 타인과 공유하고 있는데 실시간으로 생성되고 빠르게 전파되는 SNS의 특성으로 인해 개인의 포스팅은 그들과 관계를 맺고 있는 타인에게 큰 영향력을 미치고 있다. 일례로, 수학능력시험 직후 한 수험생이 성적을 비판하며 자살을 암시하는 글을 SNS에 올려 사회적으로 큰 파장을 불러일으킨 사건이 있었다. 과거에는 '일기' 수준에서 머물렀을 개인의 기록이 SNS를 통해 타인과 '공유'되고 대중의 '관심'을 얻게 됨으로써 사회적 '이슈'로 전환된 사례라 할 수 있다. 이렇듯 SNS는 청소년들이 일상생활 속에서 갖는 우울한 감정이나 스트레스, 고민들을 들을 수 있고 행태를 이해할 수 있는 공간(송태민 & 송주영, 2014)으로써 SNS상에서 나타나는 청소년의 우울과 관련된 감정표현과 위험 행태들을 분석하면 위험징후와 의미 있는 패턴을 감지함으로써 청소년 우울, 더 나아가 청소년 자살을 예방하는데 긍정적인 효과를 발휘할 수 있을 것(한국정보화진흥원, 2012)으로 기대된다.

SNS를 비롯한 소셜미디어로부터 청소년 우울 데이터를 수집하고 이를 분석하기 위해서는 청소년 우울과 관련된 방대한 지식을 수록한 청소년 우울 온톨로지가 필요하다. 온톨로지는 특정 분야의 합의된 개념과 개념간의 관계를 표현함으로써 해당 영역의 지식을 체계적으로 표현(노상규 & 박진수, 2007)하는 효과적인 도구이며 소셜데이터와 같은 빅데이터를 수집하고 분석하는데 필요한 분석틀(framework)으로써 활용된다. 청소년 우울 온톨로지는 청소년 우울과 관련하여 소셜데이터 수집 시 검색해야 할 키워드를 제공할 뿐만 아니라 수집된 데이터를 컴퓨터가 의미 있게 해석할 수 있도록 도와준다. 따라서 청소년 우울 온톨로지를 활용하면 청소년 우울 전문가가 아닌 기계가 수집된 소셜데이터 내 용어의 의미를 올바르게 파악하여 정교한 해석과 분석이 가능할 것으로 예상된다.

본 연구에서는 청소년 우울 소셜데이터 수집과 분석을 위해 청소년

우울증 온톨로지를 개발하고 개발한 온톨로지의 타당성을 평가하는 것을 목적으로 한다. 또한 청소년 우울 온톨로지 내 주요 용어의 유사어·동의어를 기술한 용어체계를 개발하여 비정형 소셜 데이터의 정교한 의미분석에 활용될 수 있도록 하였다.

## 2. 연구 목적

본 연구의 목적은 청소년 우울과 관련된 소셜데이터를 수집·분류하기 위한 분석틀로써 온톨로지를 개발하고 평가하는 것으로 구체적으로

- 1) 영국·미국·호주·한국의 청소년 우울증 임상실무지침서를 기반으로 지식을 추출하여 온톨로지를 개발한다.
- 2) 온톨로지 용어의 소셜데이터 내 출현빈도·온톨로지 개념과 우울관련 청소년 상담기록 개념의 매핑을 통한 포함력 평가·전문가 설문을 통한 온톨로지 구조 및 표현력 평가를 실시하여 개발한 온톨로지의 타당성을 평가한다.

## 3. 용어 정의

### 1) 소셜 데이터

소셜 데이터는 사용자가 본인의 의견, 생각, 경험, 관점 등을 공유하기 위해 사용하는 소셜 미디어 상 텍스트·이미지·오디오·비디오 형태의 콘텐츠를 뜻한다.

본 연구에서는 199개의 온라인 뉴스사이트, 4개의 블로그, 2개의 카페, 1개의 SNS, 9개의 게시판 등 총 215개의 소셜 미디어 채널을 통해 수집된 텍스트 기반의 담론을 의미한다.

## 2) 청소년 우울 온톨로지

임상적 우울증은 불행감 또는 흥미 상실이 지속적으로 나타나며 청소년 우울증의 경우 성인의 우울증과 달리 짜증(irritability)과 공격적(aggressive)이고 자기 파괴적(self-destructive)인 행동이 수반되는 정신과적 질환이다(USPSTF, 2009). 온톨로지란 특정분야의 지식체계를 모델링할 때 관련되는 개념들을 정의하고 개념간의 관계들로 구성하여 컴퓨터가 이해하고 처리할 수 있는 형태로 표현한 개념들의 논리적 집합이다(박정은, 정귀애, 조훈, & 김화선, 2013).

본 연구에서의 청소년 우울증 온톨로지는 임상실무지침서를 바탕으로 추출된 청소년 우울증의 위험요인, 증상 및 증후, 측정, 진단결과, 치료 및 관리 영역에 대한 개념과 개념간의 관계를 의미한다.

## 3) 청소년 우울 온톨로지의 포함력

온톨로지의 포함력은 비정형 데이터로부터 추출한 개념이 온톨로지 내 개념에 포함되는 정도를 말한다.

본 연구에서 온톨로지의 포함력은 청소년 상담기록에서 추출한 개념이 청소년 우울 온톨로지 개념으로 표현되는 정도를 말한다. 청소년 우울 온톨로지의 포함력은 청소년 상담기록으로부터 추출한 개념을 청소년 우울 온톨로지 개념에 매핑하여 확인하였으며, 매핑 유형별 매핑률을 구하였다.

## II. 문헌 고찰

본 절에서는 청소년 우울증 온톨로지를 개발하기 위해 참고한 4개국의 임상실무지침과 기존에 개발된 보건의료영역의 온톨로지 사례, 보건의료영역의 소셜데이터 수집 및 분석 연구, 온톨로지 기반의 빅데이터 분석 연구를 비롯해 용어체계 매핑 사례를 통해 온톨로지 포함력을 평가하는 문헌을 고찰하였다.

### 1. 청소년 우울증 임상실무지침

청소년 우울증 온톨로지 개발을 위한 지식 추출은 영국(NICE, 2005), 미국(USPSTF, 2009), 호주(Beyondblue, 2011), 한국(우울증 임상연구센터, 2011)의 청소년 우울증 임상실무지침을 바탕으로 수행되었다. NICE는 지난 2005년 소아청소년 우울증 관리를 위한 임상실무지침서를 발간한 이후 지난 3월에 개정판을 공개하였다(NICE, 2015). 본 임상실무지침서는 ICD-10(International Statistical Classification of Diseases)의 기준에 의거하여 소아청소년 우울증을 3단계(경증 우울증, 중등도 이상의 우울증, 정신증을 동반한 중증 우울증)로 분류하여 서술하고 있다. 경증 우울증(mild depression)은 ICD-10에서 규정한 우울증의 10가지 증상 중 4개 이하의 증상을 보이는 경우이며 중등도 이상의 우울증(moderate and severe depression)은 5개 이상의 우울증 증세를 보이거나 정신증(psychotic symptoms)은 발현되지 않은 상태이다. 마지막으로 정신증을 동반한 중증 우울증(severe depression with psychotic symptoms)은 정신증과 5개 이상의 심한 우울증세가 함께 나타나는 우울증을 의미한다(NICE, 2015). 2005년도 버전과 비교해서 2015년 개정판에서는 중등도 이상의 우울증 소아청소년의 치료와 관련

된 사항이 수정되었다. 중등도 이상의 우울증 대상자에게 최소 3개월 이상 심리 치료(인지행동치료, 대인관계치료, 가족치료 등)를 제공하고 항우울제 단독 치료를 적용하지 말 것을 권고하고 있다(NICE, 2015). 단, 심리치료가 반응하지 않은 경우 fluoxetine 투약을 심리치료와 함께 적용할 것을 권장하고 있다(NICE, 2015).

U.S. Preventive Services Task Force(이하, USPSTF)는 2002년에 이어 2009년 소아청소년의 주요 우울증 선별에 대한 진료지침서를 출간하여 우울증 선별 도구의 정확성 과 심리치료 및 약물치료의 이득과 위험성에 대한 근거를 마련하였다(USPSTF, 2009). Medline, Cochrane, PsycInfo 등 데이터베이스에서 청소년 우울증과 관련된 주요 키워드를 검색하여 체계적 문헌고찰 및 메타분석을 실시하여 청소년 우울증에 대한 총체적인 임상지식을 집약하고 있어 NICE의 임상실무지침과 더불어 국내 우울증 임상실무지침 개발에 토대가 되었다(우울증임상연구센터, 2011). 호주는 2011년에 국가차원의 우울증 협의체(the national depression initiative)인 beyondblue를 통해 청소년우울증 임상실무지침을 배포하였다. 본 임상실무지침서는 청소년 우울증의 예방, 진단, 증상, 치료 및 관리에 대한 근거기반 실무 지식을 제공하고 있으며 우울증 위험에 속한 청소년의 건강수준 향상을 목적으로 우울증 외에 기분부전 장애, 양극성 장애와 관련한 지식을 포함하고 있다(Beyondblue, 2011).

한편, 국내 우울증 임상연구센터는 USPSTF 및 NICE의 임상실무지침을 검토하여 한국적 상황과 특성을 고려한 임상실무지침서를 출판하였다(우울증임상연구센터, 2011). 본 임상실무지침서는 성인/노인과 청소년을 구분하여 기술하고 있으나 일차의료기관에 내원한 청소년을 대상으로 우울증 선별 검사를 실시하여 우울증 환자를 조기에 발견하는 것을 목적으로 하며 USPSTF 및 NICE, Beyondblue의 임상실무지침에 비해 매우 제한적인 정보를 담고 있다. 청소년 우울증의 위험요인, 증상, 예방,

치료 등 종합적인 지식을 추출하기에 한계가 있으나 국내 상황을 반영하게 위해 본 연구에 포함하였다.

## 2. 온톨로지 개발 및 평가

온톨로지는 IT 및 정보학 분야에서 지식을 표현하는 도구로 활용되며 특정 영역의 개념과 개념간 관계를 정의하여 인간뿐 아니라 소프트웨어 간 커뮤니케이션 시 특정 단어가 나타내는 개념의 의미를 이해하는 데 사용된다(노상규 & 박진수, 2007). 온톨로지의 구성요소에는 개념(class/concept), 속성(property), 관계(relationship), 제약조건(constraint), 개체(instance), 공리(axiom)가 있다. 개념은 클래스라고도 하며 현실세계에 존재하는 것에 대한 일반적이고 본질적인 인식이나 지식을 의미한다. 속성은 개념에 속해 있는 성질이며 관계는 개념들 사이의 연결/연관으로 계층구조를 형성하는 상속관계, 합성관계가 있다. 제약조건은 개념들 간의 관계나 속성의 값에 대한 제한 규칙이며 개체는 개념의 실례를 의미한다. 마지막으로 공리는 추론의 기본이 되는 명제로서 절대 '참(true)'인 문장을 말한다(노상규 & 박진수, 2007).

온톨로지는 보건의료분야의 지식의 재사용성을 높이고 환자의 의무기록과 같은 방대한 정보를 자동화하여 처리하며 환자 중심의 의료 서비스를 제공하기 위해 활용되고 있다(김흥기 & 김명기, 2003). 청소년 상담을 위해 설계된 마음의 병 온톨로지는 가치관, 인생관, 세계관의 혼란 및 부정적 생활 사건으로 발생한 병의 원인 및 증상에 대한 내담자 개인의 맞춤형 정보를 제공하는데 유용하다(백현기, 2012). 마음의 병 온톨로지는 상담 진술문으로부터 원인, 증상, 상담 영역의 102개 용어를 추출하여 ontology development 101 방법론에 따라 개발되었다. 개발된 온톨로지는 2인의 전문가 평가에서 높은 타당도(CVI 0.92)를 얻었다.

그러나 마음의 병 온톨로지는 인문학 분야에서 논의되는 개념으로 설계되어 임상적 우울증에 대한 객관적이고 총체적인 지식을 표현하는데 한계를 가지고 있다.

임상적 우울증과 관련하여 정신과의사와 심리학자간 정보 교류를 증진하고 상대 분야에 대한 폭넓은 이해를 위해 개발된 기분장애(mood disorder) 온톨로지가 있다(Haghighi, Koeda, Takai, & Tanaka, 2009). 기분장애 온톨로지는 4개의 카테고리(양극성 장애, 기분부전 장애, 순환적 장애, 우울증)로 구성되며 질환별 관련 문제(associated problems), 원인, 진단, 역학, 예방, 경과, 증상, 치료 및 유형에 대한 총체적인 정보를 담고 있다. 온톨로지를 평가하기 위해 전문가 평가를 실시하여 기존 정신진단 컴퓨터 프로그램인 Decisionbase<sup>TM</sup>와의 비교 실험을 수행하였다. 그 결과 기분장애 온톨로지 시스템은 Decisionbase<sup>TM</sup>와 비교하여 '전문적 지식'과 '정보의 정확성' 측면에서 매우 우수한 것으로 나타났다(Haghighi et al., 2009). 이밖에 우울증 발생 가능성을 추론하여 우울증 진단을 돕는 목적으로 개발된 우울증 온톨로지(Chang, Hung, & Juang, 2013)는 '환자', '질환', '증상'만을 최상위 클래스로 가지며 우울증에 대한 총체적 정보를 제공하지 못한다는 한계가 있다.

