creative commons

COMMONS DEED

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

dCollection

간호학 박사 학위논문

건강정보 온톨로지 개발 및 건강정보 콘텐츠 연계 시스템의 개발 및 평가

2023년 2월

서울대학교 대학원

간호학과 간호학 전공

강 한 나

건강정보 온톨로지 개발 및 건강정보 콘텐츠 연계 시스템의 개발 및 평가

지도 교수 박 현 애

이 논문을 간호학박사 학위논문으로 제출함
2022년 10월

서울대학교 대학원
간호학과 간호학 전공
강 한 나

강한나의 간호학박사 학위논문을 인준함
2022년 12월

위 원 장 _

부위원장 _

위 원 _

위 원 _

위 원 _

국문초록

본 연구의 목적은 국가건강정보포털의 검색 향상을 위해 건강정보 온톨로지를 건강정보 콘텐츠에 연계하는 시스템을 개발하여 평가하고, 그 시스템을 활용해 건강정보 온톨로지를 개발한 뒤 건강정보 콘텐츠에 건강정보 온톨로지 개념을 연계하며, 건강정보 온톨로지 및 건강정보 콘텐츠에 연계된 건강정보 온톨로지의 개념 및 용어의 연계 적합성을 평가하는 데 있다.

본 연구는 크게 세 단계로 나누어 진행하였다. 첫째 단계에서는 건강정보 콘텐츠를 활용하여 건강정보 온톨로지를 개발하고 건강정보 콘텐츠에 연계할 수 있는 시스템을 개발하였고, 둘째 단계에서는 해당 시스템을 활용하여 건강정보 온톨로지를 개발하였으며, 마지막으로 건강정보 콘텐츠에 온톨로지 개념 및 용어를 연계하였다.

건강정보 온톨로지 개발 및 건강정보 콘텐츠 연계 시스템은 5명의 전문가로부터 휴리스틱 평가와 PSSUQ 평가를 받았다. 휴리스틱 평가 결과 사용성 문제가 있어 반드시 수정되어야 하는 심각성 수준 3점 이상의 항목은 2개였다. PSSUQ 평가 결과 항목별 평균 점수는 3.4~5.0점이었고 전체 평균 점수는 4.3점으로 나타났다. 건강정보 온톨로지는 개념 3,823개와 동의어 12,082개로 개발되었다. 경강정보 온톨로지는 구조에 오류가 없다고 평가되었고 내용 평가에서는 평균 4.4점을 받았다. 건강정보 콘텐츠에 건강정보 온톨로지 개념을 연계한 결과 목차 단위 콘텐츠 3,241개에 온톨로지 개념 14,740개가 연계되었다. 연계 적합성에 대한 전문가 평가에서는 50개의 검색 키워드로 검색한 결과 중 3개의 키워드에서 각각 하나의 콘텐츠 연계가 부적합하다는 평가를 받았다.

본 연구에서 개발한 건강정보 온톨로지 개발 및 건강정보 콘텐츠 연계 시스템의 의의는 기존의 온톨로지 개발 및 관리 프로그램이 온톨로지에 국한되고 온톨로지가 활용될 수 있는 정보와 직접 연계하지 못한다는

한계를 보완하여 건강정보 콘텐츠와 자동으로 연계할 수 있다는 점이다. 해당 시스템을 통해 개발된 건강정보 온톨로지는 구조와 내용 두 측면에서 모두 긍정적인 평가를 받았다. 건강정보 콘텐츠 연계의 적합성은 건강정보 온톨로지와 건강정보 콘텐츠 연계 원칙에 따라 이루어졌으며 전문가들로부터 평가받은 결과에 따라 일부 결과를 개선할 수 있었다.

앞으로는 온톨로지 개발 및 건강정보 콘텐츠 연계 시스템을 콘텐츠 개발 과정에 적용할 수 있도록 CMS와 통합하는 작업이 필요하다. 또한 온톨로지 기반 검색 기능을 강화할 수 있도록 사용자 용어 수집을 자동화할 필요가 있다. 아울러 온톨로지 기반 검색 제공에 더불어 사용자의 검색 행태 기록을 활용하여 보다 더 개인화된 건강정보 콘텐츠를 제공할 필요가 있다.

주요어 : 온톨로지, 온톨로지 기반 검색, 건강정보 콘텐츠, 정보시스템 개발, 소비자 건강정보 제공
학 번 : 2018-34470

목 차

제 1 장 서론 ····· 1

제 1 절 연구의 필요성 ····· 1

제 2 절 연구 목적 ····· 4

제 3 절 용어의 정의 ····· 5

1. 건강정보 콘텐츠 ····· 5

2. 건강정보 온톨로지 ····· 5

3. 건강정보 콘텐츠에 건강정보 온톨로지 개념 연계 ····· 5

4. 건강정보 온톨로지 개발 및 건강정보 콘텐츠 연계 시스템 ····· 5

제 2 장 문헌 고찰 ····· 7

제 1 절 온라인 건강정보 제공 ····· 7

제 2 절 정보 시스템 개발 및 평가 ····· 9

1. 정보 시스템 개발 ····· 9

2. 정보 시스템 평가 ····· 9

제 3 절 온톨로지의 개발, 평가 및 활용 ····· 11

1. 온톨로지의 정의 ····· 11

2. 건강정보 온톨로지 ····· 11

3. 온톨로지 평가 ····· 11

4. 온톨로지 기반 의미론적 정보 검색 ····· 12

제 4 절 온톨로지 개발 및 관리 도구 ····· 13

제 3 장 연구 방법 ····· 15

제 1 절 건강정보 온톨로지 개발 및 건강정보 콘텐츠 연계 시스템 개발 ····· 15
1. 시스템 분석 ····· 16
2. 시스템 설계 ····· 17
3. 시스템 구현 및 적용 ····· 17
4. 시스템 사용자 평가 ····· 18
제 2 절 건강정보 온톨로지 개발 ····· 19
1. 건강정보 온톨로지 개발 ····· 19
2. 건강정보 온톨로지 평가 ····· 22
제 3 절 건강정보 콘텐츠에 건강정보 온톨로지 개념 연계 ····· 23
1. 텍스트 건강정보 콘텐츠에 온톨로지 개념 연계 ····· 23
2. 미디어 콘텐츠에 온톨로지 개념 연계 ····· 24
3. 건강정보 콘텐츠 온톨로지 개념 연계 대상 콘텐츠 ····· 25
4. 건강정보 콘텐츠와 건강정보 온톨로지 개념 연계 원칙 및 적합성 평가 ····· 26
제 4 절 연구의 윤리적 고려 ····· 27
제 4 장 연구 결과 ····· 28
제 1 절 건강정보 온톨로지 개발 및 건강정보 콘텐츠 연계 시스템 28
1. 시스템 분석 결과 ····· 28
2. 시스템 설계 결과 ····· 33
3. 시스템 구현 및 적용 결과 ····· 47
4. 사용자 평가 결과 ····· 47
제 2 절 건강정보 온톨로지 개발 및 평가 결과 ····· 50

1. 건강정보 온톨로지 개발 결과 ····· 50

2. 건강정보 온톨로지 평가 결과 ····· 53

제 3 절 건강정보 콘텐츠에 건강정보 온톨로지 개념 연계 및 적합성 평가 결과 ····· 57

1. 건강정보 콘텐츠에 건강정보 온톨로지 개념 연계 결과 ····· 57

2. 건강정보 콘텐츠에 건강정보 온톨로지 개념 연계 적합성 평가 결과 ····· 57

제 5 장 논의 ····· 61

제 1 절 건강정보 온톨로지 개발 및 건강정보 콘텐츠 연계 시스템 61

제 2 절 건강정보 온톨로지 ····· 62

제 3 절 건강정보 콘텐츠에 건강정보 온톨로지 개념 연계 ····· 63

제 6 장 결론 및 제언 ····· 65

참 고 문 헌 ····· 67

List of Tables

Table 1. Number of Health Information Content by Superclass ····· 25

Table 2. Annotation principles by content type ····· 26

Table 3. Functions for Ontology Development and content Annotation System ····· 28

Table 4. Example Rules for Categorizing Health Information Content 30

Table 5. Example Rules for Auto-Generated Keywords ····· 31

Table 6. System Functions by User Level ····· 42

Table 7. Result of Heuristic Evaluation ····· 48

Table 8. Result of PSSUQ Evaluation ····· 49

Table 9. Number of Ontology Concepts and Synonyms by Superclass 51

Table 10. Examples of Domain-Range Superclasses by Relation Type ····· 52

Table 11. Evaluation Result of Health Information Ontology ····· 55

Table 12. Result of Health Information Content and Health Information Ontology Concept Annotation ····· 57

Table 13. Examples of Health Information Content and Health Information Ontology Concept Annotation ····· 58

List of Figures

Figure 1. Process of 'Health Information Ontology Development and Content Annotation System' Development ····· 16

Figure 2. User Interface Design of Term Extraction Web Page ····· 17

Figure 3. Attribute Relationships of Health Information Ontology Super Classes ····· 22

Figure 4. An Example of Health Information Content and Ontology Concept Annotation ····· 24

Figure 5. An Example of Image Content for OCR ····· 25

Figure 6. Use-case Diagram of Ontology Development and Content Annotation System ····· 30

Figure 7. An Example of 'Content Import' ····· 35

Figure 8. An Example of Term Extraction Using OCR ····· 36

Figure 9. An Example of NLP ····· 37

Figure 10. An Example of Ontology Visualization ····· 38

Figure 11. An Example of Ontology Management ····· 39

Figure 12. An Example of Health Information Content with Ontology Concept Annotation ····· 41

Figure 13. An Example of Ontology Evaluation ····· 43

Figure 15. Entity Relation Diagram of Ontology Development and Content Annotation System ····· 46

Figure 16. Super classes of Health Information Ontology ····· 50

Figure 17. Examples of the Second-Level Classes of Health Information Ontology ····· 52

제 1 장 서론

제 1 절 연구의 필요성

인터넷 이용이 확산되고 인터넷을 이용한 건강정보탐색이 활발해지면서, 보건복지부를 비롯한 정부 기관은 인터넷을 통해 양질의 건강정보를 제공할 목적으로 40여 개의 보건복지 공공사이트를 운영하여 건강정보, 질병 정보, 약물 정보, 통계 정보, 의료기관 및 약국 정보 등을 제공하고 있다(보건복지부, 2016). 그중에 국가건강정보포털은 질병관리청이 인터넷을 통해 양질의 건강정보를 제공하고자 운영하는 사이트이다(보건복지부, 2016). 국가건강정보포털은 700여 종의 콘텐츠를 보유하고 있으며, 포털에 접속하는 페이지 조회 수는 매년 증가하여 2021년 약 2,400만 건에 달한다(질병관리청, 2021).

2017년 소비자 패널의 네이버, 헬스조선, 건강IN, 국가건강정보포털에 대한 정보 품질 및 검색 우수성 비교 평가에 따르면, 국가건강정보포털은 정보 신뢰성은 가장 높게 나타났지만 검색 성능 및 사용 방법은 상대적으로 낮게 나타났다. 이에 대한 주관식 상세 응답을 살펴보면, 콘텐츠가 잘 검색이 되지 않거나 소비자가 원하는 구체적인 수준의 정보를 찾아가기까지가 쉽지 않았다는 답변이 나왔다(보건복지부, 2018). 이는 콘텐츠가 의미 단위별로 적절한 검색 키워드가 연계되어 있지 않고, 다양한 사용자 용어로 검색되지 않기 때문으로 분석된다.

우선, 국가 건강정보 포털 콘텐츠는 의미 단위별로 다양한 검색 키워드가 연계되어 있지 않아, 콘텐츠의 제목과 정확하게 일치하는 키워드를 사용하지 않으면 관련 정보가 검색되지 않는다. 또한 이미지와 동영상 콘텐츠의 경우에도 그 내용에 등장하는 주요 개념이 검색되지 않고 오로지 제목으로만 검색이 가능하다. 예를 들어, ‘당뇨의 3대 증상’이라는 이미지는 이미지 내에 다뇨, 다식, 다갈 등의 주요 개념이 등장하지만, 이들 검색어로는 찾을 수 없고 ‘당뇨의 3대 증상’이라는 이미지 제목을 입

력해야만 찾을 수 있다. 또한 소비자들이 일반적으로 사용하는 약어 등으로는 검색되지 않는다. 일례로 '임신성 당뇨'의 줄임말로 흔히 쓰이는 '임당'과 같은 사용자 용어로 검색하면 해당 콘텐츠를 조회할 수 없다.

이러한 검색 문제를 해결하기 위해서는 국가 건강정보 포털의 콘텐츠를 활용해 건강정보의 개념과, 개념들 간의 관계로 구성된 온톨로지를 개발하고 온톨로지 개념에 대한 동의어인 사용자 용어 체계를 개발하여, 검색어로서 건강정보 콘텐츠에 연계하는 작업이 필요하다. 온톨로지는 특정 영역의 개념과 그들 간의 관계를 정의한 것이다(Noy, 2001). 온톨로지는 다양한 분야의 정보 검색의 활용에 적용할 수 있는데, 의미론적 검색을 가능하게 하고(Dung & Kameyama, 2007) 다른 검색 엔진에 비해 높은 검색 정확도와 정밀도를 나타낼 수 있다(Yu, 2019).

미국 국립 의학 도서관(NLM)의 MedlinePlus (https://medlineplus.gov/)는 건강 주제(health topics) 온톨로지를 개발하고 이를 건강정보 콘텐츠에 연계하여 온톨로지 기반 건강정보 콘텐츠 검색 결과를 제공하고 있다. MedlinePlus는 소비자가 신뢰할 수 있는 전문 건강정보에 쉽게 접근할 수 있게끔 건강정보를 온톨로지 기반으로 구조화하도록 설계되었다. 또한 의료 용어의 동의어로 소비자들이 사용하는 용어를 추가하여, 소비자가 건강 관련 질문을 할 때 잘 답변할 수 있도록 하고 있다. 포르투갈에서도 웹상에 분산된 건강 관련 정보를 구조화하고 활용성을 높이기 위해 온톨로지를 개발하고 건강정보 콘텐츠에 개념을 연계하여, 신경 질환 영역의 건강정보 검색에 의미론적 검색을 적용되었다(Mendonca, 2012).