한편, 간호진단 및 간호과정과 온톨로지를 결합하여 임상 의사결정 시스템(CDSS)을 개발한 연구(Din, Abidi, & Jafarpour, 2010; 박정은 et al., 2013; 이인근, 김화선, & 이성희, 2012)와 비만관리 온톨로지 작성 시 간호과정을 기반으로 한 연구(H.Y.Kim, H.A.Park, Y.H.Min, & E.Jeon, 2013)는 온톨로지가 간호 지식을 체계적이고 논리적으로 표현하는데 유용할 뿐 아니라 간호과정이 기본(foundational) 온톨로지로서 사용 가능함을 보여주었다. 기본 온톨로지란 특정 영역에 종속되지 않고 독립적이며 다양한 영역에 적용될 수 있는 기본적이고 보편적인 온톨로지를 의미하며(노상규 & 박진수, 2007) 간호과정이 기본 온톨로지로서



다른 분야의 온톨로지를 구축할 때 재사용될 수 있음을 뜻한다. 이밖에 온톨로지를 기반으로 항고혈압제와 식이간 상호작용 관리를 위해 구현된 임상 의사결정시스템은 전문가 평가를 통해 온톨로지의 정확성과 유용성이 검증되었고 향후 의료전문가의 효율적인 의사결정을 지원하고 약물과 식이간 상호작용으로 인한 부작용을 예방하는데 효과적일 것으로 기대된다(박정은, 김화선, 장민정, & 홍해숙, 2014).

### 3. 보건의료분야 소셜데이터 수집과 분석

소셜 데이터는 대중의 의견 및 개인의 심리적·행동적 특성을 실시간으로 드러내며(한국정보화진흥원, 2012) 사회적 트렌드를 민감하게 반영한다는 점에서 홍보 마케팅 분야는 물론 금융, 생산관리 등 다양한 산업부문에서 활용되고 있다. 보건의료분야 역시 소셜데이터를 분석하여 기존의 분석에서 얻을 수 없었던 새로운 가치와 통찰력을 얻으려는 시도가 활발히 지속되고 있다. 특히 전염병, 인플루엔자와 같이 질환의 확산 속도가 빠르나 공식적인 통계 산출에 상당한 시간이 소요되는 경우 실시간에 가까운(near real-time) 데이터 수집이 가능한 소셜미디어를 통해 질병 역학(dynamics)을 조기에 예측할 수 있다(Chunara, Andrews, & Brownstein, 2012). 2010년 아이티 지진 발생 후 콜레라 관련 소셜미디어(HealthMap, Twitter) 포스팅 건수와 아이티 보건부의 콜레라 통계 수치와의 상관분석 결과 소셜미디어 데이터 양과 정부의 공식적 통계 데이터간 유의한 상관관계가 도출되었다(Chunara et al., 2012). 이는 소셜데이터가 기존의 공식적 통계 데이터를 보완하여 전염병 발생 및 확산 추세를 조기에 예측하는 새로운 조사 도구로써 활용 가능성을 보여준다. Signorini et al(2011)은 트위터가 H1N1 인플루엔자와 관련한 대중의 관심과 우려를 의미있게 반영할 뿐 아니라 인플루엔자의 활동성을 공

식적인 통계보다 1-2주 앞서 예측하였음을 증명하였다. 트위터에서 인플루엔자와 관련한 키워드 (flu, swine, influenza, vaccine, tamiflu, h1n1, symptom, syndrome and illness)를 검색하여 데이터를 추출하였으나 키워드 선정의 배경에 대한 내용은 수록되어 있지 않다. 이는 Chunara et al.(2012)의 연구에서도 마찬가지이며 추후 데이터 검색의 민감도 (sensitivity)를 향상시키기 위해 이론적 근거를 기반으로 한 키워드 선정 과정이 보완되어야 한다. 트위터 상 소아비만과 관련한 담론을 분석한 연구(Harris, Moreland-Russell, Tabak, Ruh, & Maier, 2014)에서는 소셜미디어가 대중들의 건강정보를 교류하는 소통채널로서 활용되는 현황을 조명하고 정부 및 교육기관이 주도하여 신뢰도 높은 건강정보가 소셜미디어를 통해 활발히 제공되어야함을 강조하였다.

한편, 국내에서는 소셜데이터를 활용하여 청소년들이 자살과 관련하여 가지고 있는 생각과 행동, 심리적 특성을 파악하려는 연구가 진행되었다. 뉴스 및 블로그, 카페, SNS에 게재된 데이터를 분석대상으로 하여 버즈 분석, 키워드 분석, 감성 분석, 영향력자 분석이 시행되었다. 분석 결과 청소년들이 자살에 대해 보이는 독특한 특성이 확인되었고 온라인 상에서 청소년 자살과 관련된 언급이 폭증하는 특정 시점을 확인할 수 있었다(한국정보화진흥원, 2012). 구글의 검색트렌드를 이용하여 한국의 자살 검색량의 결정요인에 대해 다변량 분석을 실시한 연구(송태민, 2012)에서는 한국의 자살률과 구글의 자살 검색량이 비슷한 추세를 보이며 연도별 자살률이 자살 검색량을 유의하게 높이는 것으로 나타났다. 또한 스트레스 검색량이 자살 검색량에 직접적인 영향을 미치는 것을 발견하여 SNS를 통해 이용자의 키워드 검색 패턴을 분석함으로써 자살 징후를 포착하고 필요한 서비스를 제공하는 방안이 제안되었다. 그러나 자살요인 분석 시 '스트레스, 음주, 운동, 자살' 키워드만을 검색하여 자살의 관련요인을 충분히 반영하지 못했다는 한계를 가진다.

#### 4. 온톨로지 기반의 빅데이터 분석

온톨로지는 해당영역의 데이터와 지식을 결합함으로써 빅데이터를 의미 있게 해석하고 분석하는데 중요한 역할을 한다(Kuiler, 2014). 온톨로지는 비정형적 텍스트 데이터(자연어)를 분석하기 위해 수행하는 텍스트 마이닝 과정에서 해당분야에 대한 전문적 지식을 컴퓨터가 이해할 수 있게(machine-readable) 논리적이고 명시적으로 표현하여 데이터가 내포하는 의미를 컴퓨터가 정교하게 분석할 수 있도록 돕는다(Spasic, Ananiadou, McNaught, & Kumar, 2005). 온톨로지를 기반으로 텍스트 데이터로부터 도출된 지식은 Figure 1과 같이 온톨로지의 내용을 수정하고 보완하는데 다시 활용된다(Kuiler, 2014).

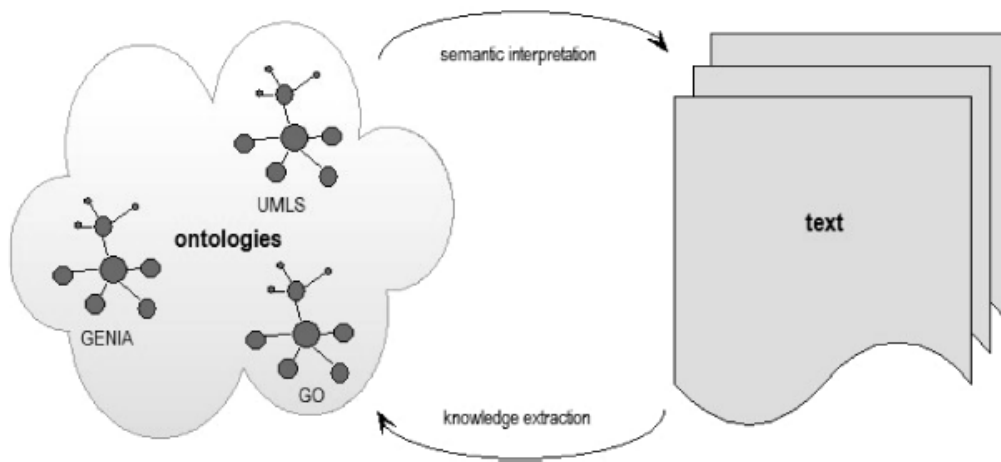


Figure 1. Role of an ontology in text mining

국내 연구에서도 온톨로지는 비정형적 텍스트 데이터를 분석하는 과정에서 발생하는 의미적 모호성을 해결하고 용어의 의미(semantic)를 파악함으로써 텍스트 마이닝의 품질을 향상시키는 것으로 나타났다(유은지, 김정철, 이춘열, & 김남규, 2012). 기존의 텍스트 분석은 주로 용어의

빈도수에 기반한 구조적(syntactic)분석에 의존하고 있어, 동음이의어 및 이음동의어를 정교하게 처리하지 못한다는 한계를 가지고 있으나 온톨로지는 하나의 용어가 여러 개념에 대응되는 문제를 해결하기에 적합한 체계라는 측면에서 텍스트 내 용어의 의미를 파악하는 장치로 사용될 수 있다. 실제 국내 대형 온라인 쇼핑몰 데이터 분석을 통해 온톨로지 기반의 텍스트 마이닝이 일반적인 텍스트 마이닝에 비해 우수한 예측력을 가짐을 증명하였다(유은지 et al., 2012).

생체임상의학 분야에서 온톨로지 기반의 텍스트 마이닝을 강조한 연구로는 Spasic et al.(2005)이 있다. 자연어로 표기된 생체임상의학 분야의 텍스트 데이터를 해석하기 위해서는 언어적 지식(linguistic knowledge), 일반적 상식(general knowledge)외에 생체임상의학 분야의 전문적 지식(domain-specific knowledge)이 필요한데 온톨로지는 생체임상의학 분야의 전문적 지식을 컴퓨터가 이해할 있는 형태로 코드화하여 표현함으로써 생체임상의학 전문가가 아닌 기계가 텍스트 데이터의 의미를 해석할 수 있다(Spasic et al., 2005). 그럼에도 불구하고 타 분야에 비해 생체임상의학 분야는 온톨로지를 바탕으로 텍스트 데이터를 검색·수집·분석하는 시스템이 부족해 텍스트 마이닝의 수준이 낮은 실정므로, 향후 온톨로지를 활용한 텍스트 마이닝이 활성화되어야 함을 강조하고 있다.

## 5. 용어체계 매핑 사례를 통한 온톨로지 포함력 평가

SNOMED CT를 비롯한 표준용어체계는 주요 영역 아래 각 개념들이 계층 구조를 형성하고 있으며 개념이 적어도 한 개 이상의 상위 구조에 포함되어 있으면서 유의어를 비롯해 다른 개념들과 관계를 맺고 있어(Eun-Young So & Park, 2011) 보건의료분야의 대표적인 온톨로지에

해당된다. 이러한 표준화된 용어체계는 다른 분야와의 정보 공유를 목적으로 용어체계 혹은 분류체계간 매핑 작업이 진행되어 왔다. 대표적인 연구로는 SNOMED CT와 ICNP ver.1의 일곱가지 축별 개념과의 매핑 작업을 통해 SNOMED CT의 포함력을 조사하는 연구(H.A.Park, Lundberg, Coenen, & Konicek, 2009)가 있으며 free text형태의 비정형 데이터를 표준용어체계와 매핑한 연구로는 서술문 간호기록을 ICNP와 매핑한 연구(Park et al., 2006), 환자문제 진술 기록과 SNOMED CT를 매핑한 연구(Elkin et al., 2006), 서술식 의사 의무기록을 SNOMED CT 및 ICNP와 매핑한 연구(Eun-Young So & Park, 2011) 등이 있다.

매핑 절차를 살펴보면, 우선 확인하려는 개념과 형태가 일치하는 개념을 타 용어/분류체계 내에서 살펴본 후 구조적으로도 일치하는가를 확인한다. 개념의 형태와 구조가 모두 일치하는 경우를 언어적 매핑(lexically mapped)으로 정의한다. 개념과 형태는 다르지만 의미가 일치하는 경우는 의미적 매핑(semantically mapped)이라고 한다. 언어적 매핑과 의미적 매핑을 합하여 완전히 매핑되었다(completely mapped)고 한다(H.A.Park et al., 2009). 부분적 매핑(partially mapped)에 해당되는 매핑 유형으로 개념이 타 용어/분류체계 내 포괄적인 개념과 매핑되는 경우(mapped to a broader concept), 용어/분류체계 내 세부적인 개념과 매핑되는 경우(mapped to a narrower concept), 개념이 타 용어/분류체계 내 여러 개념과 매핑되는 경우(one-to-many mapping)가 있다. 마지막으로 타 용어/분류체계 내 어떤 개념에도 매핑되지 않는 경우는 not mapped으로 분류한다.

### Ⅲ. 연구 방법

#### 1. 연구 설계

본 연구는 청소년 우울증 임상실무지침서를 기반으로 관련 지식을 추출하여 온톨로지를 개발한 후 소셜데이터 및 청소년 우울 상담기록 내 개념과 온톨로지 개념을 비교·분석하여 온톨로지의 포함력을 확인하고 전문가 설문을 통해 온톨로지 구조와 표현력을 평가하는 조사연구이다.

#### 2. 연구 도구

##### 1) 온톨로지 평가 도구

온톨로지의 구조와 표현력은 Kehagias, Papadimitriou, Hois, Tzovaras & Bateman(2008)이 제안한 온톨로지 평가기준에 5점 리커트 척도를 결합한 17개 항목과 개방형 질문으로 평가하였다. 온톨로지 구조 평가 항목은 개념과 관계의 수(size), 계층구조의 깊이(depth of hierarchy), 계층구조의 폭(breadth of hierarchy), 개념간 관계 수(density), 균형감(balance), 전반적인 구조의 복잡성(overall complexity), 개념간 연결성(connectivity)의 7개 항목으로 이루어져있다.

온톨로지의 표현력 평가 항목은 개념의 형식적 정의가 인지적으로 잘 수용되는지(match between formal and cognitive semantics), 개념의 형식적 정의에 모순되는 점이 없는지(consistency), 문맥 및 배경적 지식이 잘 드러나 있는지(clarity), 개념들이 분명하게 표현되어 있는지(explicitness), 개념의 정의와 설명이 잘 이해되는지(interpretability), 온톨로지 개념들이 소셜미디어에서 사용되는 개념에 잘 부합되는지

(accuracy), 온톨로지 도메인의 범위가 포괄적(comprehensiveness)이고 상세하게(granularity) 개발되었는지, 온톨로지가 소셜데이터 분석을 위한 분석틀로써 개발 목적에 맞게 설계되었는지(relevance), 온톨로지가 다루고자 하는 범위를 잘 기술하고 있는지(description)에 따라 10개 항목으로 구성되어 있다(Kehagias, Papadimitriou, Hois, Tzovaras, & Bateman, 2008). 각 항목은 1: 매우 동의하지 않음 2: 동의하지 않음 3: 보통임 4: 동의함 5: 매우 동의함으로 평가하였다. 본 연구에서는 Kehagias, Papadimitriou, Hois, Tzovaras & Bateman(2008)의 온톨로지 구조 및 표현력 평가 항목을 번역한 후 간호정보학 전문가의 확인을 받아 사용하였다.

Table 1. Evaluation Criterion of structural and representational layer

structural/architectural layer	representational/semantic layer
size (how many concepts, relations, etc.)	Match between formal and cognitive semantics (cognitive adequacy)
Density (average branching)	Consistency ( no formal contradictions)
depth of hierarch	Clarity (context, background knowledge)
breadth of hierarch	Explicitness
Balance (are all areas equally developed or are there areas of low complexity and others with high complexity)	Interpretability (formal, informal, logical specifications, documentation)
overall complexity	Accuracy
connectivity between concepts, relations, etc. (Highly interconnected? Loosely interconnected e.g., only by is a-links?)	Comprehensiveness (extent of target domain covered)
	Granularity (fine-grained coverage vs. loose coverage)
	Relevance (for application and users)
	Prescriptive vs. descriptive (i.e., does it try and describe the world, or does it regulate how some portion of the world must be?)



### 3. 연구 진행 절차

#### 1) 청소년 우울 온톨로지 개발

청소년 우울 온톨로지는 Ontology Development 101(Noy & McGuinness, 2001)방법론에 따라 개발되었다. Ontology Development 101 방법론에 근거한 청소년 우울 온톨로지 개발 과정은 다음과 같다.

##### (1) 온톨로지의 영역과 범위를 구체적으로 설정한다.

본 연구의 목적은 소셜미디어 상에 등장하는 청소년 우울과 관련된 자료를 수집·분석하는 틀(framework)로써 청소년 우울 온톨로지를 개발하는데 있다.

따라서 온톨로지는 청소년 우울과 관련하여 소셜미디어 상에 게재될 주요 영역에 대한 지식을 다루어야 하며, 주요 영역을 선정하기 위해 영국·미국·호주·한국의 청소년 우울 관련 임상실무지침을 분석하였다.

##### (2) 기존 온톨로지를 검색하여 재사용이 가능한 온톨로지 여부를 확인한다.

Pubmed, Google scholar에서 기존에 개발된 청소년 우울 온톨로지의 존재 여부를 검색하였다. 검색어는 youth depression ontology, adolescents depression ontology로 한정하였다.

##### (3) 온톨로지를 구성하는 주요 용어를 수집하고 열거한다.

영국 NICE(2005, 2015), 미국 USPSTF(2009), 호주 beyondblue(2011), 한국 우울증임상연구센터(2011)에서 발행한 임상실무지침서를 토대로 주요 용어를 수집하였다. 그밖에 청소년 우울 관련 문헌과 논문, 인터넷 검색을 병행하여 관련 용어를 수집하였다.

##### (4) 클래스를 정의하고 클래스간 분류체계를 작성한다.

수집된 용어 중 클래스로 정의할 용어를 정하고 클래스의 동의어 또는 유사어를 추가하여 용어체계(terminology)를 개발하였다. 클래스는

핵심 혹은 중심 개념으로, 개념의 특성이나 상태를 기술하는 서술자가 아닌 독립적으로 존재하는 개념을 말한다(Noy & McGuinness, 2001).

클래스를 정의한 후 클래스간 분류체계를 설계하는데 top-down기법을 활용하였다. Top-down기법은 가장 일반적이고 포괄적인 개념을 나열한 후 구체적인 개념을 배열하는 방법으로, 청소년 우울 온톨로지의 주 영역인 위험요인, 증상 및 징후, 진단방법(이후 측정), 진단결과, 치료 및 관리를 먼저 나열하고 그 아래 세부적인 개념들을 하위클래스로 분류하였다.

#### (5) 클래스의 속성(슬롯)을 정의하여 내부구조를 기술한다.

개념이 담고 있는 정보를 제공하기 위해 개념의 속성 즉, 내부구조를 기술하였다. 모든 하위클래스는 상위클래스로부터 속성을 물려받으므로 상위클래스에 표기한 속성은 하위클래스에게도 계승된다. 클래스의 속성을 정의하기 위해 여성가족부의 청소년유해환경 접촉 종합실태조사·가정폭력 실태조사, 교육부의 학교폭력 실태조사, 서울시의 서울특별시 청소년 위기 실태조사 등의 설문지 문항을 분석하였다.

#### (6) 속성값의 유형, 사상수(cardinality)를 정의한다.

속성이 가지는 값(속성값)의 유형, 사상수를 비롯해 속성값의 정보를 정의한다. 일반적으로 속성값의 유형으로 String, Number, Boolean, Enumerated, Instance 유형이 있다. String은 이름과 같이 간단한 문자의 나열 형식이며 Number는 숫자 형식, Boolean은 예/아니오 형식, Enumerated는 정해진 값 중 특정 하나를 고르는 형식(예: 증상의 정도는 mild/moderate/severe), Instance 유형은 특정 클래스의 속성값으로 특정 클래스의 개체(instance)가 되는 경우를 의미한다.

한편, 사상수란 속성이 가질 수 있는 값의 개수를 의미한다. 속성값을 하나만 가지는 경우는 단일(single)로, 속성값을 여러 개 가지는 경우 다중(multiple)으로 표현한다. 더 나아가 사상수의 최소값과 최대값을

설정할 수도 있다.

### (7) 클래스 및 개체들의 속성값을 작성한다.

클래스 또는 개체가 가지는 속성의 속성값을 기술한다.

## 2) 청소년 우울 온톨로지 평가

청소년 우울 온톨로지는 다음 세가지 방법으로 평가하였다.

### (1) 온톨로지 개념의 소셜데이터 내 출현 빈도

#### ① 평가 대상

2012년 1월 1일부터 2014년 12월 31일까지 215개의 소셜미디어 채널에 게시된 우울 관련 소셜데이터를 대상으로 하였다. 우울 관련 소셜데이터란 기록 내 우울 또는 우울증 키워드를 포함하고 있는 소셜데이터를 말한다.

#### ② 평가 및 분석 방법

1차로 개발한 온톨로지 개념의 우울 관련 소셜데이터 내 출현 빈도를 확인하였다. 한국보건사회연구원의 통계정보연구실로부터 온톨로지 용어 1,000개의 소셜데이터 출현 여부 및 빈도 결과를 제공받았고 이를 개념 단위로 재 분류·분석하여 온톨로지 개념의 소셜데이터 내 출현 여부를 빈도와 백분율로 제시하였다.

### (2) 온톨로지 개념과 청소년상담기록 개념 매핑

#### ① 평가 대상

한국청소년상담복지개발원이 홈페이지를 통해 제공하는 부모청소년고민해결 상담 기록 중 우울관련 기록 66건을 대상으로 분석하였다. 상담 기록 66건은 우울과 관련하여 청소년이 작성한 글 21건, 부모가 자녀의 우울과 관련하여 작성한 글 10건, 교사가 학생의 우울과 관련하여 작성한 글 2건, 그리고 상담가의 답변 글 33건으로 구성되어 있다.

## ② 평가 방법

청소년 우울 상담 기록 66건을 R(version3.2.2.)의 KoNLP 패키지를 사용하여 자연어처리(natural language processing, NLP)작업을 수행하였다. 자연어처리 작업에서 출현한 용어의 기록 내 문맥상 의미를 분석하여 개념을 도출하였고 도출된 개념과 온톨로지 개념을 매핑하여 매핑유형에 따라 결과를 분류하였다.

매핑 결과는 개념의 형태 및 구조가 일치하는 경우(lexically mapped), 개념의 형태는 다르지만 의미상 일치하는 경우(semantically mapped), 보다 포괄적인 개념에 매핑되는 경우(mapped to a broader concept), 보다 구체적인 개념에 매핑되는 경우(mapped to a narrower concept), 한 개의 개념이 여러 개의 개념으로 나뉘어 매핑되는 경우(one to many mapping)로 분류(H.A.Park et al., 2009)하였으며 위 과정을 도식화하면 Figure 2와 같다.

매핑 결과는 온톨로지 개발 경험이 있는 정신보건 전문가 1인으로부터 타당성을 검증 받았다. 전문가의 검증 및 자문 결과를 바탕으로 매핑 결과를 최종 수정하였다.

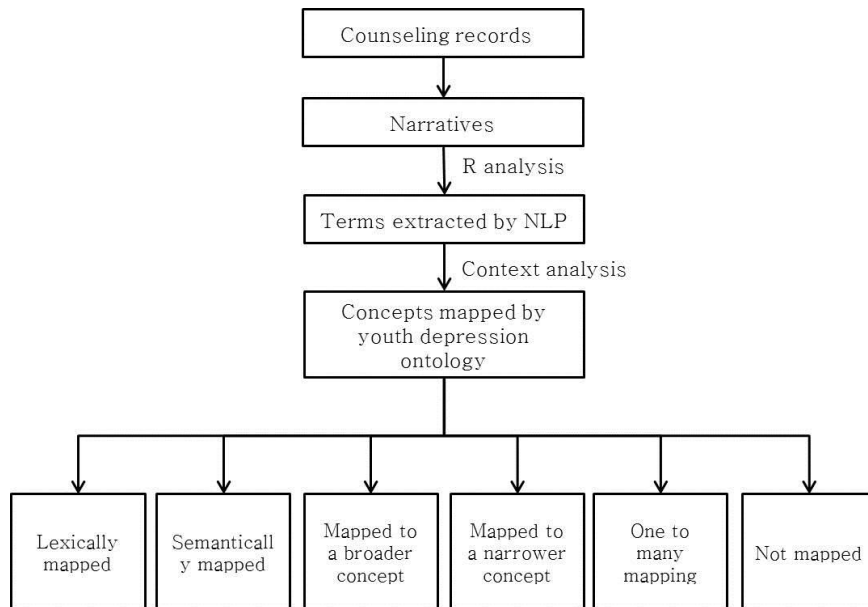


Figure 2. Mapping process

### ③ 분석 방법

청소년 우울 온톨로지의 포함력은 매핑유형별 빈도와 백분율로 제시하였다.

$$\text{매핑율 (\%)} = \frac{\text{매핑결과별 개념 수}}{\text{상담기록에서 도출된 개념 수}} \times 100$$

### (3) 온톨로지 구조 및 표현력 평가

#### ① 평가 대상

본 연구에서 개발한 청소년 우울 온톨로지의 구조와 내용을 평가 대상으로 하였다.

#### ② 평가 방법

온톨로지 구조 및 표현력은 Kehagias, Papadimitriou, Hois, Tzovaras & Bateman(2008)이 제안한 온톨로지 평가 기준을 5점 리커트 척도를 이용하여 평가하였다.

평가 절차는 간호정보학 전문가 4인(석사 이상)과 정신간호학 전문가 2인(석사 이상)으로 구성된 평가자들에게 온톨로지의 개발 목적과 평가 방법에 대해 소개하고, 평가 기준을 설명하였다.

온톨로지 구조의 타당도는 주요영역의 단계별 클래스와 클래스간 관계 유형을 표시한 온톨로지 그림과 이에 대한 설명을 바탕으로 평가하게 하였고 온톨로지 표현력은 개념별 속성과 속성값이 모두 정의된 온톨로지 모델링 결과를 제공하고 5개 영역별로 평가하도록 하였다.

#### ③ 분석 방법

평가 결과는 평가문항별로 점수의 평균과 최소값 및 최대값을 산출하였다.

## 4. 자료 수집 절차

### 1) 연구의 윤리적 고려

본 연구는 소셜데이터 수집을 비롯해 전문가 평가 과정에서 개인식별 정보를 수집하거나 기록하지 않았으며 한국청소년상담복지개발원의 공개된 자료를 이용하였으므로 생명윤리 및 안전에 관한 법률 시행규칙 제13조 제1항 제2호에 근거하여 서울대학교 생명윤리위원회로부터 심의를 면제받았다.(승인번호 E1507/001-008)

### 2) 자료 수집 기간 및 과정

본 연구를 위해 2014년 9월부터 12월까지 청소년 우울 임상실무지침서 및 관련 문헌을 통해 주요 용어를 추출하여 온톨로지를 개발하였다.

2015년 4월부터 8월까지 소셜데이터 수집을 비롯해 청소년 상담기록 분석을 통한 온톨로지 포함력을 평가하였고 이를 바탕으로 기존 온톨로지를 수정하였다. 2015년 9월 한 달간 전문가 평가를 실시하여 온톨로지의 구조와 표현력을 평가하였다.

## IV. 연구 결과

### 1. 청소년 우울 온톨로지 개발

#### 1) 온톨로지의 영역과 범위를 구체적으로 설정한다.

청소년 우울의 주요 영역을 선정하기 위해서 영국·미국·호주·한국의 청소년 우울 관련 임상실무지침서의 도메인을 참고하였다.

조사 결과, 4개국의 임상실무지침서는 청소년 우울의 위험요인, 증상 및 징후, 진단방법, 진단결과, 치료 및 관리를 공통적으로 포함하고 있었다. 따라서 본 연구에서 개발하는 온톨로지는 위 5개 주제를 포괄하도록 개발되었다.

#### 2) 기존 온톨로지를 검색하여 재사용이 가능한 온톨로지 여부를 확인한다.

Pubmed, Google scholar에서 youth depression ontology, adolescents depression ontology를 검색어로 하여 기존 연구에서 개발된 청소년 우울 온톨로지 존재여부를 확인하였다.