우리나라에서도 국가건강정보포털의 건강정보 콘텐츠를 용이하게 검색할 수 있도록 하려면 사용자 용어를 포함한 건강정보 영역의 온톨로지를 개발해야 하고, 건강정보 콘텐츠에 건강정보 온톨로지 개념을 연계하여 온톨로지를 활용한 의미 기반 검색을 제공할 필요가 있다. 따라서 온톨로지를 개발·관리하고, 건강정보 콘텐츠에 온톨로지 개념을 연계할 수 있는 시스템의 개발이 필요하다. 온톨로지 개발, 관리 및 온톨로지 개념과 건강정보 콘텐츠의 연계, 사용자 용어 수집이 수작업으로 진행되면 많은 시간과 노력을 요하므로, 이를 극복하기 위해서는 온톨로지 개발

및 콘텐츠와의 연계 작업을 반자동화할 수 있는 시스템을 개발할 필요가 있다.

따라서 본 연구에서는, 건강정보 콘텐츠 검색 기능 향상을 위한 건강정보 온톨로지 개발 및 건강정보 콘텐츠 연계 방법으로서, 건강정보 온톨로지 개발 및 국가 건강정보 콘텐츠 연계 시스템을 개발하고, 개발한 시스템을 활용해 건강정보 온톨로지를 개발하고, 건강정보 콘텐츠에 건강정보 온톨로지 개념을 연계하며, 시스템 및 온톨로지, 연계 적합성을 평가하고자 한다.

제 2 절 연구 목적

본 연구는 국가건강정보포털의 검색 향상을 위해 건강정보 온톨로지를 건강정보 콘텐츠에 연계하는 시스템을 개발하여 평가하고, 시스템을 활용해 건강정보 온톨로지를 개발한 뒤 건강정보 콘텐츠에 건강정보 온톨로지 개념을 연계하며, 건강정보 온톨로지 및 건강정보 콘텐츠에 연계된 건강정보 온톨로지의 개념 및 용어의 연계 적합성을 평가하는 것을 목적으로 한다.

본 연구의 목적을 세분하면 다음과 같다.

1) 건강정보 온톨로지 개발 및 국가 건강정보 콘텐츠 연계 시스템을 개발한다.
2) 건강정보 온톨로지 개발 및 국가 건강정보 콘텐츠 연계 시스템의 사용성을 평가한다.
3) 개발한 시스템을 활용해 건강정보 온톨로지를 개발한다.
4) 건강정보 온톨로지의 구조와 내용을 평가한다.
5) 개발한 시스템을 활용해 건강정보 콘텐츠에 건강정보 온톨로지 개념 및 용어를 연계한다.
6) 건강정보 콘텐츠에 연계된 건강정보 온톨로지 개념 및 용어의 적합성을 평가한다.

제 3 절 용어의 정의

1. 건강정보 콘텐츠

건강정보 콘텐츠란 문자, 이미지, 동영상을 이용해서 제공되는 건강정보를 말한다. 본 연구에서 건강정보 콘텐츠는 질병관리청이 운영하는 국가건강정보포털 (https://health.kdca.go.kr/)의 '건강정보' 카테고리에서 제공하는 문자, 이미지, 동영상 콘텐츠를 말한다.

2. 건강정보 온톨로지

온톨로지란 특정 영역에서 사용되고 있는 개념과, 개념 간 관계의 형식적이고 명시적인 명세(Gruber, 1993)로서 컴퓨터가 이해하고 처리할 수 있는 형태로 표현한 개념들의 논리적 집합이다(박정은 외, 2013).

본 연구에서 건강정보 온톨로지는 국가건강정보포털에서 추출한 질병, 증상 및 징후, 평가 및 검사, 치료, 자가관리, 생활습관 관리, 위험요인, 지원체계 영역의 개념과 개념들 간의 관계 및 각 개념과 동의어의 관계가 되는 용어체계를 포함한다.

3. 건강정보 콘텐츠에 건강정보 온톨로지 개념 연계

온톨로지는 더 나은 검색을 위해 데이터의 의미론적 인덱스를 설계하는데 도움이 된다. 온톨로지 기반 인덱싱은 데이터 요소를 온톨로지 개념에 결합하는 것을 말한다(Tchechmedjiev, 2018).

본 연구에서 건강정보 콘텐츠에 건강정보 온톨로지 개념 연계는 건강정보 콘텐츠에 건강정보 온톨로지에 속한 개념을 결합하는 것을 의미한다.

4. 건강정보 온톨로지 개발 및 건강정보 콘텐츠 연계 시스템

건강정보 온톨로지 개발 및 건강정보 콘텐츠 연계 시스템은 건강정보 온톨로지의 개발과 관리 및 건강정보 콘텐츠에 건강정보 온톨로지 개념 및 용어의 연계를 목적으로, 건강정보 콘텐츠에서 용어를 추출하여 온톨로지 개념 및 동의어를 생성, 수정, 삭제, 조회하고 건강정보 콘텐츠에

개발된 온톨로지에 속한 개념을 연계하는 웹베이스 시스템을 말한다.

제 2 장 문헌 고찰

본 연구는 국가건강정보포털의 검색 향상을 위해 건강정보 온톨로지를 개발하여 건강정보 온톨로지 개발 및 건강정보 콘텐츠 연계 시스템을 개발하고 평가하는 연구이다. 본 장에서는 온라인 건강정보 제공에 대해 살펴본 후, 정보 시스템 개발 및 평가, 온톨로지의 개발, 활용과 평가 및 온톨로지 개발 및 관리 도구에 대해 고찰한 결과를 기술하였다.

제 1 절 온라인 건강정보 제공

인터넷 사용이 증가하면서 인터넷이 건강정보 정보획득의 기회를 높이고 건강위험 및 미충족 정보 욕구를 채워주는 중요한 수단이 되고 있다(오영삼, & 조영은, 2019). 온라인 건강정보는 시간과 비용 측면에서 효율성이 높으며 확장성을 제공할 수 있으나, 한계도 존재한다. 오영삼과 조영은은 그 한계를 검색의 어려움, 건강정보 습득 정도, 잘못된 건강정보, 검색력, 독해력, 응용력 요인에 따라 온라인으로 분류하였다. 이에 따른 온라인 건강정보 활용 과정의 문제를 해결하기 위해 전문가들이 제시한 방안 중 하나로는 온라인 건강정보의 관리와 감독, 유통의 주체로서의 정부의 역할이 강조되었다(오영삼, & 조영은, 2019).

이에 따라 신뢰성 있는 온라인 건강정보 제공을 위해 전 세계의 다양한 국가 공공기관에서 건강정보포털을 운영하고 있다. 건강정보포털로 가장 널리 알려진 미국 국립 의학 도서관(NLM)의 MedlinePlus (https://medlineplus.gov/)는 미국 국립의학도서관(NLM: National Library of Medicine)과 미국 국립보건원(NIH), 정부 산하기관과 보건의료관련 기구들로부터 신뢰할 수 있는 정보를 제공받고 있다. 해당 포털에서는 증상, 원인, 치료와 예방 등 1,000여 개의 질병과 건강상태 그리고 웰니스 관련 콘텐츠를 제공하고 있는데, 이를 사용자가 쉽게 접근 및 이용할 수 있도록 건강정보를 온톨로지를 기반으로 구성하였다.

MedlinePlus에 포함된 건강정보를 분류하는 데 활용된 건강주제(health topics) 온톨로지는 2,146개 클래스와 13가지 속성으로 구성되어 있으며 온톨로지 내 개념은 건강정보와 연계되어 있으며 개념에 대한 다양한 사용자 용어를 매핑하고 있다. 예를 들어 당뇨의 경우 Diabetes, Diabetes mellitus, DM 등으로 매핑하여 검색어로 활용하고 있다. 해당 포털에서 제공되는 건강정보는 온톨로지 구조에 따라 신체 계통, 질병 상태, 진단 및 치료, 인구집단 특성 등으로 분류되어 있어 다양한 방법으로 정보 검색이 가능하다. 또한 MedlinePlus에서 제공되는 모든 정보자원은 건강주제, 출처 URL, 언어, 생성일자, 일련번호 등 메타데이터의 기본 요소로 색인화되어 관리되고 있다.

영국의 경우 국가건강서비스(NHS: National Health Service) 체계에서 치료보다는 예방을 강조하고 있어 만성 퇴행성 질환에 대해 관리가 잘 이뤄질 수 있는 의료 시스템을 갖고 있으며 품질과 신뢰성이 보장된 의료정보의 수집, 관리, 제공 및 정보의 일관성 유지를 목적으로 보건복지정보센터 (Health and Social Clinical Information Center; HSCIC)에서 공공 온라인 서비스(NHS, 구 NHS choice)를 운영하고 있다. NHS에서는 건강상태 및 질병 관련 정보를 알파벳(index), 성별에 따른 신체 부위, 자주 검색되는 건강 상태(common conditions) 등의 주제로 분류하여 건강상태 혹은 질병 관련 정보가 다양한 주제 분류에 포함될 수 있도록 설계되어 있어서 사용자가 다양한 경로로 원하는 건강정보를 쉽게 검색하고 조회할 수 있다.

이처럼 온라인의 수많은 건강정보 중 신뢰성 있는 건강정보가 소비자에게 잘 전달되기 위해서는, 건강정보 콘텐츠의 생산뿐 아니라 적절한 검색 기능의 제공으로 소비자가 원하는 맞춤형 콘텐츠를 제공하려는 방안이 필요하다.

제 2 절 정보 시스템 개발 및 평가

1. 정보 시스템 개발

정보 시스템이란 컴퓨터 등 정보기기들을 이용하여 효율적인 업무처리를 돕기위한 체제로, 컴퓨터에 의해 처리가 가능한 형태로 자료를 입력(input)하고, 컴퓨터에 의해 그 자료를 저장, 처리, 가공(process)하여 컴퓨터를 통해 정보를 필요한 시점에 출력(export)할 수 있도록 설계되고 구현된 정보체계를 의미한다(허원실, 2008).

이러한 정보시스템의 개발 모형은 시스템 개발 생명주기 모형, 프로토타이핑 모형, 나선형 모형 등이 있다. 그중 전통적인 시스템 개발 모형인 시스템 개발 생명주기 모형(Software Development Life Cycle; SDLC)은 분석 단계(정의 단계), 설계 단계, 구현 단계, 테스트 단계, 유지보수 단계의 5단계로 진행된다. 먼저 분석 단계는 시스템의 요구 분석 단계로, 시스템이 갖추어야 할 기능을 구체적이고 명확하게 정의하는 과정이다. 설계 단계는 기능 구현 방법을 기술적으로 어떻게 구현할 것인지 인터페이스 및 데이터베이스 등을 설계하는 과정이다. 구현 단계는 설계를 바탕으로 직접 프로그램을 작성하는 과정이며, 테스트 단계는 시스템 구동을 시험적으로 운영해 보며 테스트하는 과정이다. 유지보수 단계는 시스템을 운영하며 시스템 일부 기능을 개선하는 과정이다.

2. 정보 시스템 평가

개발한 정보 시스템의 품질은 정확성, 유지보수성, 무결성, 사용성으로 평가하는데 그중 시스템 사용성은 사용자가 쓰기에 편리하다는 것을 의미한다. 사용성은 시스템을 배우는데 요구되는 물질적 및 지적 노력, 시스템의 사용이 적합한 효율을 갖는 데 걸리는 시간, 보통 사람이 시스템을 사용할 때 측정되는 생산성의 순수 증가, 사용자의 시스템에 대한 주관적 평가로 구성된다. 시스템 사용성을 평가하는 도구는 Minnesota대학에서 개발 및 관리하는 Questionnaire on information structure (QUIS) (Skopeteas et al, 2006), Davis(1989)가 개발한 지각된 유용성 및 지각된

사용 용이성 측정 도구, Lewis(1992)가 개발하고 개정한 Post-Study System Usability Questionnaire (PSSUQ), 소수의 사용성 평가 전문가를 대상으로 사용성 평가 원칙에 따라 시스템의 사용자 화면을 평가하는 Neilson (1994)의 휴리스틱 평가(Heuristic Evaluation) 등이 있다. 본 연구에서는 시스템 사용 대상자 및 도구의 문항 수 등을 고려하여 휴리스틱 평가와 PSSUQ를 사용한 사용성 평가를 진행할 예정이다.

1) 휴리스틱 평가

휴리스틱 평가는 일반적으로 사용성이 높은 시스템이 공통적으로 갖는 특징으로부터 추출한 10개의 원칙을 통해 평가하는 것이다. Neilson(1994)이 제시한 10개의 원칙은 시스템 상태의 사각화, 시스템과 실 시계의 일치, 제어의 자유, 일치와 표준, 오류 방지, 기억보다 인식, 융통성과 효율성, 간소화된 디자인, 에러 인식, 도움말과 문서화이다. 휴리스틱 평가는 빠른 시간 내에 쉽게 사용성의 문제를 발견할 수 있다는 장점이 있다.

2) PSSUQ (Post-Study System Usability Questionnaire)

PSSUQ는 사용자의 인지적 측면을 반영한다는 점에서 휴리스틱 평가의 제한점을 보완할 수 있다. 표준화된 사용성 질문지를 이용하는 평가는 컴퓨터 시스템을 사용하는 사용자가 인지한 만족도를 평가하기 위해 개발되었고, 1992에 Lewis에 의해 19문항으로 이루어진 첫 번째 버전이 개발된 이래 2002년 16문항으로 이루어진 세 번째 버전(Lewis, 2002) 까지 개정되었다.

정보시스템은 이와 같이 개발 단계를 잘 따라 설계 및 개발되어야 하고, 정보시스템의 사용 목적에 따라 적절한 평가 도구를 사용해 평가를 진행해 수정 및 유지 보수해 나가야 할 필요가 있다.