검색 결과, 우울증 온톨로지는 기분장애 온톨로지(Haghighi et al., 2009)정신건강 온톨로지(Hadzic, Chen, & Dillon, 2008)의 서브온톨로지로서 개발되어 있었고 우울증의 관련 요인과 치료만을 주제로 다루고 있어 범위가 제한적이거나(Hadzic et al., 2008), 범위는 포괄적(Haghighi et al., 2009)임에도 청소년 우울이 가지는 특정 요인(학교, 교우관계 등)과 증상(무단결석, 짜증 등)등을 포함하지 않아 청소년 우울을 설명하기에 많은 제한점이 있었다.

이밖에 마음의 병 온톨로지(백현기, 2012)는 청소년 상담을 위해 인문학적 관점에서 개발되어 임상적 지식을 제공하는 데는 한계가 있었다.



따라서 소셜미디어에서 나타나는 청소년 우울의 전 영역(위험요인, 증상 및 징후, 진단방법, 진단결과, 치료 및 관리)을 설명하고 분석하는데 적합한 새로운 온톨로지를 개발하기로 하였다.

### 3) 온톨로지를 구성하는 주요 용어를 수집하고 열거한다.

청소년 우울 임상실무지침서 4종(Table 2) 및 관련 문헌과 논문, 인터넷 검색을 바탕으로 청소년 우울 위험요인, 증상 및 징후, 진단방법, 진단결과, 치료 및 관리와 관련된 용어를 중복여부를 고려하지 않고 수집한 후 열거하였다.

이 과정에서 총 2,073개의 용어가 수집되었다.

Table 2. Youth depression guidelines

No.	Title	Publication	Country	Issued
1	Depression in children and young people	NICE(National Institute for Health and Care Excellence)	U.K	2005 ,2015
2	Screening for child and adolescent depression in primary care settings	USPSTF(U.S. Preventive Services Task Force)	USA	2009
3	Clinical Practice Guidelines : Depression in adolescents and young adults	Beyondblue	Australia	2011
4	일차의료기관에서 우울증 선별에 대한 진료지침	우울증임상연구센터	Korea	2011

#### 4) 클래스를 정의하고 클래스간 분류체계를 작성한다.

청소년 우울 온톨로지는 5개 영역, 총 452개의 클래스로 구성되었다. 주제별로 살펴보면, 위험요인 영역은 5단계·197개 클래스로 이루어져있으며 증상 및 징후 영역은 3단계·63개 클래스, 진단방법을 설명하는 측정 영역은 2단계·20개 클래스, 진단결과 영역은 1단계·3개 클래스, 치료 및 예방을 포함하는 치료 및 관리 영역은 5단계·169개 클래스로 구성되어 있다. Table 3는 증상 및 징후 영역을 구성하는 클래스와 분류체계를 나타낸 것이다.

Table 3. Example of the sign and symptoms hierarchy

category	Class(level1)	Class(level2)	Class(level3)
증상 및 징후	감정변화	우울, 불안, 슬픔, 무감정, 불행, 짜증, 분노, 적개심, 즐거움 상실, 무가치함, 죄책감, 의욕상실	
	행동변화	잦은 울음, 우유부단, 학습부진, 홀로 지냄/위축, 개인위생 저하, 외모에 대한 무관심, 자해, 자살 계획, 자살시도,	
		대인관계 저하	가족관계 저하 교우/또래관계 저하
		중독행동	인터넷중독, 게임중독, TV중독, 스마트폰중독
		비행 <sup>1)</sup>	흡연, 음주, 약물오용 및 남용 성관계, 등교거부, 가출, 공격적 행동
	인지변화	자기비판적 사고, 집중력 저하, 자존감 저하, 부정적 신체상, 죽음에 대한 생각, 자살에 대한 생각	
	신체/물리적변화	피로, 통증,	

		성욕저하	
		식욕변화	식욕증가, 식욕감소
		체중변화	체중증가, 체중감소
		수면리듬변화	불면, 과다수면
		정신운동변화	정신운동지체, 정신운동초조

1) 한국청소년패널조사에서는 비행의 정의를 ‘흡연, 음주, 무단결석, 가출, 성관계, 폭력 패싸움, 남의 돈/물건 뺏기, 남의 돈/물건 훔치기, 원조교제, 남을 놀리거나 조롱하기, 협박, 집단 따돌림, 성폭행/성희롱’으로 정의하고 있다. 본 정의에 입각하여 청소년 우울 증상 중 ‘비행’에 해당하는 개념을 하위클래스로 나타내었다.

수집된 용어와 추가적인 검색(Daum, Naver 검색엔진 어학사전 이용)을 바탕으로 클래스의 유사어/동의어를 추가하여 총 1,680개 용어를 수록한 용어체계를 개발하였다.

Table 4. Example of synonyms in hospital management care

Class(level1)	Class(level2)	Class(level3)	Synonyms
병원치료			병원관리, 전문적 치료, 전문의료기관서비스 의료기관관리
	정신심리치료		정신치료, 심리치료, Psychotherapy
		인지행동치료	인지치료, CBT
		대인관계치료	IPT
		정신역동치료	정신분석치료
	물리적치료		Physical treatment, 물리치료
		전기경련요법	ECT, 전기요법, 전기치료, 경련요법,
		경두개자극	TMS

**5) 클래스의 속성을 정의하여 내부구조를 기술한다.**

클래스의 속성을 정의하기 위하여 여성가족부, 교육부를 비롯한 정부 부처의 실태조사 보고서 내 설문 문항을 분석하고 클래스 개념과 관련한 전문서적 및 논문을 참고하였다. 예를 들어, 측정도구 중 Hamilton Depression Rating Scale 클래스는 측정 대상, 측정자, 도구 사용 연령, 검사항목 수, 총 점수, 절단점, 점수해석범위, 측정항목 등 아홉 개의 속성으로 상세히 표현될 수 있으며 자살시도 클래스는 발생시점, 자살시도 장소, 자살시도 목적, 자살시도 방법 등의 열두 개의 속성으로 표현될 수 있다.

하위클래스는 상위클래스의 속성을 물려받으며(Noy & McGuinness, 2001) 추가적인 속성을 갖도록 설계하였다. 이에 대한 예시는 Table 5 와 같다.

Table 5. Example of properties of behavioral change class

Class(level1)	Class(level2)	Class (level3)
행동변화 - Presence/증상여부	대인관계저하 - Presence/증상여부 - Severity/정도	또래관계저하 - Presence/증상여부 - Severity/정도 - Type/유형 - No. of friends in the past/ 과거 친하게 지낸 친구 수 - No. of friends in present/ 현재 친하게 지내는 친구 수

**6) 속성값의 유형, 사상수를 정의한다.**

Table 6는 증상 및 징후 중 자살시도와 진단결과 중 Depression 클래스의 속성값의 유형과 사상수를 표현하였다.

자살시도의 속성 중 발생시점은 free text로 자살시도 과거력과 정신

질환 과거력은 [있음|없음]과 같은 코드화된 텍스트 값으로, 정신질환 종류 및 자살시도 동기, 자살시도 방법은 코드화된 여러 텍스트 값으로 유형을 정의하였다. Depression 클래스의 속성 중 정도는 [mild | moderate | severe]와 같은 코드화된 텍스트 값으로, 유형은 멜랑콜리아 우울증, 비전형우울증, 정신병적우울증, 계절성우울증, 만성우울증, 긴장성우울증, 산후우울증, 다른 것에 속하지 않는 주요우울증과 같은 코드화된 텍스트 값으로 표현되도록 정의하였다.

자살시도의 속성 중 정신질환 종류와 자살시도 동기는 코드화된 텍스트 값 중 여러 개를 속성값으로 가질 수 있으므로(예를 들어, 불안장애와 조현병이 함께 있을 수 있음) 사상수는 multiple로, 자살시도 및 정신질환 과거력 여부를 묻는 속성은 [있음|없음] 중 하나의 값만을 가지므로 사상수를 single로 표현하였다. Depression의 정도와 유형 속성 또한 단일한 속성값만을 가지므로 single로 표현하였다.

Table 6. Example of the facets of the properties

class	property	value type	value set	cardinality
자살시도	자살시도 과거력	Boolean	[있음 없음]	single
	정신질환 과거력	Boolean	[있음 없음]	single
	발생시점	String	unit : 년/월/일	multiple
	정신질환 종류	Enumerated	[우울장애 불안장애 양극성장애 인격장애 적응장애 물질장애 정신지체 학습장애 발달장애 조현병]	multiple
	자살시도 동기	Enumerated	[정신과적증상 가족관계문제 친구관계문제 학교내왕따 학업성적 만성적빈곤 외로움 신체적질병 물리적폭행 성폭행]	multiple

	자살시도 방법	Enumerated	[음독 목매 날카로운물질로 찌르기 둔탁물질로때리기  투신 자동차/기차로뛰어들 거나앞에서있음 가스흡입  익수]	multiple
Depression	정도	Enumerated	[mild moderate severe]	single
	유형	Enumerated	[멜랑콜리아우울증 비전형 우울증 정신병적우울증 계 절성우울증 만성우울증 긴 장성우울증 산후우울증 다 른것에속하지않는주요우울 증]	single

### 7) 클래스 및 개체들의 속성값을 기술한다.

속성값의 유형이 Boolean이나 enumerated인 경우 코드화된 텍스트 값 중에서 속성값을 가지므로 미리 속성값 리스트를 설정해야 한다.

예를 들어, 자살시도 클래스의 속성 중 증상여부(presence)의 경우 속성값의 유형이 boolean이므로 속성값으로 [있음|없음]을 설정하였다. 정신질환 종류(type of psychiatric disorder)속성의 경우 속성값의 유형이 enumerated이므로, [우울장애|불안장애|양극성장애|인격장애|적응장애|물질장애|정신지체|학습장애|발달장애|조현병]과 같이 속성값 리스트를 정의하였다.

### 8) Protégé를 사용해 온톨로지를 표현한다.

본 연구에서는 Protégé를 활용하여 온톨로지의 주요 클래스와 클래스 간 관계를 나타내었다. Protégé는 스탠포드 대학교 의료정보학 팀에서 개발된 오픈소스용 온톨로지 제작 소프트웨어로 온톨로지의 계층구조를 설계하고 개념간의 관계를 나타내기에 적합하다. Figure 3은 청소년 우울 온톨로지를 구성하는 위험요인, 증상 및 징후, 측정, 진단결과와 치료

및 관리 영역을 1단계 하위 클래스까지 전개한 후 클래스간 관계를 표현한 그림이다.

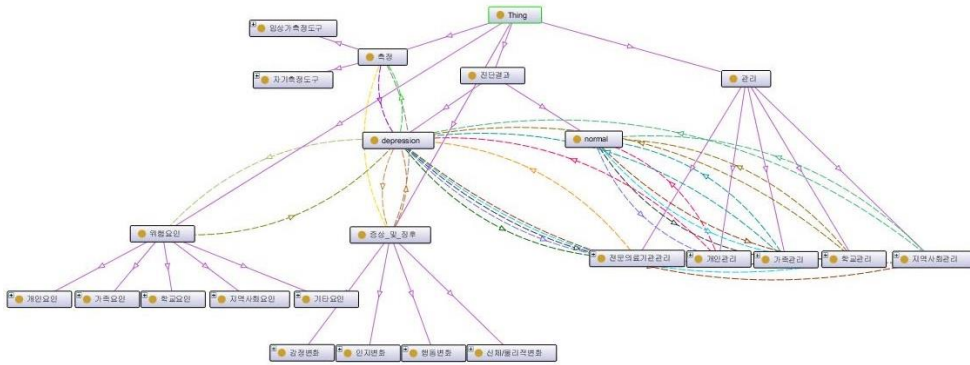


Figure 3 Youth depression ontology

위 그림에서 실선은 상위클래스와 하위클래스간 관계를 의미하며 점선은 5개 영역간 속성관계(object property relationship)를 나타낸다. 예를 들어, 진단결과의 하위클래스인 depression은 속성으로 위험요인, 증상 및 징후, 측정도구, 치료 및 관리방법을 가지고 있어 5개 영역인 이들과 속성관계를 맺고 있다 (Figure 4). 다시 말해 depression은 위험요인과 hasRiskFactors 관계를, 증상 및 징후와 hasSignAndSymptoms 관계를, 측정도구와 hasMeasurementTools 관계를, 치료 및 관리방법과 hasIndividualManagementCare, hasFamilialManagementCare, hasSchoolManagementCare, hasHospitalManagementCare, hasCommunityManagementCare를 맺고 있다. 또한 이에 대응하는 역 속성(Inverse property) 관계를 가지는데 위험요인과 depression은 isRiskFactorsOf 관계, 증상 및 징후와 depression은 isSignAndSymptomsOf 관계, 측정도구와 depression은

isMeasurementToolOf관계, 치료 및 관리방법과 depression은 isIndividualManagementCareOf, isFamilialManagementCareOf, isSchoolManagementCareOf, isHospitalManagementCareOf, isCommunityManagementCareOf 관계를 맺고 있다.

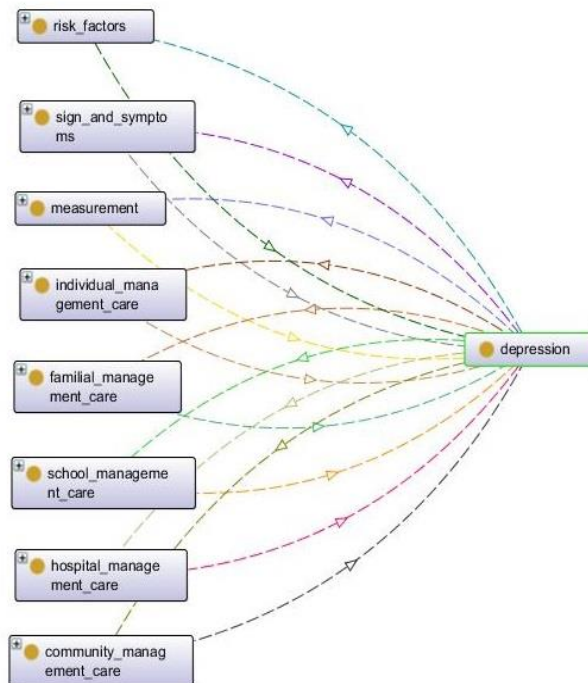


Figure 4. Example of an property relationship : Depression

증상 및 징후와 측정은 isAssessedBy, evaluatesSignAndSymptoms 관계를 가지고 Normal과 치료 및 관리방법은 Depression과 치료 및 관리방법간 속성관계 중 hasHospitalManagementCare, isHospitalManagementCareOf를 제외한 나머지 속성관계를 가진다.

청소년 우울 온톨로지는 Table 7와 같이 5개 영역간 19개 유형의 관



계(object property)가 존재한다.

Table 7. Relationship-linking classes in the ontology

Relationship name	Defined at class
Part of	상위클래스와 하위클래스간 관계
Has risk factors	Depression과 위험요인간 관계
Is risk factors of	위험요인과 Depression간 관계
Has measurement tool	Depression과 측정간 관계
Is measurement tool of	측정과 Depression간 관계
Has sign and symptoms	Depression과 증상 및 징후간 관계
Is sign and symptoms of	증상 및 징후와 Depression간 관계
Evaluates sign and symptoms	측정과 증상 및 징후간 관계
Is assessed by	증상 및 징후와 측정간 관계
Has individual management care	Normal 혹은 Depression과 개인관리방법간 관계
Is individual management care of	개인관리방법과 Normal 혹은 Depression간 관계
Has familial management care	Normal 혹은 Depression과 가족관리방법간 관계
Is familial management care of	가족관리방법과 Normal 혹은 Depression간 관계
Has school management care	Normal 혹은 Depression과 학교관리방법간 관계
Is school management care of	학교관리방법과 Normal 혹은 Depression간 관계
Has hospital management care <sup>1)</sup>	Depression과 병원치료방법간 관계
Is hospital management care of	병원치료방법과 Depression간 관계
Has community management care	Normal 혹은 Depression과 지역사회관리방법간 관계
Is community management care of	지역사회관리방법과 Normal 혹은 Depression간 관계

1) 병원치료방법은 병원차원의 우울증 치료방법에 대한 내용을 기술하였다.

## 2. 청소년 우울 온톨로지 평가

### 1) 온톨로지 용어의 소셜데이터 내 출현 빈도

한국보건사회연구원의 도움을 받아 sk텔레콤의 웹크롤러를 사용해 215개 소셜미디어 채널로부터 약 370만건의 우울 관련 소셜데이터가 수집되었다.

소셜데이터의 수집 기간 및 수집사이트 정보는 Table 8과 같다.

Table 8. Information of social data collected

수집기간	2012.01.01 ~ 2014.12.31
수집사이트 (215) <sup>1)</sup>	
뉴스(199)	AP연합뉴스·bnt뉴스·CBS노컷뉴스 등
블로그(4)	네이트·티스토리·다음·네이버 블로그
카페(2)	네이버·다음 카페
SNS(1)	트위터
게시판(9)	다음 미즈넷_생활속 에피소드, 다음 TIP, 다음 아고라, 네이버 지식인, 루리웹_커뮤니티 사회/정치/경제 정보, 세티즌, 디스이즈게임_커뮤니티/자유게시판, youtube, MLBPARK_BULLPAN

1) 국내에서 수집 가능한 소셜미디어 채널 전수임.

2) 한국보건사회연구원은 여성가족부·미래창조과학부의 ‘창조비타민 프로젝트’ 일환으로 ‘빅데이터 분석 기반의 위기청소년 예측 및 적시대응기술 개발 연구’를 통해 소셜 빅데이터를 수집하였음.

출처 : 송태민(2015). 소셜 빅데이터를 활용한 우리나라 청소년의 우울 현황과 위험요인 분석, 보건복지 Issue&Focus 제296호

수집된 소셜데이터를 대상으로 1차로 개발된 온톨로지 용어 1,000개의 출현 여부를 분석한 결과, 657개의 용어(65.7%)가 소셜데이터에 출현하였다.

용어의 동의어·유사어를 묶어 개념으로 처리한 결과 1,000개의 용어는 596개의 개념으로 분류되었고 이 중 465개의 개념(78.3%)이 소셜데이터에 출현하였다.

한편, 청소년이 작성한 것으로 판단되는 161,581건의 소셜데이터에 서는 온톨로지 용어 605개(60.5%)가 출현하였고 개념단위 분석결과, 596개 온톨로지 개념 중 434개(72.8%)의 개념이 소셜데이터에 등장한 것으로 나타났다. (Table 9)

Table 9. The frequency of terms and concepts appeared in the social data

	3,703,100 social data		161,581 social data	
	Term unit	Concept unit	Term unit	Concept unit
Freq=0	343(34.3)	131(21.7)	395(39.5)	162(27.2)
freq>0	657(65.7)	465(78.3)	605(60.5)	434(72.8)
Total	1000(100.0)	596(100.0)	1000(100.0)	596(100.0)

370만건의 소셜데이터에 출현한 온톨로지 개념 중 16만건의 소셜데이터에서 등장하지 않은 개념은 31개(SARI, SDQ, 규칙적 수면, 도파민 분비 이상 등)이다.

청소년이 작성한 것으로 판단되는 16만건의 소셜데이터는 370만건의 전체 데이터 중 4.3%에 불과하나 포함하고 있는 온톨로지 개념 수는 전체 소셜데이터와 비교하였을 때 큰 차이가 없었다.

## 2) 온톨로지 개념과 청소년상담기록 개념 매핑

청소년 우울 상담 기록 66건을 작성자에 따라 청소년 글, 부모 글, 교사 글, 상담가 글로 나누어 분석한 후 전체를 통합하여 분석하였다.

### (1) 청소년 글

청소년 글 21건 전문을 텍스트 파일로 전환한 후 R의 KoNLP패키지를 활용하여 자연어처리한 결과 277개의 용어가 출현하였다 (Figure 5).

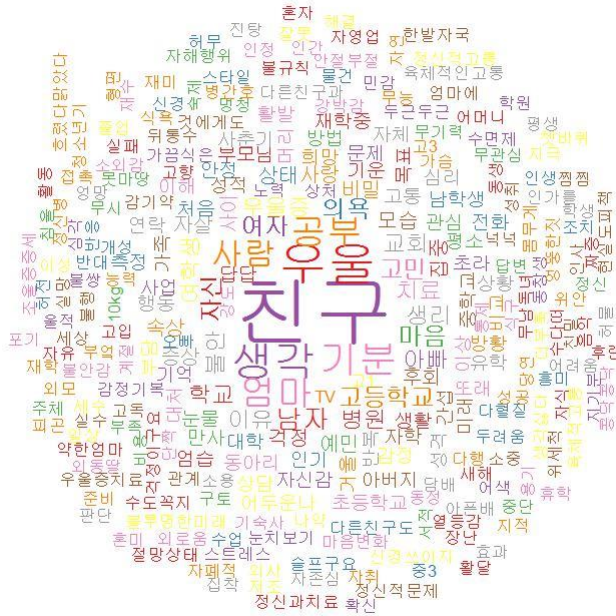


Figure 5. Terms extracted from youths' records

청소년 글에서 가장 많이 출현한 용어는 친구 총 63회, 우울 33회, 생각 26회, 기분 22회, 엄마 21회, 공부 20회 순으로 나타났다.

용어의 상담기록 내 의미를 분석하여 186개의 고유개념이 도출되었다. 도출된 186개 개념을 온톨로지 개념과 매핑한 결과, 언어 및 구조상으로 모두 일치하는 경우 (lexically mapped)가 43.5%, 의미상 일치하

는 경우(semantically mapped)가 38.2%, 보다 포괄적인 개념으로 매핑된 경우(mapped to a broader concept)가 11.3%, 보다 좁은 개념으로 매핑된 경우(mapped to narrower concept)가 1.6%, 한 개념이 여러 개념으로 나뉘어 매핑된 경우(one to many mapping)가 1.6%로 나타났다. 한편, 전혀 매핑되지 않은 개념은 3.8%로 계절변화·고민·방황·사춘기·생리·청소년기·향수병이었다. 이 개념들은 기존 온톨로지가 포함하지 않은 개념으로 최종 수정 사항에 반영하여 추가하였다.

Table 10. Mapping the concept of youth records to ontology concepts

Type of mapping	No. of unique concept (%)
<b>Completely mapped</b>	<b>152(81.7)</b>
Lexically mapped	81(43.5)
Semantically mapped	71(38.2)
<b>Partially mapped</b>	<b>27(14.5)</b>
Mapped to a broader concept	21(11.3)
Mapped to narrower concept	3(1.6)
One to many mapping	3(1.6)
<b>Not mapped</b>	<b>7(3.8)</b>
<b>Total</b>	<b>186 (100.0)</b>

## (2) 부모 글

부모가 작성한 자녀의 우울과 관련된 상담 기록 10건을 자연어처리한 결과 200개의 용어가 출현하였다. 부모 글에서 가장 빈번히 등장한 용어는 아이 49회, 친구 15회, 우울 14회, 걱정 12회 순으로 나타났다 (Figure 6).

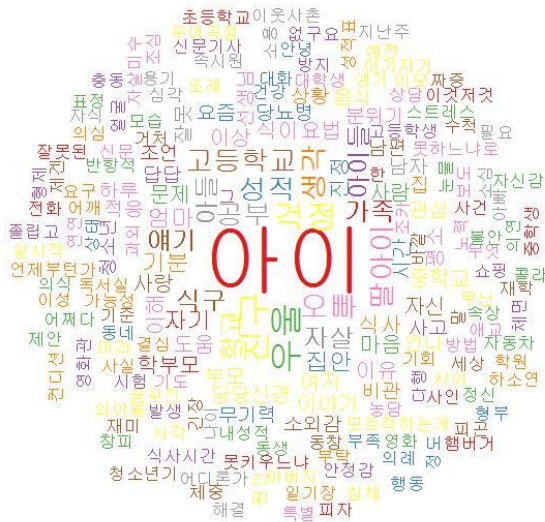


Figure 6. Terms extracted from parents' records

총 95개의 고유개념이 부모의 상담기록으로부터 도출되었고 이 개념과 온톨로지 개념을 매핑한 결과 언어 및 구조상으로 모두 일치하는 경우가 50.5%, 의미상 일치하는 경우가 30.5%, 보다 포괄적인 개념으로 매핑된 경우가 11.6%, 보다 좁은 개념으로 매핑된 경우가 1.1%, 한 개념이 여러 개념으로 나뉘어 매핑된 경우가 2.1%로 나타났다. 매핑되지

않은 4.2%는 온톨로지에 추가 반영하였다.

Table 11. Mapping the concept of parents' records to ontology concepts

Type of mapping	No. of unique concept (%)
<b>Completely mapped</b>	<b>77 (81.0)</b>
Lexically mapped	48 (50.5)
Semantically mapped	29 (30.5)
<b>Partially mapped</b>	<b>14 (14.8)</b>
Mapped to a broader concept	11 (11.6)
Mapped to narrower concept	1 (1.1)
One to many mapping	2 (2.1)
<b>Not mapped</b>	<b>4 (4.2)</b>
Total	95 (100.0)





Table 12. Mapping the concept of teachers' records to ontology concepts

Type of mapping	No. of unique concept (%)
<b>Completely mapped</b>	<b>20(86.9)</b>
Lexically mapped	9(39.1)
Semantically mapped	11(47.8)
<b>Partially mapped</b>	<b>2(8.7)</b>
Mapped to a broader concept	2(8.7)
Mapped to narrower concept	0(0.0)
One to many mapping	0(0.0)
<b>Not mapped</b>	<b>1(4.3)</b>
Total	23 (100.0)

#### (4) 상담가 글

청소년·부모·교사의 고민상담글에 대한 상담가의 답변글 33건을 자연어처리한 결과 사람 106회, 우울 104회, 생각 103회, 자신 88회, 우울증 54회, 친구 51회, 자녀 39회, 어머니 32회, 증상 32회 순서로 나타났다. Figure 8는 상담가 글에서 빈도 5회 이상 출현한 용어의 워드클라우드이다.



Figure 8. Terms extracted from counselors' records(freq≥5)

총 499개의 용어 중 293개의 고유개념이 도출되었고 이를 온톨로지 개념과 매핑한 결과 언어 및 구조상으로 모두 일치하는 경우가 37.2%, 의미상 일치하는 경우가 29.7%, 보다 포괄적인 개념으로 매핑된 경우가 17.4%, 보다 좁은 개념으로 매핑된 경우가 3.4%, 한 개념이 여러 개념으로 나뉘어 매핑된 경우가 4.1%로 나타났다. 매핑되지 않은 8.2%는

온톨로지에 추가 반영하였다.

Table 13. Mapping the concept of counselors ‘records to ontology concepts

Type of mapping	No. of unique concept (%)
<b>Completely mapped</b>	<b>196(66.9)</b>
Lexically mapped	109(37.2)
Semantically mapped	87(29.7)
<b>Partially mapped</b>	<b>73(24.9)</b>
Mapped to a broader concept	51(17.4)
Mapped to narrower concept	10(3.4)
One to many mapping	12(4.1)
<b>Not mapped</b>	<b>24(8.2)</b>
Total	293(100.0)

(5) 전체

66건의 상담기록에서 총 774개의 용어가 출현했으며 476개의 고유 개념이 도출되었다. 476개의 고유개념을 온톨로지 내 개념과 매핑한 결과 언어 및 구조상으로 모두 일치하는 경우는 37.4%, 의미상 일치하는 경우 35.7%, 보다 포괄적인 개념으로 매핑되는 경우 15.1%, 보다 좁은 개념으로 매핑되는 경우 2.7%, 한 개념이 여러 개념으로 나뉘어 매핑된 경우 3.2%, 어느 개념에도 매핑되지 않는 경우가 5.9%로 나타났다 (Table 14).