제 3 절 온톨로지의 개발, 평가 및 활용

1. 온톨로지의 정의

온톨로지는 어떤 영역에서 사용되고 있는 용어와, 용어 간 관계의 형식적이고 명시적인 명세(Gruber, 1993)이다. 온톨로지는 사람뿐 아니라 컴퓨터가 이해 가능한 정의를 제공할 수 있으며 온톨로지 개발의 목적은 해당 영역의 지식을 분석하고, 해당 영역의 가정을 명시적으로 만들고, 해당 영역의 지식과 운영 지식을 분리하고, 해당 영역의 지식을 재사용 가능하도록 하고, 사람 또는 시스템 간 정보 구조에 대한 공통된 이해를 공유할 수 있도록 하는 데 있다(Noy, 2001).

2. 온톨로지 개발

온톨로지 개발 방법 중 수작업 개발 방식은 많은 시간과 인력이 소요되며 다른 주제로의 확장이 어려운 단점이 있다. 이런 문제를 극복하기 위해 온톨로지를 자동으로 구축하는 시도가 이루어지고 있으며, 온톨로지를 반자동(Semi-automatic process)으로 구축하는 방안으로 텍스트의 분석결과로 얻어지는 단어들의 분포를 이용하여 온톨로지를 구축하고 검토하면서 확장하는 방법이 활용(Kim et al., 2019; 임수연 외, 2005; Ko et al., 2019; Mizuno et al., 2016)되고 있다. 자연어 처리 기반 온톨로지 개발에서 더 나아가 이 과정을 시스템으로 개발한 온톨로지 자동 구축 연구들(심준 & 이홍철, 2009; Ruiz-Martínez et al., 2011; Cahyani, & Wasito, 2017)도 있다.

3. 온톨로지 평가

온톨로지 평가 방법론은 기존 온톨로지와의 비교, 활용될 애플리케이션에 적용, 원천데이터와의 적합성과 관련성 및 미리 정의된 지표나 표준을 통한 전문가 평가로 분류할 수 있다(김성훈, & 오삼균, 2015). 그중 지표를 사용하는 평가 방법으로 널리 사용되는 Gomez-Perez의 Criteria to Evaluate Ontologies에서는 대분류를 온톨로지의 구조, 온톨로지의 내

용으로 나누어 평가한다. 온톨로지의 구조는 온톨로지 구조의 비일관성과 불완전성을 평가하는 항목으로 구성되어 있으며, 온톨로지의 내용은 온톨로지의 일관성, 완성도, 간결성, 확대 가능성, 민감성을 평가하는 항목으로 구성되어 있다(Gomez-Perez, 2004).

4. 온톨로지 기반 의미론적 정보 검색

온톨로지는 매우 다양하게 활용되나, 웹에 있는 방대한 양의 데이터를 구조화된 형식으로 조직화하여 정보를 제공할 때 온톨로지를 활용할 수 있다. 온톨로지 기반의 정보 검색 시스템은 사용자의 검색어로 인덱싱된 데이터 간 의미론적 유사성을 기반으로 데이터를 검색할 수 있도록 한다(Asim, 2019).

그중 웹 페이지의 정보를 온톨로지로 연결하여 의미론적 검색이 가능하도록 하는 방법에 대한 연구에서 Dung과 Kameyama (Dung, & Kameyama, 2007)는 기존 품사 기반 정보 추출 시스템의 한계점을 극복할 수 있는 온톨로지 기반의 의학 정보 추출 시스템을 제안하였다. 이 시스템에서는 '의미론적 요소 추출 알고리즘'을 이용하여 웹기반 건강정보에서 개념(concept)과 개념의 서술어(description), 개념과-서술어 쌍, 질병명을 추출하여 추출된 의미론적 요소를 '새로운 의미론적 요소 학습 알고리즘'을 사용하여 온톨로지를 개선하는 데 사용하였다. 또한 웹기반 건강정보에서 의미론적 요소를 추출한 후에 추출된 단어와 웹기반 건강정보를 지식베이스로 저장하여 나중에 건강정보를 검색하는 데 활용할 수 있도록 하였다. 이 시스템을 이용한 결과 의미론적 추출의 정확도가 증가하였고, 온톨로지 업그레이드의 효율성이 증가하였다. 이 연구를 통해 품사 기반의 정보 추출 방식보다 건강정보 온톨로지를 활용하는 접근이 검색 기능을 개선에 도움이 된다는 것을 확인할 수 있었다.

Yu (2019)는 도메인 온톨로지 기반의 검색 모델을 개발하여 IEEE Electronic Library에 포함된 1,000개의 학술자료를 사용하여 평가한 결과 온톨로지 기반의 검색 모델이 기존의 검색 모델보다 사용자의 요구에 정밀도와 재현율(recall rate)이 더 높은 검색결과를 제공하는 것으로 나

타났다. 온톨로지 기반의 검색 모델에서는 개념 간의 관계를 정의하여 개념 간 의미론적 네트워크(semantic network)를 구성하고 이를 통해 개념간 연관 검색, 유사어 확장 검색(synonym expansion retrieval)이 가능하도록 하였다.

온톨로지를 기반으로 한 의미론적 검색 엔진을 개발한 또 다른 연구에서는(Sayed, 2017) 개발한 온톨로지 기반 검색 엔진을 구글 및 다른 검색 엔진과 비교한 결과 검색의 정확도와 정밀도가 더 높게 나타났으며, 특히 단순 쿼리보다 복합 쿼리에서 더 나은 결과를 보였다.

이처럼 온톨로지를 개발하여 정보 검색에 활용하는 접근은 의미론적 정보 제공에 효과적이나, 온톨로지 개발 및 데이터 요소에 온톨로지 개념을 연계하는 것은 수작업으로 이루어지는 경우가 많아 많은 노력과 시간이 소요된다는 한계가 있다.

제 4 절 온톨로지 개발 및 관리 도구

온톨로지는 다양한 개념, 용어 및 관계로 이루어져 있기 때문에 단순한 문서 형태로 관리가 어려워, 지속적인 개발 및 관리를 위한 적절한 도구가 필요하다. 온톨로지 개발 및 관리 툴에 요구되는 주요 기능은 개념· 동의어· 관계의 생성· 수정· 삭제, 개념들의 관계 시각화, 협업을 위한 사용자 관리· 추적, 콘텐츠의 입출력 등이다. 현재까지 다양한 기관에서 온톨로지 개발 및 편집, 관리를 위한 도구들을 개발하여 사용하고 있다.

가장 널리 사용되는 온톨로지 개발 및 관리 툴은 스탠포드 대학교(Stanford University)에서 개발한 프로테제(Protégé)이다(Stanford University, 2022). 프로테제는 데스크탑용과 웹용으로 제공되어 36만명 이상의 사용자들이 무료로 사용하고 있다. 단순한 인터페이스와 협업에 최적화된 관리 기능 및 특정 포맷을 이용한 콘텐츠의 대량 업로드 및 다운로드 기능을 제공하며, 반응형으로 제공되는 온톨로지 시각화 기능도 제공한다.

다른 온톨로지 개발 툴로는 키메라(Chimæra)가 있다. 키메라의 특징은

여러 온톨로지를 통합하고, 각각의 온톨로지를 평가할 수 있다는 점이다(McGuinnes, 2000). 또한 온톨로지를 조회하고 수정하는 등의 기본적인 기능 및 온톨로지를 시각화하여 보여주는 기능도 제공한다.

맨체스터 대학(University of Manchester) 에서 개발한 OilED는 온톨로지 편집 도구로, 사용자 친화적 인터페이스를 제공하는 동시에 웹 온톨로지 언어 OIL을 지원한다(Bechhofer, 2001). 또한 이전의 온톨로지 관리 도구들의 기능에 더하여 개념의 정의 수준(primitive/ defined)를 구분하는 기능 등 다른 기능들을 제공한다. 다만 그래픽적 시각화는 지원하지 않는다.

마지막으로 웹온토(WebOnto)는 웹에서 온톨로지를 공유여 협업하고자 개발된 도구이다(Domingue, 1998). 온톨로지 개발 기능뿐 아니라, 온톨로지의 개념 및 관계에 대한 의견을 교환할 수 있도록 코멘트 기능을 지원한다.

위와 같은 온톨로지 개발 및 관리 도구는 공통적으로 온톨로지의 개념 및 관계를 생성하고 관리하며 시각화 하는 기능을 지니고 있으나, 온톨로지 개념을 데이터 요소에 인덱싱하는 기능은 제공하지 않는다.

제 3 장 연구 방법

본 연구는 크게 세 단계로 나누어 진행하였다. 첫째 단계에서는 건강정보 콘텐츠를 활용하여 건강정보 온톨로지를 개발하고 건강정보 콘텐츠에 연계할 수 있는 시스템을 개발하고 둘째 단계에서는 해당 시스템을 활용하여 건강정보 온톨로지를 개발하고 마지막으로 건강정보 콘텐츠에 온톨로지 개념 및 용어를 연계하였다.

제 1 절 건강정보 온톨로지 개발 및 건강정보 콘텐츠 연계 시스템 개발

건강정보 온톨로지 개발 및 건강정보 콘텐츠 연계 시스템은 Figure 1에 제시된 바와 같이, 프로세스, 데이터, 인터페이스 영역에 대해 분석, 설계, 구현, 평가, 유지보수 단계의 5단계에 따라 개발하였다.

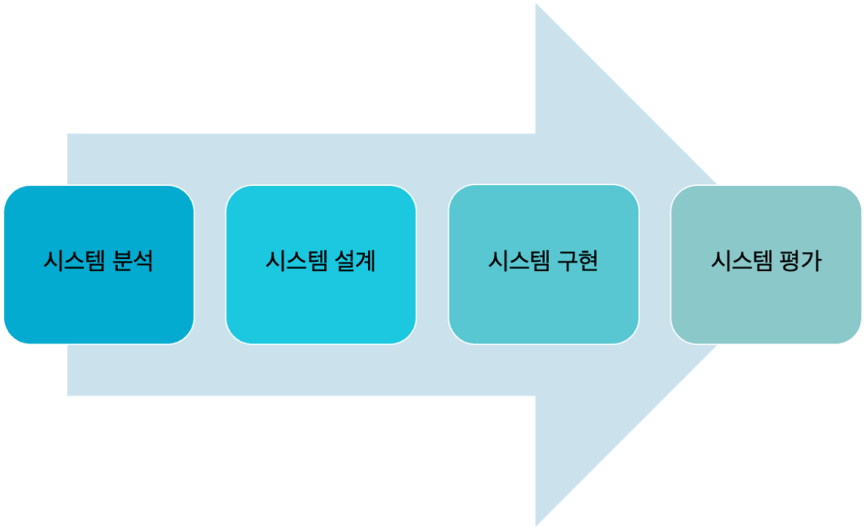


Figure 1. Process of 'Health Information Ontology Development and Content Annotation System' Development

1. 시스템 분석

시스템 분석 단계는 건강정보 온톨로지 개발 및 건강정보 콘텐츠 연계 시스템에 필요한 기능 및 데이터 항목을 분석하는 단계이다.

시스템을 사용할 연구원들을 대상으로 하여, 건강정보 온톨로지를 개발하고 건강정보 온톨로지의 개념을 건강정보 콘텐츠에 연계하는 작업을 위한 필수 기능을 추출하고 이를 사용자와 시스템, 사용사례를 표현하는 유스케이스 다이어그램으로 표현하였다.

필수 기능을 구현하기 위해 어떤 화면을 설계할지 다른 온톨로지 관리 시스템 및 사이트를 참고하여 기능 구현의 상세 화면을 분석하였다. 예를 들어, 온톨로지 개념 생성 및 관계 정의를 위해 온톨로지 개발 및 관리 도구인 프로테제, 웹온토 등을 참고하였다.

사용자 요구사항에 따른 기능들에 사용될 데이터의 항목 및 타입을 분석하여 데이터베이스 설계에 어떤 컬럼 및 테이블이 필요한지 분석하였다.

2. 시스템 설계

시스템 설계 단계는 시스템의 기능이 어떻게 구현될지 정의하는 단계이다. 시스템 기능이 작동할 수 있도록 기능을 설계하고 사용자의 상호작용을 위한 사용자 화면을 Microsoft Powerpoint 프로그램을 사용해 Figure 2와 같이 설계하였다. 시스템에서 사용되는 데이터를 설계하여 개체 관계 다이어그램(Entity Relation Diagram, ERD)으로 나타냈다.

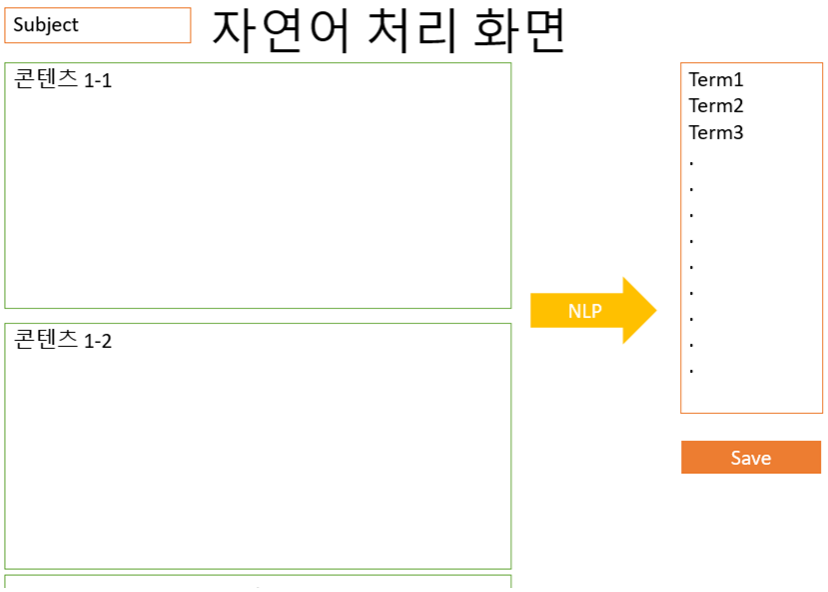


Figure 2. User Interface Design of Term Extraction Web Page

3. 시스템 구현 및 적용

시스템 코딩은 소프트웨어 엔지니어가 작업한 뒤 연구자가 검토하여 및 수정 요청하는 방식으로 구현하였다.

사용자 측 구현은 HTML (HTML5), JavaScript (Ajax, jQuery), CSS를 사용하고, 서버 측 개발언어로는 PHP와 Python을 사용하였다. 데이터 베이스는 ERD에 따라 MySQL을 이용하여 개발하고, PHP의 PDO (PHP Data Objects)인터페이스와 Python의 MySQL-Connector-Python 모듈을 이용하여 데이터베이스를 관리하도록 구현하였다.

4. 시스템 사용자 평가

1) 휴리스틱 평가

시스템의 휴리스틱 평가는 간호정보학 박사 3인 및 의료정보학 박사 2인, 총 5인의 전문가가 휴리스틱 평가 도구로 평가하였다.

휴리스틱 평가는 Neilson (1995)이 제안한 양식을 이용하여 보건의료정보학 전문가(박사 이상) 5인을 대상으로 진행하였다. 먼저 휴리스틱

평가 방법을 평가자들에게 소개하고, 휴리스틱 평가 원칙 리스트 및 심각성 수준 척도를 설명한 후 평가자들이 시스템을 사용할 수 있는 웹사이트 및 아이디, 비밀번호를 제공하고 수행해보아야 할 과제를 소개하였다. 평가자들이 수행해야 할 과제는 시스템 필수 기능에 준해 목록을 제공하였다.

시스템 휴리스틱은 Nielsen (1995)의 휴리스틱 10원칙에 심각성 수준 도구(Severity ranking scale, SRS)를 연계한 10개 문항과 개방형 질문으로 측정하였다. 휴리스틱 10원칙은 시스템 상태의 시각화; 시스템과 실세계의 일치; 사용자 제어 및 자율성; 일관성 및 표준; 오류 예방; 기억보다는 인식; 사용의 유연성 및 효율성; 공학적으로 간소화된 디자인; 오류 인식, 진단 및 복구 지원; 도움말 및 설명서로 구성되어있다. 각각의 원칙은 심각성 수준을 0: 사용성 문제 없음 1: 사용성 문제가 있으나 반드시 수정할 필요는 없음 2: 사용성 문제가 있으며 수정이 필요하지만 우선 순위가 상대적으로 낮음 3: 사용성 문제가 있으며 수정이 필요하고 우선 순위가 상대적으로 높음 4: 사용성 문제가 있으며 반드시 수정되어야 함으로 구성된 5점 척도로 평가된다.

2) PSSUQ 평가

시스템의 사용성 평가는 간호정보학 박사 3인 및 의료정보학 박사 2인, 총 5인의 전문가가 PSSUQ 평가도구로 평가하였다.

먼저 PSSUQ 도구를 평가자들에게 소개하고, 시스템의 목적인 건강정보 온톨로지 개발과, 건강정보 콘텐츠에 온톨로지 개념 연계에 대해 설명해 시스템의 두 가지 목적에 대해 모두 평가할 수 있도록 하였다. 평가자들이 시스템을 사용할 수 있는 웹사이트 및 아이디, 비밀번호를 제공하고 수행해보아야 할 시스템 기능 목록을 제공하였다.

시스템 사용성은 웹 환경 시스템의 사용편의성 평가를 위해 Lewis(1995)가 개발하고 2002년에 개정된 총 16문항의 PSSUQ 설문지 3판을 활용하여 측정하였다. 시스템의 유용성은 6개 항목, 정보의 질은 6개 항목, 인터페이스의 질 3개 항목, 전반적 평가는 1개 항목으로 이루어

져있다. 각 항목은 Lewis가 리커트 7점으로 평가한 것을 편의를 위하여 5점 척도로 변경하여 5: 매우 동의함 4: 동의함 3: 보통 2: 동의하지 않음 1: 매우 동의하지 않음으로 구성된 5점 척도로 평가된다. 본 연구에서는 Lewis의 PSSUQ 3판을 전은주(2013)가 한국어로 번역한 것을 보건의료학 전문가의 확인을 받아 사용하였다.

제 2 절 건강정보 온톨로지 개발

1. 건강정보 온톨로지 개발

건강정보 온톨로지는 개발된 건강정보 온톨로지 개발 및 건강정보 콘텐츠 연계 시스템을 활용하여, Ontology Development 101(Noy & McGuinness, 2001)의 방법론에 따라 개발하였다. 이 방법론에 근거한 건강정보 온톨로지 개발 단계는 다음과 같다.

1) 온톨로지의 영역과 범위 설정

본 연구의 목적은 국가건강정보포털의 검색을 개선하기 위한 건강정보 온톨로지의 개발 및 건강정보 콘텐츠와의 연계이다. 따라서 건강정보 온톨로지의 영역은 국가건강정보포털에서 제공하는 건강정보로 설정하였다. 건강정보 온톨로지의 범위는 국가건강정보포털에서 제공되고 있는 콘텐츠의 주제를 분석하여 질병 및 장애, 건강 문제, 진단 및 검사, 치료, 생활습관, 자가관리, 지원체계로 설정하였다.

2) 기존 온톨로지의 재사용 가능 여부 확인

Biomedical ontology, Google Scholar에서 기존에 개발된 온톨로지 중 재사용이 가능한 관련 온톨로지의 존재 여부를 검색하였다. 검색 결과 보건의료 영역 온톨로지 중 건강정보와 관련하여 MedlinePlus 온톨로지는 건강정보 주제의 상위 분류를 확인하는 데 도움이 되어 본 연구의 건강정보 온톨로지 개발에 재사용할 수 있었다. 또한 SNOMED CT는 보건의료 용어체계이자 온톨로지로서 매우 광범위한 용어 및 관계를

가지고 있어서 재사용 가능성을 확인하였다.

3) 온톨로지를 구성하는 주요 용어의 수집 및 열거

건강정보 온톨로지 개발 및 건강정보 콘텐츠 연계 시스템을 활용하여 국가건강정보포털의 건강정보 텍스트, 이미지, 동영상 콘텐츠로부터 주요 용어를 수집하였다. 텍스트 및 동영상 콘텐츠에서는 텍스트 및 동영상 스크립트를 자연어 처리하여 주요 용어를 수집하였다. 이미지 콘텐츠에서는 이미지의 제목 및 이미지에 속한 용어를 광학 문자 인식 방법을 활용해 수집하였다. 건강정보 텍스트, 이미지, 동영상 콘텐츠에서 시스템이 추출하지 못한 용어는 직접 검토하여 수집하였다.

그 밖에 건강정보 온톨로지의 완전성 및 균형성을 위해, 온톨로지 범위에 필요한 주요 용어들은 보건의료분야에서 널리 사용되는 용어체계 및 온톨로지에서 추가로 수집하였다. 예를 들어, MeSH (Medical Subject Headings) 용어체계나 SNOMED CT (Systematized Nomenclature of Medicine - Clinical Terms)를 활용하여 용어를 수집하였다.

4) 개념 및 개념의 계층 구조 정의

수집한 용어 중 핵심, 혹은 중심이 되는 독립 개념을 정의하고 개념의 동의어를 추가하여 용어체계(terminology)를 개발하였다. 수집한 용어를 열거한 뒤 같은 의미를 가진 여러 용어를 하나의 영문 개념명으로 정하고 한글 용어 중 하나는 한글 대표용어로 지정하였다.

개념을 정의한 후에는 기존의 관련 온톨로지의 상위 구조를 참고하여 Top-down 기법으로 개념 간의 상위 계층 구조를 설계하였다. 예를 들어, 가장 상위의 개념(슈퍼클래스)로는 SNOMED CT의 Top-level hierarchy 등 다른 온톨로지 및 분류체계를 참고하여 '질병 및 장애', '건강문제 및 증상', '진단 및 평가', '치료', '생활습관 관리', '자가관리', '지원 체계', '위험 요인' 등의 개념을 정의하고 각 슈퍼클래스의 하위 개념 관계도 상위 개념에 대해 'is-a' 관계가 성립할 수 있도록 정의하였

다. 예를 들어 '질병 및 장애' 의 하위로는 '심혈관계 질병 및 장애', '호흡기계 질병 및 장애', '내분비계 질병 및 장애', '비뇨기계 질병 및 장애' 등을 정의하였다. 또한 상위 수준과 하위 수준의 개념들 간 관계를 맺을 때는 SNOMED CT를 참고해 중간 개념을 검토하고 필요 시 그루핑 개념으로 추가하였다. 예를 들어, '유전 및 선천성 질병 및 장애'와 '귀의 선천성 형태 이상' 간 계층 관계를 맺을 때 중간 개념으로 '선천성 기형'을 추가해 계층 구조를 정의하였다.

5) 개념의 속성 관계 정의

개념의 속성 관계는 개념과 개념 간 상하 관계(is-a)가 아닌 다른 관계를 의미한다. 온톨로지에서 사용할 속성 관계의 종류를 정의하고, 이를 사용하여 개념 간의 관계를 정의할 수 있다. 예를 들어, 슈퍼클래스 간에는 Figure 3과 같은 속성 관계를 정의하였다. '질병 및 장애' 는 '건강문제 및 증상' 과 'hasSymptoms'의 속성 관계를, '질병 및 장애' 및 '건강문제 및 증상' 은 '치료'클래스와 'hasTreatment'의 속성 관계를 정의하였다.

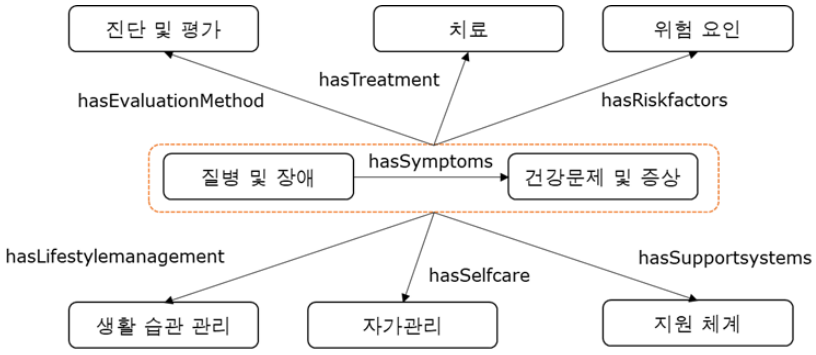


Figure 3. Attribute Relationships of Health Information Ontology Super Classes

6) 온톨로지 개념에 대한 동의어 수집

개념의 동의어는, 참여형 온라인사전인 우리말샘, 인터넷 커뮤니티 및 관련 문헌을 검색하여 가능한 많이 수집하였다. 개념의 한글 동의어는 건강정보 콘텐츠를 자연어 처리하여 추출하거나, 국가 암 정보 센터, 의료기관의 건강정보 제공 사이트와 네이버 등 포털에서 사용되고 있는 사용자 용어를 수집하여 추가하였다.

2. 건강정보 온톨로지 평가

온톨로지 평가는 간호정보학 전문가 4인(박사 수료 이상) 및 의료정보학 박사 1인, 총 5인의 전문가가 수행하였다. 온톨로지 평가 방법을 평가자들에게 소개하고, 평가자들이 온톨로지를 조회할 수 있는 웹사이트 및 아이디, 비밀번호를 제공하고 온톨로지를 조회한 뒤 평가 점수 및 의견을 작성하도록 하였다.

온톨로지의 구조와 내용은 Gomez-Perez의 온톨로지 평가 기준(Gomez-Perez, 2010)에 오류 여부 및 5점 리커트 척도를 결합한 14개 항목과 개방형 질문으로 평가하였다. 온톨로지 구조 평가 항목은 비일관성과 불완전성의 3개 항목으로 이루어져 있다. 각 항목은 '오류 있음' 혹은 '오류 없음'으로 평가하였다. 온톨로지의 내용 평가 항목은 일관성, 완성도, 간결성, 확대 가능성, 민감성의 11개 항목으로 이루어져 있다. 각 항목은 1: 매우 그렇지 않다 2: 약간 그렇지 않다 3: 보통이다 4: 약간 그렇다 5: 매우 그렇다로 평가하였다.

이 과정에 발생할 수 있는 질문에 답을 얻을 수 있도록 연구자의 연락처를 공개하였다.

제 3 절 건강정보 콘텐츠에 건강정보 온톨로지 개념 연계

1. 텍스트 건강정보 콘텐츠에 온톨로지 개념 연계

개발된 건강정보 온톨로지 개발 및 건강정보 콘텐츠 연계 시스템을

활용하여 텍스트 건강정보 콘텐츠에 건강정보 온톨로지 개념을 연계하였다. 시스템의 자동 연계 기능을 활용하여 건강정보 콘텐츠 주제 개념을 연계하고, 주제 개념과 목차명을 연결하여 생성한 키워드를 연계하고, 추가로 시스템의 수동 연계 기능을 활용하여 검색에 활용될 개념들을 연계하였다.

하나의 목차별 건강정보 콘텐츠에서 추출한 개념 연계 사례는 다음과 같다. '복부팽만' 이라는 제목을 가진 건강정보 콘텐츠의 주제 온톨로지 개념은 '복부 팽만(Abdominal Distension, Gaseous)'으로 모든 목차별 건강정보 콘텐츠에 자동으로 이 주제 개념을 연계하였다(Figure 4, A). 다음으로 목차인 '건강문제의 원인' 에서는 '복부 팽만 원인'을 키워드로 생성하고 '건강문제의 병태생리'에서는 '복부 팽만 병태 생리'를 키워드로 생성하여 연계하였다(Figure 4, B). 자연어 처리를 통해 추출한'Irritable Bowel Syndrome', 'Functional Dyspepsia', 'Constipation'와 같은 개념은 온톨로지 개념으로 추가하고, 주제 개념인 '복부 팽만'과 'hasCausativeAgent' 관계로 연계하였다(Figure 4, C). 또한 이들 개념의 유사어/동의어로 '과민성 장 증후군', '소화 장애', '변비'와 같은 용어를 검색에 활용될 수 있도록 추가하였다.