Table 14. Mapping the concept of counseling records to ontology concepts

Type of Mapping	No. of Terms	No. of Concept (%)	No. of Unique concept (%)
<b>Completely mapped</b>		<b>529(74.9)</b>	<b>348(73.1)</b>
Lexically mapped		301(42.6)	178(37.4)
Semantically mapped		228(32.3)	170(35.7)
<b>Partially mapped</b>		<b>136(19.3)</b>	<b>100(21.0)</b>
Mapped to broader concept		100(14.2)	72(15.1)
Mapped to narrower concept		16(2.3)	13(2.7)
One to many mapping		20(2.8)	15(3.2)
<b>Not mapped</b>		<b>41(5.8)</b>	<b>28(5.9)</b>
<b>Total</b>	<b>774<sup>1)</sup></b>	<b>706(100.0)</b>	<b>476(100.0)</b>

1) 청소년, 부모, 교사, 상담가 기록 내 중복되어 나타난 용어는 제거하였음.

### 3) 온톨로지 구조 및 표현력 평가

온톨로지 전체 구조 평가 결과 문항별 평균은 개념·관계의 수, 계층구조의 깊이, 계층구조의 폭, 개념간 관계 수, 균형감, 전반적인 구조의 복잡성, 개념간 연결성의 문항별로 각각 4.75, 5.00, 4.75, 4.50, 5.00, 5.00, 4.75점으로 나타났다.

7개 항목 모두 4점 이상을 받았으며, 계층구조의 깊이·균형감·전반적 구조의 복잡성 항목의 경우 평가자 전원에게 5점을 받았다.(Table 15).

Table 15. Evaluation Scores of the structural layer (N=4)

Item	Score Mean	Minimum	Maximum
Size	4.75	4.00	5.00
Depth of hierarchy	5.00	5.00	5.00
Breadth of hierarchy	4.75	4.00	5.00
Density	4.50	4.00	5.00
Balance	5.00	5.00	5.00
Overall complexity	5.00	5.00	5.00
Connectivity between concepts, relations	4.75	4.00	5.00

반면, 5개 영역별로 시행한 온톨로지 표현력 평가에서 5개 영역 모두 형식적 정의의 인지적 수용 여부, 형식적 정의의 일관성, 문맥 및 배경적 지식의 명확성, 개념 표현의 명시성, 개념의 정의 및 설명의 해석력, 온톨로지 포함력의 상세성 총 6개 문항에서 평가자 6인에게 4점 이상을 받았다 (Table 16).

평가 설문지를 통한 평가 이외에 기타 체언 및 건의 사항에 대한 의견을 개방형 질문으로 받은 결과, 평가자 2인에게서 규칙적인 운동의 하위클래스(예: 자전거타기, 수영, 등산 등)와 취미활동 클래스의 하위클래

스(예: 전시회감상, 영화감상 등)의 경우 다른 클래스에 비해 세분화되어 있다는 의견, 증상 및 징후 내 클래스인 잦은 울음의 경우 '잦은'이라는 형용사가 빈도(frequency) 속성과 중복적인 의미를 가지므로 클래스명을 울음으로 수정하자는 의견, 위험요인 내 성관계경험 클래스에 파트너 연령 속성을 추가하여 원조교제 여부를 파악할 수 있는 정보를 추가해야 한다는 의견이 있었다.

또한 일부 클래스의 표현 방식에 수정이 필요하여(예 : 정신분열병→조현병) 이를 최종 온톨로지에 반영하였다.

Table 16. Evaluation scores of the representational layer (N=6)

	Risk Factors			Sign & Symptoms			Measurement			Diagnostic Result			Management Care		
	Mean	Min	Max	Mean	Min	Max	Mean	Min	Max	Mean	Min	Max	Mean	Min	Max
<b>Match</b>	4.83	4.00	5.00	4.83	4.00	5.00	4.67	4.00	5.00	4.67	4.00	5.00	4.67	4.00	5.00
<b>Consistency</b>	4.67	4.00	5.00	4.50	4.00	5.00	4.67	4.00	5.00	4.67	4.00	5.00	4.67	4.00	5.00
<b>Clarity</b>	4.67	4.00	5.00	4.67	4.00	5.00	4.83	4.00	5.00	4.67	4.00	5.00	4.67	4.00	5.00
<b>Explicitness</b>	4.67	4.00	5.00	4.83	4.00	5.00	4.67	4.00	5.00	4.67	4.00	5.00	4.83	4.00	5.00
<b>Interpretability</b>	4.83	4.00	5.00	4.50	4.00	5.00	4.67	4.00	5.00	4.83	4.00	5.00	4.67	4.00	5.00
<b>Accuracy</b>	4.33	2.00	5.00	4.50	3.00	5.00	4.67	4.00	5.00	4.67	4.00	5.00	4.50	4.00	5.00
<b>Comprehensiveness</b>	4.33	3.00	5.00	4.67	4.00	5.00	4.83	4.00	5.00	4.83	4.00	5.00	4.50	4.00	5.00
<b>Granularity</b>	4.67	4.00	5.00	4.67	4.00	5.00	4.83	4.00	5.00	4.83	4.00	5.00	4.83	4.00	5.00
<b>Relevance</b>	4.50	3.00	5.00	4.67	4.00	5.00	4.50	4.00	5.00	4.50	4.00	5.00	4.67	4.00	5.00
<b>Description</b>	4.83	4.00	5.00	4.67	4.00	5.00	4.83	4.00	5.00	4.67	3.00	5.00	4.67	4.00	5.00

## V. 논의

### 1. 청소년 우울 온톨로지 개발

청소년 건강문제는 신체적·정신적·사회적 문제와 관련되어 있으며 개인적 특성을 비롯해 가족·학교·지역사회의 환경적 특성과 밀접하게 연계되어 있다(정호지, 박미정, 서상훈, 신의진, & 한태룡, 2011). 따라서 청소년 우울은 개인 차원에서부터 가족·학교·지역사회 차원까지 통합적 관점에서 분석되어야 하며 다차원적인 해결방안이 모색되어야 한다. 본 연구에서는 이러한 관점을 반영하여 개인·가족·학교·지역사회 차원에서 청소년 우울의 위험요인과 관리방안을 설명하는 온톨로지를 개발하였다.

온톨로지 개발 방법에는 Cyc, KACTUS, METHONTOLOGY, TOVE, OTKM, DOLCE, Ontology Development 101 등이 있으나 Ontology Development 101을 제외하고 온톨로지 구축 계획·설계·유지관리 등 대용량 온톨로지 구축을 위한 방법론(장우혁, 2015)으로서 본 연구에서 개발하고자 하는 온톨로지의 사용 목적과 범위를 벗어나는 방법론이다. 따라서 청소년 우울 온톨로지는 비교적 작은 규모로 쉽게 구축이 가능한 Ontology Development 101 방법론의 7단계에 근거하여 개발되었다.

청소년 우울 온톨로지의 용어와 주요 개념을 수집하는 과정에서 영국 NICE(2005, 2015), 미국 USPSTF(2009), 호주 beyondblue(2011) 및 국내 우울증임상연구센터(2011)의 임상실무지침서를 참조하였다. 국내 청소년 우울의 현황과 고유한 특성을 파악하기 위해 국내 우울증임상연구센터의 임상실무지침서를 우선 선정하였고 국내 임상실무지침서 개발 시 모태가 된 영국 NICE(2005)와 미국 USPSTF(2009)의 임상



실무지침서를 포함하였다. 마지막으로 본 연구에서 개발하고자 하는 청소년 우울 온톨로지의 5개 영역을 포괄적으로 기술한 호주 beyondblue(2011)의 임상실무지침서를 추가 선정하였다. 이를 바탕으로 용어를 추출하던 중 NICE 임상실무지침서가 개정되어 개정판을 최종 포함시켰다.

임상실무지침서를 중심으로 추출한 용어 중에서 클래스를 정의하고 분류체계를 정립하는 과정에서는 스트레스 온톨로지(Khoozani & Hadzic, 2010), 정신건강 온톨로지(Hadzic et al., 2008), 만성질환 재택관리를 위한 온톨로지(Riaño et al., 2012)를 참고하였다. 스트레스 온톨로지는 스트레스를 유발하는 위험요인(원인 및 매개요소)을 시작으로 스트레스를 진단(측정)하고 치료하는 일련의 과정과 함께 스트레스로 인한 인지적·정서적·물리적 기능 변화를 설명하고 있으며 만성질환 재택관리를 위한 온톨로지 역시 질환(진단결과)별 증상 및 징후, 사정 방법(진단방법), 중재(치료) 등의 카테고리로 구성되어 있다. 따라서 본 연구에서는 이를 근거로 청소년 우울의 **위험요인, 증상 및 징후, 진단결과, 진단방법, 치료 및 관리 영역**으로 온톨로지를 개발하였고 영역별 클래스와 분류체계를 정의하였다.

한편, 용어를 클래스로 정의하는 단계에서 여러 번의 수정이 있었다. 예를 들어, 치료 및 관리 영역 중 자전거 타기, 수영과 같은 구체적인 운동 종류를 클래스로 정의할 것인지 상위 클래스인 규칙적인 운동의 속성값으로 정의할 것인지에 대해 여러 번의 수정을 거쳤다. (Noy & McGuinness, 2001)는 개념을 클래스 또는 속성값으로 정의하는 문제에 있어 정답은 해당 영역의 관심 범위(scope)에 달려 있다고 하였다. 자전거 타기 및 수영과 같은 구체적인 운동 종류들이 다른 개념들과 관계를 맺으면서 고유한 의미를 갖는다면 클래스로 정의하고 그렇지 않으

면 상위클래스의 속성값으로 정의하는 것이 바람직한데, 본 온톨로지에 서는 규칙적인 운동 수행 자체가 구체적인 운동 종류보다 청소년 우울 을 관리하는데 중요한 의미를 가지므로 최종적으로 상위 클래스의 속성 값으로 분류하였다.

## 2. 청소년 우울 온톨로지 평가

### 1) 온톨로지 용어의 소셜데이터 내 출현 빈도

우울 관련 소셜데이터를 대상으로 온톨로지 용어 1,000개의 출현여 부를 확인한 결과 65.7%의 용어가 수집된 소셜데이터 내에서 출현하였 다. 출현하지 않은 34.3%의 용어를 분석하는 과정에서, 동의어 및 유사 어의 소셜데이터 내 출현 여부를 고려하여 개념단위로 재 분류하였다. 예를 들어, 용어단위 분석 시 흡입약은 소셜데이터 상에서 출현하지 않 은 반면, 흡입약의 동의어인 흡입제는 출현한 것으로 나타났다. 온톨로 지는 두 용어를 하나의 개념으로 처리함으로써 이음동의어를 정교하게 처리할 수 있다는 특징을 가지므로(유은지 et al., 2012), 흡입약과 흡입 제는 동일한 개념으로 처리되어야 한다.

용어단위 분석 시 소셜데이터에 출현하지 않은 343개의 용어(34.3% 에 해당)를 개념단위로 재분석한 결과, 126개 용어가 소셜데이터에 출현 한 것으로 분류되어 온톨로지 용어의 소셜데이터 출현율이 78.3%로 증 가하였다.

한편, 개념단위 분석으로도 소셜데이터 상에서 출현하지 않은 것으로 분류된 176개 용어를 확인한 결과 우울증의 생물학적 및 생화학적 위험 요인(13.6%, 24개), 항우울제를 포함한 전문의약품명(9.7%, 17개), 우 울증 스크리닝 도구명(8.5%, 15개) 등 전문용어가 대다수 포함되어 있 었다. 이는 임상실무지침서를 비롯해 관련 논문을 바탕으로 온톨로지 용

어가 수집됨으로써 대중들이 소셜미디어상에서 흔히 사용하는 용어를 반영하지 못한 결과로 해석된다. 그러나 전문지식이 대중화됨에 따라 향후 전문용어가 소셜미디어 내에 출현할 가능성이 높으므로 온톨로지는 소셜데이터의 분석틀로써 포괄적인 내용을 다루어야 한다.

본 연구에서는 한국보건사회연구원 통계정보연구실로부터 온톨로지 용어의 소셜데이터 내 출현 여부 및 빈도수 자료만을 제공받아 분석함으로써 원시자료(raw data)에 대한 내용과 사용된 온톨로지 용어의 의미 분석이 함께 이루어지지 않았다는 한계를 가진다.

## 2) 온톨로지 개념과 청소년상담기록 개념 매핑

자연어처리를 거쳐 청소년 우울 상담기록으로부터 출현된 774개 용어(중복제거)는 용어자체로서도 의미 있으나 상담기록 문맥상 용어가 어떤 의미로 쓰인 것인지를 파악하는 것이 중요하다. 예를 들어, 친구라는 용어는 상담기록 내에서 '친구와 있어도 즐겁지 않고'나 '친구가 유학을 갔거든요'와 같이 서로 다른 의미를 전달하기 위해 쓰였는데, 전자의 경우 즐거움을 상실했다는 핵심 의미를 전달하기 위해 친구가 수식어로 사용됐음을 알 수 있으며 후자의 경우 친구와의 이별이라는 부정적 사건을 표현하기 위해 친구 용어가 사용되었음을 알 수 있다.