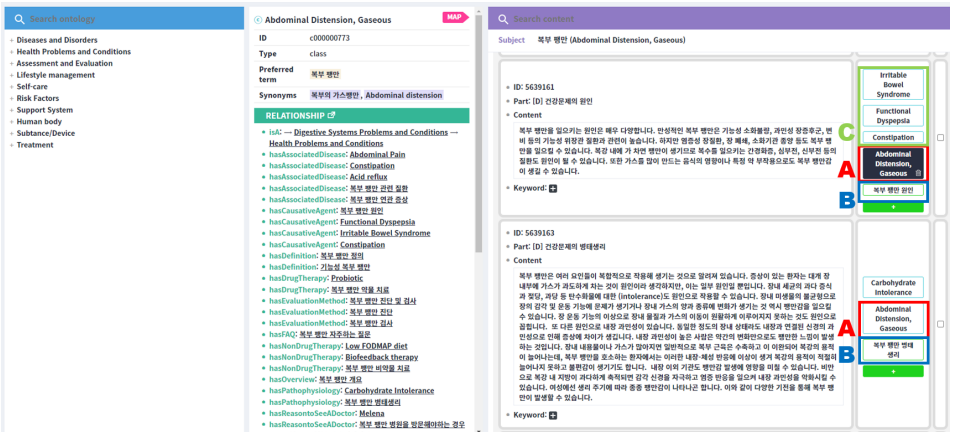


Figure 4. An Example of Health Information Content and Ontology Concept Annotation

2. 미디어 콘텐츠에 온톨로지 개념 연계

국가건강정보포털 건강정보 콘텐츠에 포함된 이미지, 동영상 등 미디어의 웹 접근성 강화를 위해 미디어 콘텐츠에 온톨로지 개념을 연계하였다. 시스템의 이미지의 문자 인식 기능을 통해 이미지 내 텍스트를 추출한 후 추출한 텍스트에서 추출한 개념을 연계하고, 동영상 스크립트의 자연어처리를 통해 추출한 개념을 연계하였다.

예를 들어, Figure 5와 같이 OCR을 이용하여 이미지 <그림. 골수의 구조>에 포함된 텍스트에서 '골수의 구조', '골수 세포'를 추출하고 연계해 검색이 가능하도록 하였다.

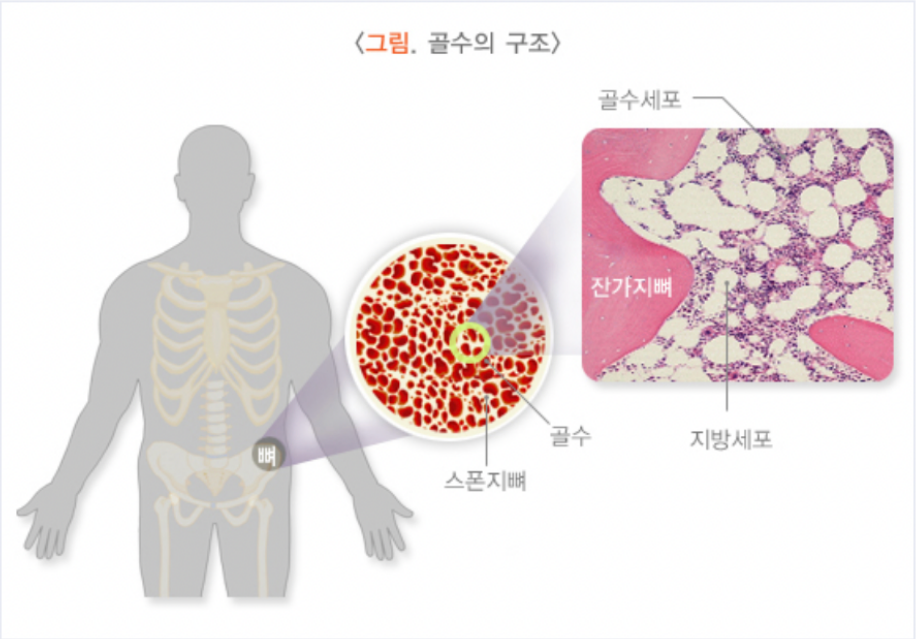


Figure 5. An Example of Image Content for OCR

3. 건강정보 콘텐츠 온톨로지 개념 연계 대상 콘텐츠

연계 대상 건강정보 콘텐츠는 2021년 신규 집필 및 재집필이 완료된 316개를 대상으로 하였다. 316개의 대상 콘텐츠를 슈퍼클래스 별로 살펴보면 Table 1과 같다.

Table 1. Number of Health Information Content by Superclass

슈퍼 클래스	콘텐츠 수
질병 및 장애	132
건강 문제	61
진단 및 검사	54
치료	47
생활 습관	2
자가 관리	5
위험 요인	12
지원 체계	3
합계	316

4. 건강정보 콘텐츠와 건강정보 온톨로지 개념 연계 원칙 및 적합성 평가

1) 건강정보 콘텐츠의 건강정보 온톨로지 개념 연계 원칙

국가건강정보포털 건강정보 콘텐츠의 연계는 최종적으로 연구자가 검토하고 확정하는 과정을 거쳤으며 이 때, Table 2와 같은 원칙에 따라 연계하도록 참여 연구자들을 가이드하였다. 또한 연계 결과를 다른 연구자가 재확인하는 과정을 거쳐 적절한 개념이 연계될 수 있도록 하였다.

Table 2. Annotation principles by content type

콘텐츠 종류	원칙
텍스트	-연계되는 개념 검색 시 조회되었을 때 의미상으로 사용자가 원하는 내용일지를 고려한다. -목차별 콘텐츠는 단락 수를 기준으로 주요 개념을 연계하여 연계되는 개념 수가 단락 수와 크게 차이나지 않도록 한다. -지나치게 많은 개념을 연계하여 검색 결과에 불필요하게 조회되지 않도록 한다.
이미지	-이미지 콘텐츠는 텍스트 콘텐츠와 같이 텍스트 기반 검색이 되지 않으므로, 상대적으로 더 많은 수의 개념을 연계할 수 있다. -이미지 내에서 추출된 용어를 모두 연계하지는 않도록 하되, 이미지 내에서 추출된 용어가 아니더라도 이미지를 잘 설명할 수 있는 주요 개념을 연계할 수 있다. -이미지 제목은 그대로 추출하여 연계한다.
동영상	-동영상 콘텐츠는 텍스트 콘텐츠와 같이 텍스트 기반 검색이 되지 않으므로, 상대적으로 더 많은 수의 개념을 연계할 수 있다. -동영상 스크립트에서 추출된 용어가 아니더라도 동영상의 주제 및 내용을 잘 표현하는 개념을 연계한다. -동영상 제목은 그대로 추출하여 연계한다.

2) 건강정보 콘텐츠에 건강정보 온톨로지 개념 연계 적합성 평가

건강정보 콘텐츠에 건강정보 온톨로지 개념을 연계한 적합성의 평가는 장인호(2011)의 연구를 참고하여 간호정보학 전문가 4인(박사 수료 이상) 및 의료정보학 박사 1인, 총 5인의 전문가가 수행하였다. 먼저 건강정보 콘텐츠의 건강정보 온톨로지 연계 원칙을 평가자들에게 소개하

고, 평가자들이 연계 결과를 조회할 수 있는 웹사이트 및 아이디, 비밀번호를 제공하였다. 먼저 평가자들에게 10개의 검색 키워드(온톨로지 개념의 한글 대표용어 및 동의어)를 제공하고, 시스템 Administration 탭의 Evaluation 기능을 이용하여 검색 키워드를 입력한 뒤, 출력되는 연계된 콘텐츠 목록을 확인하도록 하였다. 연계된 콘텐츠를 각각 조회하여 해당 콘텐츠와 검색 키워드의 연계가 연계가 적합한지에 대해 예 혹은 아니오로 답변할 수 있도록 안내하였다.

이 과정에 발생할 수 있는 질문에 답을 얻을 수 있도록 연구자의 연락처를 공개하였다.

제 4 절 연구의 윤리적 고려

본 연구는 전문가 평가 설문을 위해 서울대학교 생명윤리위원회(Institutional Review Board)의 심사를 받고 연구를 승인받았다(IRB No. 2207/003-002).

본 연구는 전문가 평가 과정에서 개인 식별 정보를 수집하거나 기록하지 않아, 서면 동의 면제를 승인받아 구두로 동의를 받았다.

제 4 장 연구 결과

제 1 절 건강정보 온톨로지 개발 및 건강정보 콘텐츠 연계 시스템

건강정보 온톨로지 개발 및 온톨로지 개념과 건강정보 콘텐츠 연계 시스템은 크게 콘텐츠 입력, 온톨로지 개발 및 관리, 건강정보 콘텐츠에 온톨로지 개념 연계의 세 가지 목적으로 구분하여 개발하였다. 해당 시스템은구조화된 건강정보 콘텐츠를 각각의 유형에 따라 import 하고, 해당 콘텐츠에서 자연어처리나 광학문자인식 기능을 활용하여 온톨로지를 개발하고, 관리할 뿐 아니라 개발된 온톨로지의 개념을 건강정보 콘텐츠와 연계하고 그 결과를 출력할 수 있는 시스템이다.

1. 시스템 분석 결과

1) 시스템 기능 분석 결과

시스템이 필요로 하는 기능을 사용 사례 중심으로 분석한 결과 크게 온톨로지 개발 및 관리, 건강정보 온톨로지 연계, 시스템 관리의 세 가지 분류로 필수 기능이 도출되었다. 해당 분류에 따라 사용자가 직접 수행하는 기능과 시스템 내에서 자동으로 이루어지는 기능을 구분하여 나타낸 결과는 표 3과 같이 정의하였다.

Table 3. Functions for Ontology Development and content Annotation System

분류	기능	
	사용자	시스템
온톨로지 개발 및 관리	개념 추가/수정/삭제	개념 관계 시각화
	한글 대표 용어 추가/수정/삭제	의학용어 사전과 비교 결과 제시
	개념의 동의어 추가/수정/삭제	추출한 용어 빈도수 제시
	개념 및 동의어 검색	추출한 용어의 우선순위 제시
	콘텐츠 파일 import	이미지에서 용어 추출
	자연어처리 패키지 선택	콘텐츠 목차에 따른 개념 간 관계 자동

		설정
	말뭉치에 용어 추가	텍스트 콘텐츠의 자연어처리를 통한 용어 추출
	온톨로지에 용어 및 관계 추가	동영상 스크립트의 자연어처리를 통한 용어 추출
	개념 간 관계 수동 설정	기존 개발된 온톨로지 개념과 비교 결과 제시
	개념에 연계된 콘텐츠 조회	
콘텐츠에 온톨로지 개념 연계	자연어 처리 용어 선택 연계	목차별 콘텐츠 목차명 활용 자동 연계
	콘텐츠에 연계된 개념 수정	콘텐츠 주제 자동 연계
	온톨로지 개념 검색하여 콘텐츠에 연계	콘텐츠 목차 식별
		콘텐츠 슈퍼클래스 분류
		콘텐츠 주제 분류
시스템 관리	산출물 출력	온톨로지 평가 쿼리 수행
	온톨로지 파일 업로드	
	사용자 관리	
	온톨로지 생성, 삭제	

또한 정의된 기능을 Figure 6의 유스케이스 다이어그램(Use case Diagram)과 같이 표현하였다.

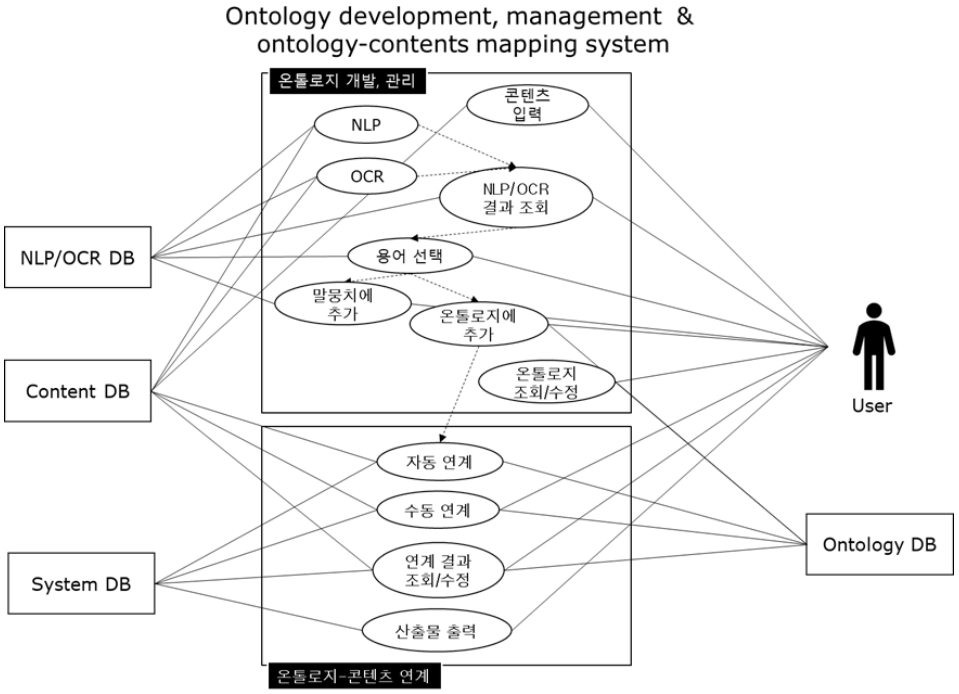


Figure 6. Use-case Diagram of Ontology Development and Content Annotation System

건강정보 콘텐츠의 주제를 성별, 연령, 질병 행태, 유전성/선천성 여부, 만성질환 여부, 신체계통, 검사 및 치료 유형으로 분류하여 검색할 수 있는 기능 구현을 위해서는, Table 4와 같은 규칙을 정의하였다.

Table 4. Example Rules for Categorizing Health Information Content

Category	Code	Description	Rules
성별	M	남성	IF subject concept has [Men Diseases and Disorders] as a parent concept THEN classify sex group of subject as 남성 [M]
	F	여성	IF subject concept has [Women Diseases and Disorders] as a parent concept THEN classify sex group of subject as 여성 [F]
연령	P0M--P6Y	영유아	IF subject concept has [Infants and Toddler Diseases and Disorders] as a parent concept

			THEN classify age group of subject as 영유아 [P0M--P6Y]
	P6Y--P19Y	소아/청소년	IF subject concept has [Children and Adolescents Diseases and Disorders] as a parent concept THEN classify age group of subject as 소아/청소년 [P6Y--P19Y]
	P19Y--P65Y	청년	IF subject ontology concept has [Young Adult Diseases and Disorders] as a parent concept THEN classify age group of subject as 청년 [P19Y--P65Y]
	P65Y--P9999Y	노인	IF subject concept has [Seniors Diseases and Disorders] as a parent concept THEN classify age group of subject as 노인 [P65Y--P9999Y]
질병 행태	C00-D48	암	IF subject concept has [Cancers] as a parent concept THEN classify morphology group of subject as 암 [C00-D48]
	A00-B99	감염	IF subject ontology concept has [Infections] as a parent concept THEN classify morphology group of subject as 감염 [A00-B99]

건강정보 콘텐츠의 구조는, 위와 같은 내용 분류 및 상세 분류에 따라 구분되어 표준 목차를 따라 구조화되어 있다. 콘텐츠에 따라 모든 표준 목차가 사용될 수도, 아닐 수도 있으나 기본 구조는 모두 표준 목차에 따라 동일하다. 따라서 이 구조를 사용하여 건강정보 콘텐츠를 시스템에 입력(import)하고 연계(annotation)하는 기능을 지원할 수 있도록 Table 5와 같이 기본 검색 키워드 및 동의어 키워드 규칙을 정의하였다.

Table 5. Example Rules for Auto-Generated Keywords

목차 유형	목차 명	기본 키워드	동의어
Type K 검사	검사의 개요	* 개요	* 배경, * 무엇인가, * 무엇인가요, * 이란
	검사 목적	* 목적	* 왜, * 이유
	검사 적응증 및 금기증	* 적응증 및 금기증	* 언제 할 수 있고 또 언제 하면 안되는지
	검사 적응증 및 금기증	* 적응증	* 어떤 때, * 언제, * 언제 하는지
	검사 적응증 및 금기증	* 금기증	* 안 되는 경우, * 금지, * 받을 수 없는 경우, * 검사

		불가 경우
검사 검체	* 검체	* 검사재료
검사 항목	* 항목	* 검사 종류
검사 장비	* 장비	* 기구, * 도구, * 기기, * 에 사용되는 기구, * 에 사용되는 장비, * 할 때 어떤 장비 이용
검사 준비	* 준비	* 전에 해야할 일, * 준비사항, * 미리 준비
검사 절차	* 절차	* 순서, * 진행 과정, * 과정
검사 결과 해석	* 결과 해석	* 결과, * 해석, * 결과 분석, * 결과 의미
검사 관련 주의사항 및 부작용	* 주의사항 및 부작용	* 조심해야할 점 및 나타날 수 있는 부작용
검사 관련 주의사항 및 부작용	* 주의사항	* 주의할 점, * 주의점, * 조심해야할 점
검사 관련 주의사항 및 부작용	* 부작용	* 이상 반응, * 이상 증상
검사 관련 자주하는 질문	* 자주하는 질문	* FAQ, * 많이 묻는 질문, * 대표 질문, * 관련 많이 물어보는 질문, * 궁금증, * 궁금한 점

자연어처리를 한 결과 추출된 용어를 사용자가 선택할 때 도움이 되도록 결과 제시의 우선순위를 정의하였다. 먼저 추출 빈도 순으로 나열한 뒤 첫째, 건강정보 온톨로지에 기존에 존재하는 개념 및 용어와 일치하는 용어 둘째, 의학용어 사전에 존재하는 용어와 일치하는 용어를 각각 다른 색의 하이라이트로 표시되도록 규칙을 정의하였다.

시스템 기능에 따른 화면 분석결과로 Content Import 화면, Term Extraction 화면, Ontology 화면, Annotation 화면, Administarion 화면이 필요한 것으로 분석되어 이를 바탕으로 시스템 인터페이스를 도출하였다.

2) 시스템 데이터 분석 결과

시스템 데이터 분석 결과로 61개의 데이터 컬럼이 도출되었으며 이를 데이터 베이스 설계 시 12개의 테이블로 구성하여 반영하였다.

2. 시스템 설계 결과

1) 시스템 기능 설계 결과

자연어처리의 기능 구현을 위해서는 한국어 정보처리를 위한 파이썬 패키지 KoNLPy (코엔엘파이)를 활용하였다. 코엔엘파이에서 제공하는 말뭉치를 사용하고, 이에 추가로 콘텐츠의 특성을 고려하여 KMLE 의학 용어 사전의 용어를 수집하여 말뭉치에 추가하여 자연어처리 시 용어가 더 잘 추출될 수 있게 하였다. 또한 시스템을 사용하며 사용자가 온톨로지에 수동으로 추가하는 용어도 말뭉치에 추가하는 방법을 통해 지속적으로 말뭉치가 확장되도록 설계하였다. 형태소 분석에는 서울대학교 연구실에서 개발한 Kkma (꼬꼬마)와 Shineware에서 개발한 Komaran (코모란) 패키지를 선택적으로 활용하도록 설계하였다. 꼬꼬마는 다른 형태소 분석기에 비래서 분석의 질이 뛰어난 장점이 있으며, 코모란은 처리 속도가 빠른 장점이 있다.

또한 이미지 콘텐츠에 대한 개념 추출을 위해서는 광학 문자 인식(OCR)을 활용한 용어 추출 방안을 모색하였다. 해당 기능에는 HP에 의해 개발되고 오픈 소스화되어 Google에 의해 개발되고 있는 Tesseract를 활용하였다.

시스템 화면은 다음과 같이 'Content import', 'Term extraction', 'Ontology', 'Annotation', 'Administration' 화면(탭)으로 설계하였다.

가. Content import 화면

콘텐츠 입력 화면은 워드 포맷에 작성된 건강정보 콘텐츠를 입력(import)하면 제목에 따라 온톨로지 개념을 검색하여 연계하고, 정의된 규칙에 따라 온톨로지 계층 구조를 활용하여 인구학적 유형, 질병 행태 유형, 유전성/선천성 여부, 만성질환 여부, 신체계통, 치료 유형, 검사 유형으로 자동으로 분류되도록 구현하였다.

예를 들어, Figure 7의 혈관조영술은 제목에 따라 온톨로지 개념 Angiography에 연계되고, 온톨로지 계층 구조에 따라 자동으로 검사 유형 중 '영상의학검사'로 분류된다.

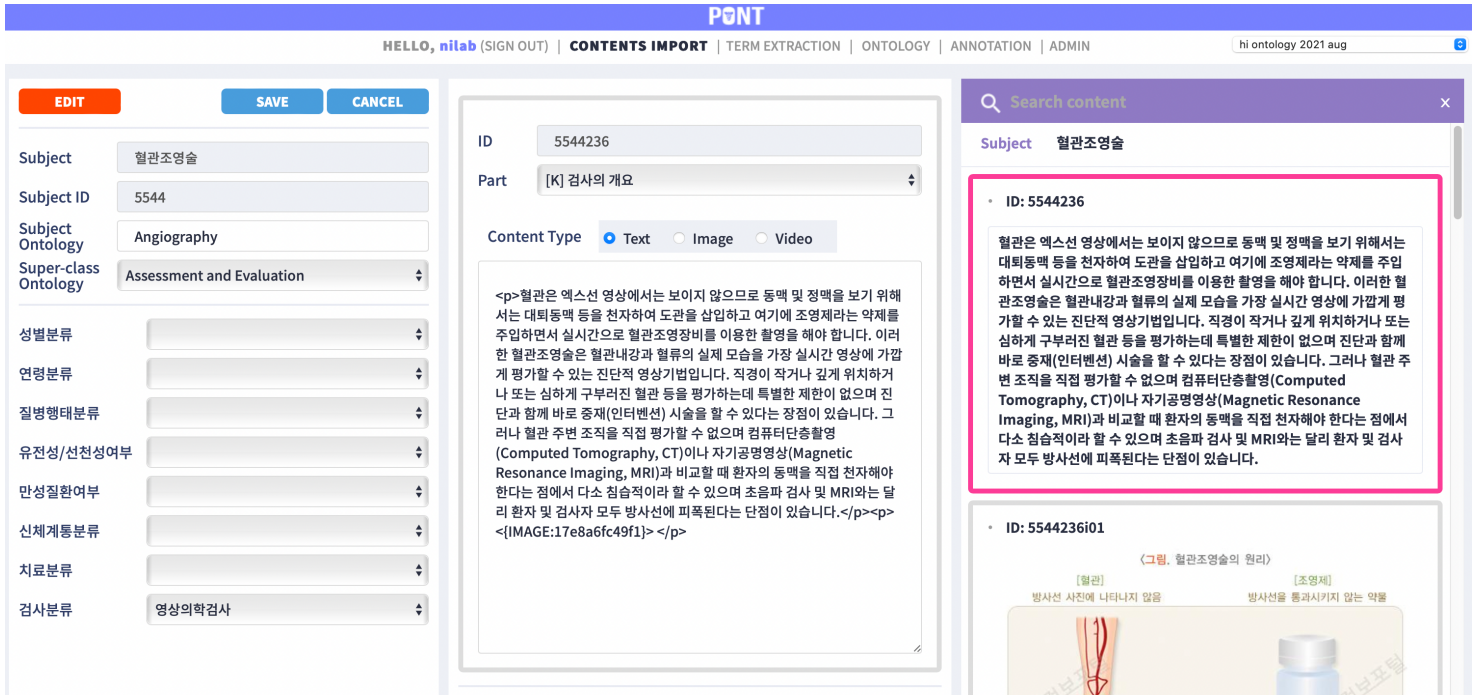


Figure 7. An Example of ‘Content Import’

나. Term extraction 화면

Term extraction 화면에서는 목차별 건강정보 텍스트 콘텐츠를 입력받아 두 가지 자연어 처리 패키지 '꼬꼬마'와 '코모란' 중 하나를 이용하여 자연어처리를 한 후 그 결과를 조회하고 필요한 용어를 선택할 수 있도록 구현하였다(Figure 9). 자연어처리 결과 추출된 용어는 먼저 빈도순으로 정렬한 뒤 가장 위에는 현재까지 개발된 온톨로지에 포함되어 있는 용어를 가장 위에 주황색으로 나타내고 다음으로 의학용어사전 말뭉치에 포함되어있는 용어를 다음으로 파란색으로 나타내어 우선순위 규칙에 따라 조회할 수 있도록 하였다. 우측 옵션에서는 선택한 용어를 말뭉치에만 추가할 것인지, 온톨로지 개념으로 관계유형과 함께 추가할 것인지, 목차별 건강정보 콘텐츠에 연계할 것인지 선택할 수 있다. 해당 화면에서는 텍스트 콘텐츠의 자연어처리뿐 아니라 동영상 스크립트의 자연어처리도 시행하고, 이미지 콘텐츠에서 광학 문자 인식 기술을 이용하여 Figure 8과 같이 용어를 추출하여 콘텐츠에 연계하는 기능을 구현하였다.