다시 말해, 전자는 친구라는 용어를 사용함으로써 우울의 주요 증상인 즐거움 상실을, 후자는 우울을 유발하는 위험 요인 중 친구와의 이별이라는 개념을 언급하고 있다. 따라서 같은 용어라 할지라도 다른 목적을 가지고 상이한 의미를 전달하고 있으므로 이를 구분하여 용어가 말하고자 하는 핵심 개념(key concept)을 도출하는 것이 중요하다. 상담기록 문맥상 용어의 의미를 바탕으로 추출한 핵심 개념과 온톨로지 개념을 매핑한 결과 언어 및 구조상으로 모두 일치하는 경우는 전체 글을 대상

으로 분석한 결과 37.4%로 나타났다. 의미상 일치하는 경우는 35.7%로 해당되는 개념 모두 온톨로지 개념의 유사어·동의어에 추가함으로써 용어체계를 수정 보완하였다. 한편, 기존의 온톨로지 구조와 내용·의미 안에서 어느 개념에도 매핑되지 않은 5.9%의 경우는 온톨로지 구조를 수정하거나 새로운 클래스를 생성함으로써 포함할 수 있었다. 예를 들어 기후변화(계절변화)의 경우, 임상실무지침서 및 관련문헌에서 추출할 수 없었으나 상담기록으로부터 도출된 우울증의 위험요인이다. 그러나 개인요인/가정요인/학교요인/지역사회요인으로 분류하여 정의한 기존의 위험요인 영역 내에서는 기후변화(계절변화)를 포함할 수 없어 기타요인을 추가함으로써 이를 포함하였다. 마찬가지로 여성의 월경 또한 기존 문헌에서 발견할 수 없었던 우울증의 위험요인이었지만 상담기록으로부터 새로 도출되어 위험요인 영역 내 개인요인 중 생물학적 요인 안에 클래스를 추가하여 포함시켰다. 이밖에 보다 포괄적/세부적인 개념으로 매핑되는 경우를 비롯해 여러 개념으로 나뉘어 매핑된 경우도 온톨로지 구조를 일부 수정함으로써 평가 결과를 반영하였다.

### 3) 온톨로지 구조 및 표현력 평가

온톨로지 전체 구조와 5개 영역별 표현력에 대한 전문가 평가에서 항목 모두 평균 4점 이상을 받아 본 연구에서 개발한 온톨로지가 청소년 우울 영역을 설명하는데 비교적 우수한 것으로 나타났다. 그러나 위험요인 영역의 경우, 표현력 문항 중 accuracy와 comprehensiveness 문항에서 다른 영역 보다 상대적으로 낮은 점수(평균 4.33)를 받았는데 이는 위험요인 영역을 표현하는 용어가 소셜미디어 상에서 사용되는 용어라고 보기 어려운 측면이 일부 존재하였고 온톨로지 도메인의 범위 중 정신질환 가족력 클래스에서 조현병 가족력이 누락되어 있다는 점을 지

적 받았기 때문이다.

또한 전문가 의견을 바탕으로 치료 및 관리 영역 내 클래스 수준을 일부 조정(하위클래스를 상위클래스의 속성값으로 이동)함으로써 온톨로지 구조의 균형감을 높이고 온톨로지가 다양한 개념을 포괄적으로 수용할 수 있도록 향상시켰다.

## VI. 결론 및 제언

최근 보건의료영역에서도 SNS 또는 소셜미디어 상의 메시지를 분석하여 전통적인 분석방법에서 얻을 수 없었던 가치 있는 정보와 새로운 통찰력을 발견하려는 시도들이 늘어나고 있다. 이러한 과정에서 온톨로지는 해당영역의 데이터와 지식을 결합함으로써 데이터를 의미 있게 해석하는 분석틀으로써 중요한 역할을 하고 있다(Kuiler, 2014).

본 연구에서는 우리 사회의 심각한 문제 중 하나인 청소년 우울과 관련하여 소셜미디어상의 데이터를 수집·분석하기 위해 필요한 온톨로지를 개발하였다. 온톨로지는 Ontology Development 101 방법론에 근거하여 개발하였으며 온톨로지를 구성하는 용어는 영국·미국·호주·한국의 임상실무지침서를 비롯해 관련 문헌을 고찰하여 수집하였다. 그 결과, 총 2,073개의 용어가 수집되었고 452개의 클래스, 19개 유형의 관계, 1,228개의 동의어 및 유사어로 구성된 용어체계가 개발되었다.

온톨로지는 온톨로지 용어의 소셜데이터 내 출현빈도, 온톨로지 개념과 청소년 상담기록 개념 매핑, 전문가 설문을 통한 온톨로지 구조 및 표현력 평가를 통해 타당성을 검증하였다. 평가결과, 청소년 우울 온톨로지는 소셜데이터를 비롯해 청소년 우울 상담기록 내 개념을 상당부분 반영(소셜데이터 78.3%, 상담기록 73.1%)하고 있었으며 온톨로지의 구조와 표현력이 우수한 것으로 나타났다. 평가결과를 반영하여, 온톨로지 개념의 유사어와 동의어를 추가함으로써 용어체계를 보강하였고 온톨로지의 포함력을 확대하기 위해 일부 구조를 수정하였다.

본 연구에서 개발한 온톨로지는 개인·가족·학교·지역사회 차원의 통합적 관점을 반영한 최초의 청소년 우울 온톨로지이며, 전문가 평가라는 전통적인 온톨로지 평가 방법 이외에 자연어처리기법을 이용한 키워

드 분석과 개념 매핑을 적용하여 온톨로지의 포함력을 평가했다는 점에서 의미가 있다. 최근, 소셜 빅데이터를 활용한 우리나라 청소년의 우울 현황과 위험요인을 분석한 연구(송태민, 2015)에서는 본 연구에서 개발한 온톨로지를 청소년 우울 관련 소셜 빅데이터의 분석틀로 활용하여 주제분석과 감성분석을 실시하였다. 연구 결과, 청소년 우울과 관련하여 부정적인 감정을 나타내는 문서의 발생 추이(2012-2015년)가 청소년 건강행태 온라인조사의 우울감 경험률 추이와 비슷한 것으로 나타나 소셜 빅데이터 분석이 청소년 우울의 현황을 파악하는 정부통계조사 결과를 보완하고 향후 청소년 우울 발병을 조기에 예측하는 조사 도구로써 활용가능성이 높음을 보여주었다. 송태민(2015)의 연구에서처럼 본 연구에서 개발한 청소년 우울 온톨로지는 해당 주제의 소셜 데이터를 체계적으로 분류하고 분석하는데 이론적 근거를 제공할 뿐 아니라 세종 전자사전과 같은 기존의 용어사전이 포함할 수 없었던 특정 주제와 관련된 다양한 동의어와 유사어를 포함함으로써 소셜 빅데이터의 정교한 분석을 가능하게 할 것으로 기대된다.

한편, 본 연구는 다음과 같은 제한점을 가진다.

첫째, 온톨로지 개발과정에서 BFO(Basic Formal Ontology), DOLCE(Descriptive Ontology for Linguistic and Cognitive Engineering)와 같은 상위 온톨로지의 사용을 검토하지 못하였다. 본 연구에서 개발한 청소년 우울 온톨로지와 기존에 개발된 혹은 앞으로 개발될 다양한 온톨로지와의 상호운용성을 향상시키기 위해서는 상위온톨로지를 기반으로 본 온톨로지를 재설계하는 작업이 필요하다.

둘째, 청소년 우울 상담 기록을 자연어처리하는 과정에서 활용한 R의 KoNLP패키지는 세종 전자 사전에 등록된 용어를 바탕으로 분석하므로 사전에 등록되지 않은 신조어 및 청소년 은어는 연구자가 직접 프로그램에 입력하였다. 이 과정에서 추가되지 않은 용어는 자연어처리 과정에서

분석되지 않아 개념 매핑이 이루어지지 않았을 것으로 예상된다.

셋째, 소셜데이터 수집 시 우울 혹은 우울증 키워드를 포함하고 있는 전체 문서를 대상으로 하여 우울과 관련이 적은 포스팅이 평가 대상으로 포함되었을 가능성이 크다.

이러한 제한점에도 불구하고 본 연구는 소셜빅데이터 분석을 위해 온톨로지 중심의 분석틀을 개발하고 실제 분석에 적용하여 유의한 결론을 도출한 최초의 시도이며 청소년 자살, 청소년 비행과 같은 개인·가족·학교·지역사회 차원의 통합적 관점이 필요한 다양한 청소년 관련 문제에 일반화되어 사용되는 등 활용가치가 높을 것으로 기대된다..



## 참고문헌

- Beyondblue. (2011). *Clinical Practice Guidelines: Depression in adolescents and young adults*.
- Chang, Y.-S., Hung, W.-C., & Juang, T.-Y. (2013). *Depression Diagnosis based on Ontologies and Bayesian Networks* Paper presented at the 2013 IEEE International Conference on Systems, Man, and Cybernetics.
- Chunara, R., Andrews, J. R., & Brownstein, J. S. (2012). Social and news media enable estimation of epidemiological patterns early in the 2010 Haitian cholera outbreak. *The American Journal of Tropical Medicine and Hygiene*, *86*(1), 39–45.
- Din, M. A., Abidi, S. S. R., & Jafarpour, B. (2010). Ontology based modeling and execution of Nursing Care Plans and Practice Guidelines. *Studies in Health Technology and Informatics*, *160*(2), 1104–1108.
- Elkin, P. L., Brown, S. H., Husser, C. S., Bauer, B. A., Wahner-Roedler, D., Rosenbloom, S. T., & Speroff, T. (2006). Evaluation of the Content Coverage of SNOMED CT: Ability of SNOMED Clinical Terms to Represent Clinical Problem Lists. *Mayo Clinic Proceedings*, *81*(6), 741–748.
- Eun-Young So, & Park, H.-A. (2011). Exploring the Possibility of Information Sharing between the Medical and Nursing Domains by Mapping Medical Records to SNOMED CT and ICNP. *Health Informatics Research*, *17*(3), 156–161.

- H.A.Park, Lundberg, C. B., Coenen, A., & Konicek, D. J. (2009). Evaluation of the content coverage of SNOMED-CT to represent ICNP Version1 catalogues. *Studies in Health Technology and Informatics*, 146, 303–307.
- H.Y.Kim, H.A.Park, Y.H.Min, & E.Jeon. (2013). Development of an obesity management ontology based on the nursing process for the mobile-device domain. *Journal of Medical Internet Research*, 15(6).
- Hadzic, M., Chen, M., & Dillon, T. S. (2008). Towards the Mental Health Ontology. *IEEE International Conference on Bioinformatics and Biomedicine*, 284–288.
- Haghighi, M., Koeda, M., Takai, T., & Tanaka, H. (2009). Development of clinical ontology for mood disorder with combination of psychomedical information. *Journal of Medical and Dental Sciences*, 56(1), 1–15.
- Harris, J. K., Moreland-Russell, S., Tabak, R. G., Ruh, L. R., & Maier, R. C. (2014). Communication About Childhood Obesity on Twitter. *American Journal of Public Health*, 104(7), 62–69.
- Kehagias, D. D., Papadimitriou, I., Hois, J., Tzovaras, D., & Bateman, J. (2008). A Methodological Approach for Ontology Evaluation and Refinement. *ASK-IT International Conference*.
- Khoozani, E. N., & Hadzic, M. (2010). Designing the human stress ontology: A formal framework to capture and represent knowledge about human stress. *Australian Psychologist*,

45(4), 258–273.

- Kuiler, E. W. (2014). From Big Data to Knowledge: An Ontological Approach to Big Data Analytics. *Review of Policy Research*, 41(4), 311–318.
- NICE. (2005). *Depression in children and young people: Identification and management in primary, community and secondary care.*
- NICE. (2015). *Depression in children and young people: Identification and mangement in primary, community and secondary care.*
- Noy, N. F., & McGuinness, D. L. (2001). *Ontology Development 101: A Guide to Creating Your First Ontology.*
- Park, I., Shin, H., Kim, E., Park, H., Kim, Y., & Jo, E. (2006). Mapping Nursing Statements with the ICNP and Its Practical Use in Electronic Nursing Records. *Studies in Health Technology and Informatics*, 122, 989–990.
- Riaño, D., Real, F., López–Vallverdú, J. A., Campana, F., Ercolani, S., Mecocci, P., . . . Caltagirone, C. (2012). An ontology–based personalization of health–care knowledge to support clinical decisions for chronically ill patients. *Journal of Biomedical Informatics*, 45, 429–446.
- Spasic, I., Ananiadou, S., McNaught, J., & Kumar, A. (2005). Text mining and ontologies in biomedicine: making sense of raw text. *Brief Bioinformatics*, 6(3), 239–251.
- USPSTF. (2009). *Screening for Child and Adolescent Depression In*

*Primary Care Settings.*

- 권호인. (2012). 아동 청소년기 우울증의 예방 개입에 대한 고찰. *인지 행동치료, 12*(1), 21-41.
- 김흥기, & 김명기. (2003). 의료정보학에서의 온톨로지 기술. *대한의료 정보학회지, 9*(3), 213-219.
- 노상규, & 박진수. (2007). *인터넷 진화의 열쇠 온톨로지*. 서울: 가즈토이.
- 박정은, 김화선, 장민정, & 홍해숙. (2014). 항고혈압제-식이 상호작용 관리를 위한 온톨로지 기반의 임상 의사결정 지원 시스템 구현. *대한간호학회지, 44*(3), 294-304.
- 박정은, 정귀애, 조훈, & 김화선. (2013). 산부인과 간호단위의 간호과정과 SNOMED CT를 이용한 간호진단 온톨로지의 구축. *여성건강 간호학회지, 19*(1), 1-12.
- 백현기. (2012). 청소년 상담을 위한 마음의 병 온톨로지 설계에 관한 연구. *Journal of Digital Contents Society, 13*(4), 547-557.
- 보건복지부. (2005). *자살예방 5개년 종합대책*.
- 송태민. (2012). 빅데이터를 활용한 자살요인 다변량 분석: Google 검색 트렌드 적용.
- 송태민. (2015). *소셜 빅데이터를 활용한 우리나라 청소년의 우울 현황과 위험요인 분석*.
- 송태민, & 송주영. (2014). *빅데이터 연구 한 권으로 끝내기: 한나래출판사*.
- 우울증임상연구센터. (2011). *일차의료기관에서 우울증 선별에 관한 진료지침*.
- 유은지, 김정철, 이춘열, & 김남규. (2012). 시맨틱 텍스트 마이닝을 위한 온톨로지 활용 방안. *정보시스템연구, 21*(3), 137-161.