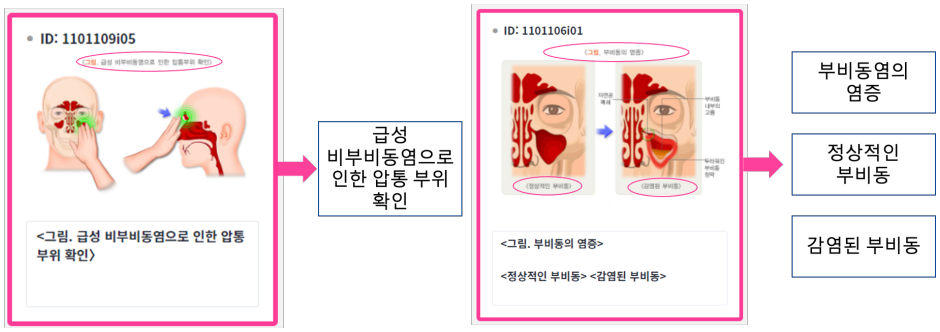


Figure 8. An Example of Term Extraction Using OCR

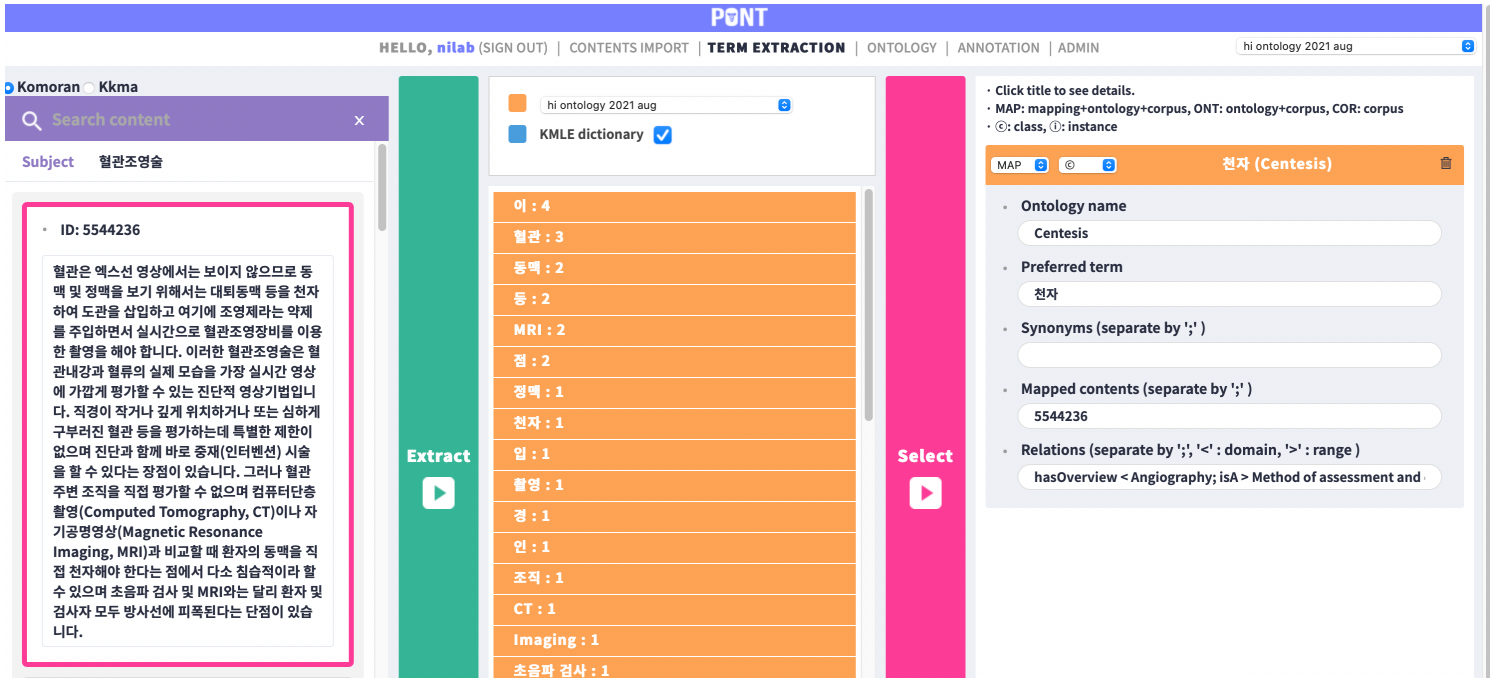


Figure 9. An Example of NLP

다. Ontology 화면

온톨로지에서는 건강정보 콘텐츠에서 추출한 개념 및 연구자가 수동으로 추가한 온톨로지 개념을 활용하여 온톨로지의 개념, 동의어, 관계를 관리하고 시각화할 수 있도록 Figure 11과 같이 구현하였다. 해당 화면에서는 화면에서 개념 및 동의어, 관계를 생성, 수정, 삭제하고 계층 구조를 따라 조회하고 연계된 콘텐츠를 조회할 수 있다. 개념이 맺고 있는 관계의 시각화는 Figure 10과 같이 해당 개념을 출발점으로 하여 맺고 있는 다양한 계층 및 속성 관계를 보여주는 기능이다.

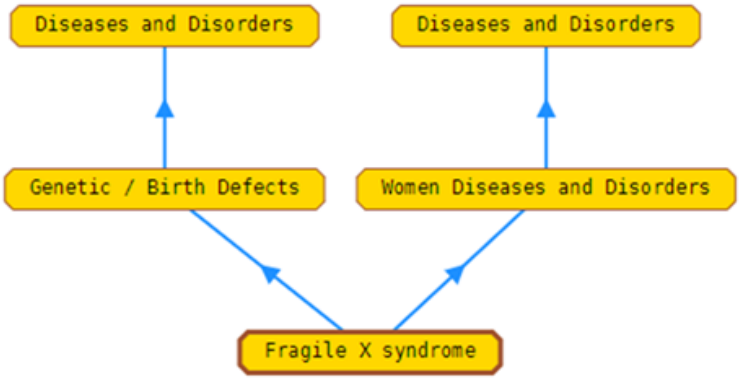


Figure 10. An Example of Ontology Visualization

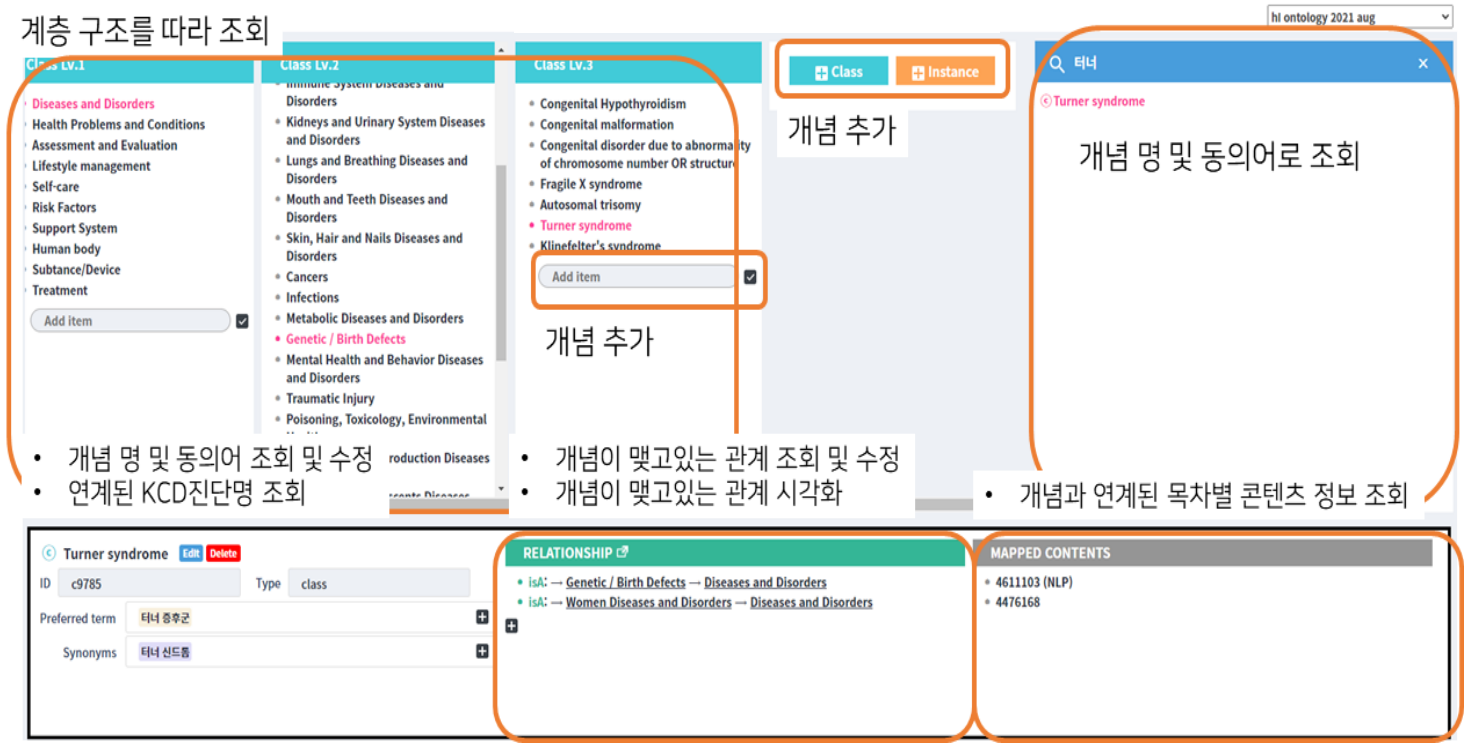


Figure 11. An Example of Ontology Management

라. Annotation 화면

Annotation 화면에서는 자동으로 연계된 건강정보 콘텐츠의 주제 온톨로지 개념 및 목차별로 자동으로 생성되어 연계된 온톨로지 개념을 검토하고, 연구자가 추가로 온톨로지 개념을 조회하여 연계할 수 있도록 Figure 12와 같이 구현하였다. 건강정보 콘텐츠의 주제 개념인 '치주질환'은 모든 목차별 건강정보 콘텐츠에 자동으로 연계되어 조회되고, 자연어 처리에서 추출되어 연계된 개념인 '치은염', '치주염'도 해당 목차별 콘텐츠에서 조회된다. 또한 목차별 건강정보 콘텐츠에서 자동으로 생성된 '치주질환 종류'도 연계된 것을 확인할 수 있다. 또한, 좌측과 같이 온톨로지에서 개념을 검색하면 동의어, 다른 개념과의 관계뿐 아니라, 연계된 건강정보 콘텐츠를 확인할 수 있다.

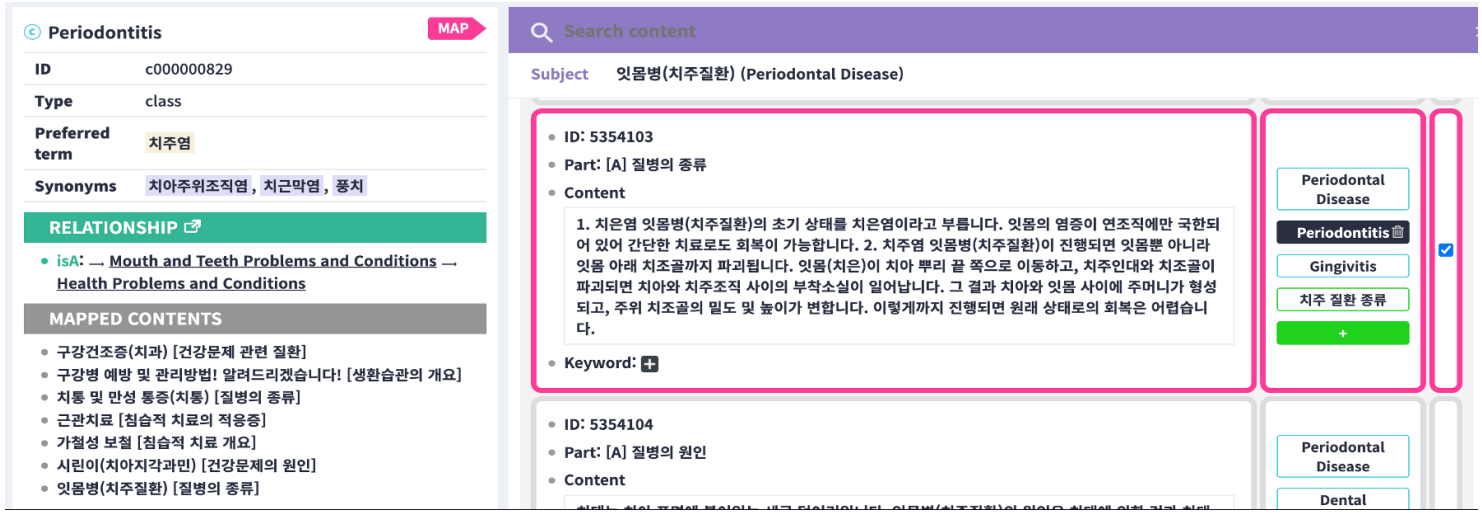


Figure 12. An Example of Health Information Content with Ontology Concept Annotation

마. Administration 화면

Administration 화면에서는 우선 사용자를 생성, 삭제하거나 권한을 구분하여 설정하는 기능을 구현하였다. 사용자 권한은 Table 6과 같이 세가지로 구분하여 구현하였다. Guest는 개발된 온톨로지 및 콘텐츠 연계 결과를 조회할 수 있는 권한을 가지며, 온톨로지 및 콘텐츠 연계 평가자 들에게 해당 권한을 부여하였다. User는 온톨로지 개발 및 콘텐츠 연계 조회뿐 아니라 작업을 수행하는 권한을 가지며, 작업을 수행하는 연구자들에게 해당 권한을 부여하였다. Administrator는 조회, 작업 수행 기능에 추가로 작업 산출물을 출력하고, 사용자를 등록하고 권한을 부여하는 관리 권한을 가진다.

Table 6. System Functions by User Level

User level	Function		
	조회	작업 수행	산출물 출력/ 사용자 관리
Guest	O		
User	O	O	
Administrator	O	O	O

온톨로지 평가 기능은, 쿼리를 이용해 특정 개념의 부모 개념, 자식 개념, 속성 관계를 맺고 있는 개념 및 연계된 콘텐츠를 조회할 수 있도록 구현하였다. Figure 13과 같이 4가지 중 자식 개념을 조회하는 조건을 선택하고, ‘Hypertension’ 개념명을 입력하면 우측과 같이 자식 개념들을 조회하여 온톨로지의 내용을 보다 쉽게 평가할 수 있다.

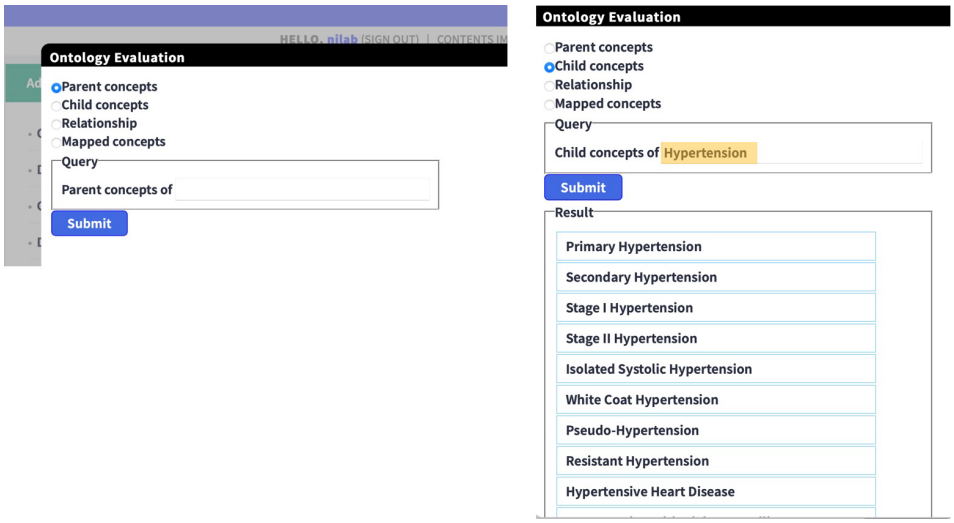


Figure 13. An Example of Ontology Evaluation

산출물 출력 기능은 Figure 14와 같이 건강정보 온톨로지, 건강정보 온톨로지와 콘텐츠 연계 결과 등을 테이블 단위로 다운로드할 수 있도록 구현하고, 업로드 기능은 엑셀 파일 형태로 작업하여 개발된 온톨로지를 시스템에 업로드하여 관리할 수 있도록 구현하였다.