- 이인근, 김화선, & 이성희. (2012). 온톨로지에 기반한 간호진단 지식모델의 설계. *Journal of Korean Institute of Intelligent Systems*, 22(4), 468--475.
- 장우혁. (2015). 무기체계 부품국산화 정보의 온톨로지 구축방안 연구. *Journal of Korea Multimedia Society*, 18(7), 873-885.
- 전영주. (2001). 우울증을 매개로 한 청소년의 자살구상에 관한 가족 및 학교환경의 경로분석 모델. *대한가정학회지*, 39(1), 151-167.
- 정효지, 박미정, 서상훈, 신의진, & 한태룡. (2011). 청소년 건강증진 중 합대책 방안 연구.
- 조성연, 김혜원, & 김민. (2010). 부모-자녀의사소통 방식과 학교생활만족도가 청소년의 자살생각에 미치는 영향과 우울의 중재효과. Paper presented at the 한국가정관리학회 2010년 춘계학술대회.
- 최규련. (2010). 가족환경과 학업 및 친구요인이 청소년의 우울에 미치는 영향. *한국가정관리학회지*, 28(6).
- 최인재, 모상현, 강지현, & 김윤희. (2011). 아동청소년 정신건강 증진을 위한 지원방안 연구 1. 서울:
- 통계청, 여. (2014). *2014 청소년 통계*.
- 한국정보화진흥원. (2012). 소셜 분석으로 살펴본 청소년 자살예방정책의 시사점. *IT&Future Strategy*, 17.

## 부록1. 생명윤리위원회 (IRB) 심사결과 통지서

### 심의면제 통보서

#### 수신

책임연구자	이름: 정혜실	소속: 서울대학교 간호학과	직위: 석사과정
지원기관	해당없음		

#### 과제정보

승인번호	IRB No. E1507/001-008
연구과제명	청소년 우울 관련 소셜데이터 수집과 분석을 위한 온톨로지 개발 및 평가
연구종류	학위 논문 연구, 설문조사, 공개된 정보를 이용하는 연구
면제일자	2015-07-06
검토의견	본 연구는 대면하지만 개인식별정보를 수집 기록하지 않는 설문 연구로 「생명윤리 및 안전에 관한 법률」 시행규칙 제13조 제1항 제2호에 근거하여 심의를 면제함. 다만 소셜데이터를 분석하는 연구의 경우에는 연구자에게 연구 참여자 개인에 대한 자료가 전달되는 것이 아니고, 용어의 출현 여부 와 빈도표 만이 전달되기에 인간을 대상으로 하는 연구가 아닌 것으로 판단함.
심의결과	면제승인

상기 연구과제에 대하여 본 위원회에서는 심의면제대상임을 확인합니다.

모든 연구자들은 아래의 사항을 준수하여야 합니다.

1. 연구자께서는 제출하신 계획서에 따라 연구를 수행하여야 하며, 이와 다르게 연구를 진행하실 경우 다시 심의를 진행하셔야 함을 유의하시기 바랍니다.
2. 위원회의 요구가 있을 때에는 연구의 진행과 관련된 보고를 위원회에 제출하여야 합니다.
3. 연구윤리를 위하여 관련부처가 필요시 조사 및 감독 차원에서 현장점검을 실시할 수 있습니다.
4. 연구와 관련된 기록은 연구가 종료된 시점을 기준으로 최소 3년간 보관하여야 합니다.

2015년 07월 06일

서울대학교 생명윤리위원회 위원장



## 부록2. 온톨로지 평가용 설문지

### 전문가 평가 설문지

**연구 제목 : 청소년 우울 관련 소셜데이터 수집과 분석을 위한  
온톨로지 개발 및 평가**

본 연구에서는 청소년 우울과 관련된 소셜데이터를 수집하고 분석하기 위한 분석틀로써 청소년 우울 온톨로지를 개발하였습니다. 온톨로지 개발에 경험이 있는 전문가를 대상으로 본 연구에서 개발된 온톨로지의 정확성과 타당성을 평가하고자 합니다.

본 평가는 크게 세 개의 주제로 구성되어 있습니다.

먼저, 온톨로지 모형의 구조(graph structure)를 평가하고 온톨로지 내 도메인과 컨셉, 컨셉들 간의 관계를 5점 리커트 척도로 평가해주시시오. 마지막으로 온톨로지 개발 전반에 대한 조언을 자유롭게 기술해주시면 됩니다.

본 설문과 관련한 문의사항은 02-740-8838 (정혜실, 간호정보학 석사과정)로 주시기 바랍니다. 감사합니다.

온톨로지 구조는 5개 주제에 대한 전체 구조를 평가해주시시오.

평가항목	항목 해설	동의하지않음		보통임		동의함	
		1	2	3	4	5	
<b>Structural/architectural layer</b>							
Size	컨셉과 관계의 수가 적절한가? -총 컨셉 : 452개, 총 관계 : 19개 유형, 478개						
Depth of hierarchy	주제별 계층구조의 깊이가 적절한가? -위험요인 : 5단계 -중상및중후 : 3단계 -측정 : 2단계 -진단결과 : 2단계 -관리 : 5단계						
Breadth of	주제의 폭이 청소년 우울을 설명하는데 적절한가? -위험요인 : 개인/가족/학교/지역사회/기타요인						

평가항목	항목 해설	동의하지않음		보통임		동의함	
		1	2	3	4	5	
hierarchy	-증상및증후:감정변화/행동변화/인지변화/신체변화 -측정 : 임상가측정도구/자기측정도구 -진단결과 : 정상/우울증 -관리 : 개인/가족/학교/지역사회/전문의료기관						
Density	컨셉을 연결하는 평균 가짓수(average branching)? -is-a관계 총 개수:889개 -p/r 총 개수:52개 -총 941개, 개념당 평균 가짓수 2.08개						
Balance	5개 주제가 고르게 개발되었는가? 특정 주제에 치우쳐 개발되진 않았는가?						
Overall complexity	전반적인 구조의 복잡성은 적절한가?						
Connectivity between concepts, relations	개념간 연결이 잘 맺어져 있는가?						

온톨로지 표현력은 5개 주제 각각에 대해 평가해주시기 바랍니다.

① 진단결과

평가항목	항목 해설	동의하지않음		보통임		동의함	
		1	2	3	4	5	
<b>Representational/Semantic layer</b>							
Match between formal and cognitive semantics	온톨로지 개념의 형식적 정의가 인지적으로 받아들여지는가						
Consistency	형식적 정의의 모순이 없는가?						
Clarity	문맥 및 배경적 지식이 명확히 드러나 있는가						
Explicitness	개념이 분명하게 표현되었는가						
Interpretability	개념의 정의와 설명이 잘 이해 되는가						
Accuracy	개념이 소셜미디어에서 사용되는 용어라고 할 수 있는가						
Comprehensiveness	다루는 도메인의 범위가 포괄적인가						



평가항목	항목 해설	동의하지않음		보통임		동의함	
		1	2	3	4	5	
Granularity	온톨로지가 상세수준까지 커버하고 있는가						
Relevance	온톨로지가 개발 목적에 잘 부합하는가						
Description	온톨로지가 청소년 우울의 진단결과 영역을 잘 기술하고 있는가						

② 측정

평가항목	항목 해설	동의하지않음		보통임		동의함	
		1	2	3	4	5	
<b>Representational/Semantic layer</b>							
Match between formal and cognitive semantics	온톨로지 개념의 형식적 정의가 인지적으로 받아들여지는가						
Consistency	형식적 정의의 모순이 없는가?						
Clarity	문맥 및 배경적 지식이 명확히 드러나 있는가						
Explicitness	개념이 분명하게 표현되었는가						
Interpretability	개념의 정의와 설명이 잘 이해 되는가						
Accuracy	개념이 소셜미디어에서 사용되는 용어라고 할 수 있는가						
Comprehensiveness	다루는 도메인의 범위가 포괄적인가						
Granularity	온톨로지가 상세수준까지 커버하고 있는가						
Relevance	온톨로지가 개발 목적에 잘 부합하는가						
Description	온톨로지가 청소년 우울의 측정방법 영역을 잘 기술하고 있는가						

③ 증상 및 증후

평가항목	항목 해설	동의하지않음		보통임		동의함	
		1	2	3	4	5	
<b>Representational/Semantic layer</b>							
Match between formal and cognitive semantics	온톨로지 개념의 형식적 정의가 인지적으로 받아들여지는가						
Consistency	형식적 정의의 모순이 없는가?						
Clarity	문맥 및 배경적 지식이 명확히 드러나 있는가						
Explicitness	개념이 분명하게 표현되었는가						
Interpretability	개념의 정의와 설명이 잘 이해 되는가						
Accuracy	개념이 소셜미디어에서 사용되는 용어라고 할 수 있는가						
Comprehensiveness	다루는 도메인의 범위가 포괄적인가						
Granularity	온톨로지가 상세수준까지 커버하고 있는가						
Relevance	온톨로지가 개발 목적에 잘 부합하는가						
Description	온톨로지가 청소년 우울의 증상 및 증후 영역을 잘 기술하고 있는가						

④ 관리

평가항목	항목 해설	동의하지않음		보통임		동의함	
		1	2	3	4	5	
<b>Representational/Semantic layer</b>							
Match between formal and cognitive semantics	온톨로지 개념의 형식적 정의가 인지적으로 받아들여지는가						
Consistency	형식적 정의의 모순이 없는가?						
Clarity	문맥 및 배경적 지식이 명확히 드러나 있는가						
Explicitness	개념이 분명하게 표현되었는가						
Interpretability	개념의 정의와 설명이 잘 이해 되는가						

평가항목	항목 해설	동의하지않음		보통임		동의함	
		1	2	3	4	5	
Accuracy	개념이 소셜미디어에서 사용되는 용어라고 할 수 있는가						
Comprehensiveness	다루는 도메인의 범위가 포괄적인가						
Granularity	온톨로지가 상세수준까지 커버하고 있는가						
Relevance	온톨로지가 개발 목적에 잘 부합하는가						
Description	온톨로지가 청소년 우울의 관리 영역을 잘 기술하고 있는가						

⑤ 위험요인

평가항목	항목 해설	동의하지않음		보통임		동의함	
		1	2	3	4	5	
<b>Representational/Semantic layer</b>							
Match between formal and cognitive semantics	온톨로지 개념의 형식적 정의가 인지적으로 받아들여지는가						
Consistency	형식적 정의의 모순이 없는가?						
Clarity	문맥 및 배경적 지식이 명확히 드러나 있는가						
Explicitness	개념이 분명하게 표현되었는가						
Interpretability	개념의 정의와 설명이 잘 이해 되는가						
Accuracy	개념이 소셜미디어에서 사용되는 용어라고 할 수 있는가						
Comprehensiveness	다루는 도메인의 범위가 포괄적인가						
Granularity	온톨로지가 상세수준까지 커버하고 있는가						
Relevance	온톨로지가 개발 목적에 잘 부합하는가						
Description	온톨로지가 청소년 우울의 위험요인 영역을 잘 기술하고 있는가						

이밖에 온톨로지 전반에 대한 조언 부탁 드립니다. (자유형식)

Abstract

# Development and Evaluation of Youth Depression Ontology for Analyzing Social Data

Hyesil Jung

College of Nursing

The Graduate School

Seoul National University

In this study, youth depression ontology was developed and evaluated to be used for collecting and analyzing social media data.

Youth depression ontology was developed based on the 'ontology development 101' methodology; Step 1, the domain and scope of the ontology was determined by making a list of questions that the ontology should be able to answer. Step 2, existing ontologies representing youth depression were searched on the web (Google scholar, Pubmed) and in the literatures. Step 3, key terms were

extracted from four clinical practice guidelines (CPGs) published by NICE(2005, 2015), USPSTF(2009), beyondblud(2011) and Korean Clinical Research Center For Depression(2011). Step 4, the classes and the class hierarchy were defined using top-down process. First of all, the most general classes were categorized and more detailed classes were specialized appropriately. Step 5, the properties of classes and values of properties were defined. To extract the properties and values, the questionnaires such as 'Korea National Survey of Domestic Violence and Sexual Violence', 'Korea National Survey of School Violence', 'Study of Media Use among Children and Youth' were analyzed. Step 6, the ontology representing concepts of classes, properties, values, relationships between concepts was designed. Also terminology with synonyms for the class concepts was developed. Step 7, the ontology was represented by using Protégé.

As a result, 2,073 terms were extracted from four CPGs and 452 classes were defined out of these terms. Youth depression ontology consists of five sub-ontologies including Risk factors, Sign and symptoms, Measurement, Diagnostic result and Management care.

The ontology was evaluated in three ways.

First, the frequency of ontology concepts appeared in depression-related social media data was examined. As a result, 161,571 depression-related postings were searched from 215 social media channels during the last years 2012-2014 and out of 596 concepts in the ontology, 72.8% appeared in the postings of

social media.

Second, the content of coverage was examined by mapping concepts extracted from the counseling records of the Korean Youth Counseling and Welfare Institutes to the ontology concepts. In this process, key concepts from the counseling records were extracted by KoNLP package in R. In total, 1,574 narratives were extracted from 66 youth depression counseling records and 1,028 terms emerged from KoNLP. Out of these terms, 476 unique concepts were extracted by analyzing meaning of the terms. As a result of mapping, 73.1% were completely mapped.

Third, the structural and representational layers of the ontology were evaluated by six experts with experience in ontology design and psychiatric nursing. As a result, structural layer score was 4.8 of 5, with depth of hierarchy, balance and overall complexity criteria scoring 5 point. And representational layer score was 4.6 of 5.

The ontology was improved by adding new terms used in social media and counseling records. Also the ontology was revised by modifying the level of structure.

Even though youth depression ontology proved to be effective framework for collecting and analyzing social data, a further study aligning the ontology classes with top level ontology such as BFO and DOLCE is needed to improve its interoperability with other ontologies. Furthermore, a study mapping the ontology terms to standardized terminologies like SNOMED CT is also needed.

Keywords: youth depression, ontology, social media data, ontology  
development 101, natural language process, youth counseling record  
Student Number: 2013-20442