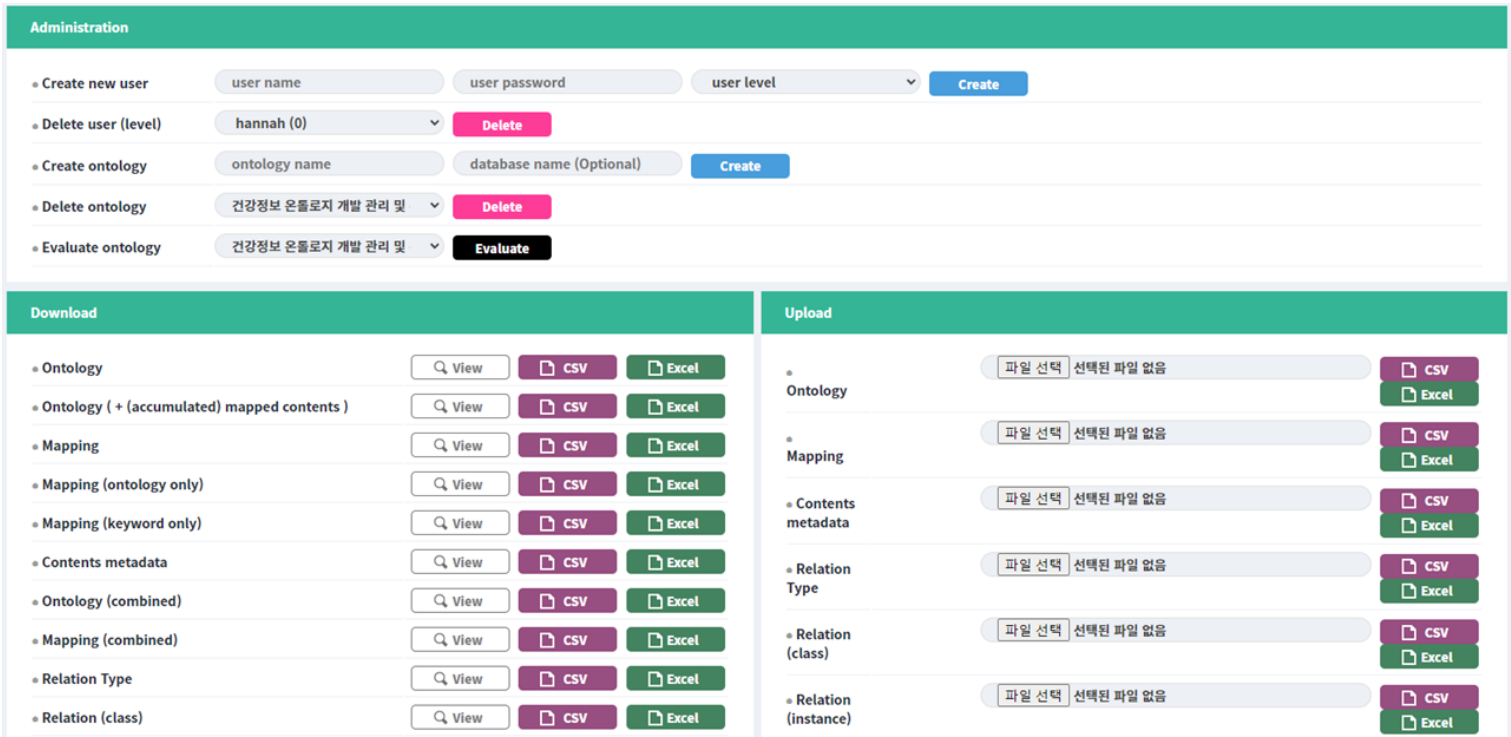


Figure 14. An Example of Administration

2) 시스템 데이터 설계 결과

시스템 데이터 설계 결과로는 요구사항에 따라 필요한 데이터 항목을 정의하고 데이터의 구조를 보여줄 수 있는 개체 관계 다이어그램을 설계하였다.

온톨로지 데이터는 개념 정보, 동의어 정보, 관계 정보로 이루어져 있다. 건강정보 콘텐츠 데이터는 콘텐츠 종류 정보, 콘텐츠 주제 정보, 콘텐츠 목차 정보로 이루어져 있다. 연계 데이터는 온톨로지 개념과 건강정보 콘텐츠의 관계정보가 해당하며 시스템 관리 정보는 데이터베이스 정보 및 사용자 정보 등으로 이루어져 있다. 이에 따라 시스템의 데이터 구조를 보여줄 수 있는 개체 관계 다이어그램을 Figure 15와 같이 설계하였다. 데이터의 내용에 따라 개체를 구분하고 이름을 지은 뒤 그 관계를 표현하였다. 본 연구의 개체 관계 다이어그램은 크게 온톨로지 영역과 콘텐츠 영역으로 구성된다. 우선 온톨로지 영역은 온톨로지 개념 테이블, 용어 테이블, 관계 종류테이블 및 개념과 개념 간 관계 테이블로 구성하였다. 콘텐츠 영역은 콘텐츠 아이디 및 이름 테이블, 콘텐츠의 메타데이터 테이블, 텍스트 콘텐츠 테이블, 이미지 콘텐츠 테이블 및 동영상 콘텐츠 테이블로 구성하였다.

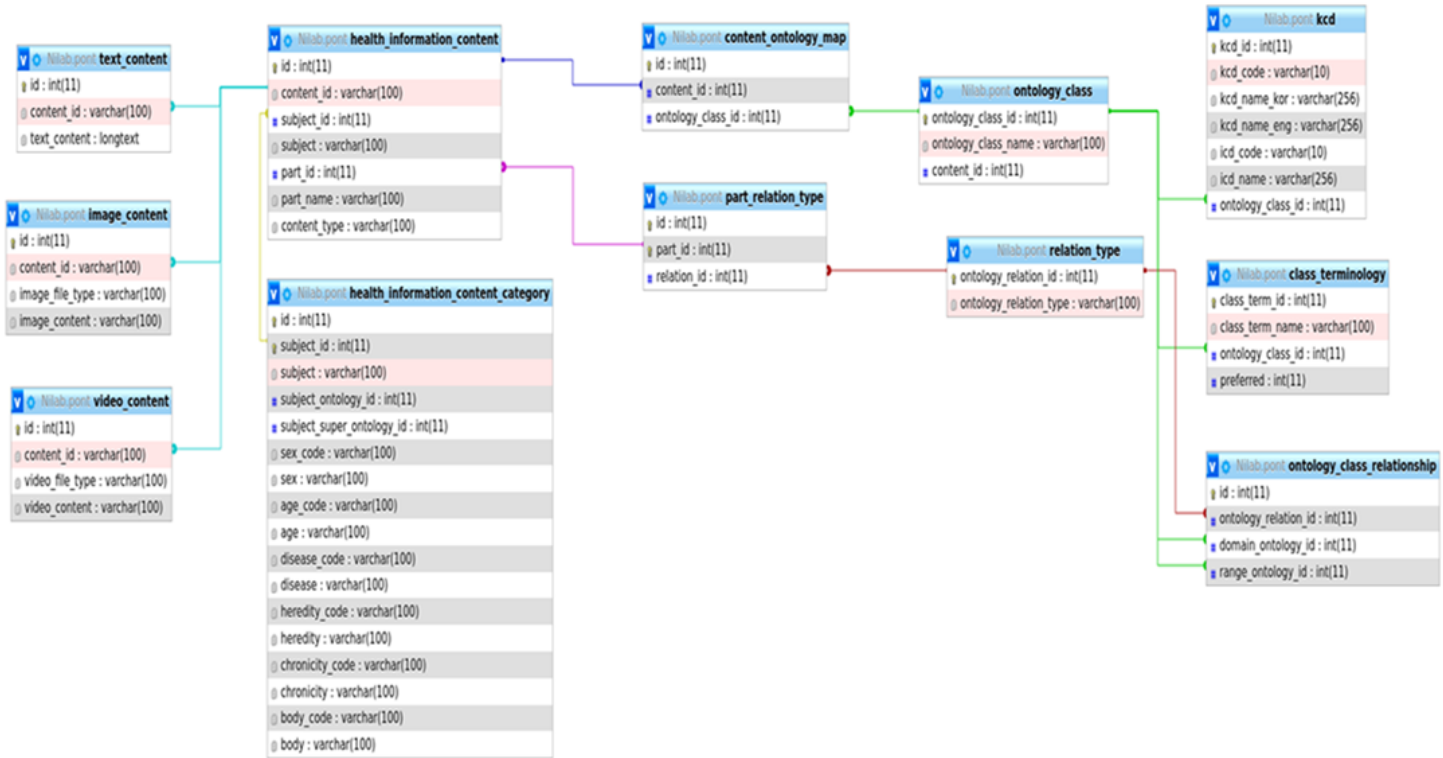


Figure 15. Entity Relation Diagram of Ontology Development and Content Annotation System

3. 시스템 구현 및 적용 결과

본 시스템은 일반적인 웹 환경에서 웹브라우저(Internet Explorer, Edge, Chrome, Safari, Firefox 등)를 이용하여 구동하도록 개발되었다.

MS Excel 포맷의 온톨로지 및 콘텐츠 데이터베이스의 입출력은 PhpSpreadsheet 라이브러리를 이용하여 처리하도록 개발되었다.

4. 사용자 평가 결과

1) 평가자 특성

휴리스틱 평가는 총 5명의 간호정보학 및 의료정보학 전문가(박사 이상)가 시행하였다. 간호정보학 전공자 3명, 의료정보학 전공자 2명은 모두 시스템 개발 경험이 있는 전문가이다.

2) 평가 결과

가. 휴리스틱 평가 결과

휴리스틱 평가 결과는 평가 의견 및 점수로 도출되었다. 휴리스틱 원칙 10가지에 대해서, 평가 점수는 Table 7과 같이 평가자 별, 항목별로 도출되었고, 사용성 문제가 있어 반드시 수정되어야 하는 심각성 수준 3점 이상의 결과는 휴리스틱 6번 기억보다는 인식, 10번 도움말 및 설명서 항목 2개이었다. 10번 항목에 대한 서술형 의견으로는, 시스템이 제공하는 기능에 대한 설명이 있으면 좋을 것 같다는 것이 있었다. 그 외에 5명의 평가자 중 과반 이상에게서 문제가 있는 것으로 평가 받은 원칙은 3번 사용자 제어 및 자율성, 5번 오류 예방, 9번 오류 인식, 진단 및 복구 지원이 있었다. 5번, 9번 항목의 오류 예방 및 인식에 대해서, 용어를 추출하는 과정에서 기 등록된 용어를 제외하여 보여주거나 별도로 표시하여 중복하여 작업하지 않도록 하는 기능이 있으면 좋겠다는 서술형 의견이 있었다. 9번 오류 인식, 진단 및 복구 지원에 대해서는 오류 발생 시 연락 가능한 정보가 있으면 좋겠다는 의견이 있었다. 그 외 서술형 의견에서는, 시스템이 직관적이고 효율적이라는 평가가 있었다.

Table 7. Result of Heuristic Evaluation

Heuristic/ 평가자	1	2	3	4	5
1 시스템 상태의 시각화	0	0	0	0	0
2 시스템과 실세계의 일치	0	0	0	0	1
3 사용자 제어 및 자율성	2	1	0	1	0
4 일관성 및 표준	0	1	0	0	0
5 오류 예방	2	1	0	1	1
6 기억보다는 인식	1	3	0	0	0
7 사용의 유연성 및 효율성	1	1	0	0	0
8 공학적으로 간소화된 디자인	0	0	0	0	0
9 오류 인식, 진단 및 복구 지원	2	1	1	1	0
10 도움말 및 설명서	3	4	1	1	0

나. PSSUQ 평가 결과

PSSUQ 평가 결과는 항목 별 평균 점수가 Table 8과 같이 3.4~5.0점으로 나타났으며 전체 평균 점수는 4.3점으로 나타났다. 가장 낮은 점수를 얻은 항목은 5번 '이 시스템은 배우기가 쉽다'였으며 전반적인 시스템 만족도를 묻는 16번 항목의 점수는 4.4점이다. 서술형 의견에서는 정보의 질과 관련하여 온톨로지를 개발했을 때 제대로 되었는지 평가할 수 있는 기능이 있으면 좋겠다는 의견이 있었다.

시스템 전반에 대한 서술형 의견에서는, 온톨로지 개발 및 콘텐츠 연계 측면에서 사용자 인터페이스가 단순하고 직관적이어서 다양한 정보를 빠르게 생성하고 쉽게 관리할 수 있을 것으로 보인다는 의견이 있었으며. 전반적으로 온톨로지 개발과 건강정보 콘텐츠 연계를 효율적으로 수행할 수 있도록 설계되어 있으며 사용자 편의성에 많은 노력을 기울인 것으로 보인다는 평가를 받았다.

Table 8. Result of PSSUQ Evaluation

No	Item	Mean
1	전반적으로 나는 온톨로지 개발 및 콘텐츠 연계 시스템의 쉬운 사용방법에 만족한다.	3.8
2	온톨로지 개발 및 콘텐츠 연계 시스템은 사용하기에 간단하다.	4.4
3	나는 이 시스템으로 온톨로지 개발 및 건강정보 콘텐츠 연계를 빨리 끝마칠 수 있다.	4.2
4	나는 이 시스템을 사용하는 것이 편하다.	3.8
5	이 시스템은 배우기가 쉽다.	3.4
6	나는 이 시스템을 이용해서 온톨로지 개발 및 건강정보 콘텐츠 연계를 잘 할 수 있다고 생각한다.	4.2
7	에러 메시지는 문제를 해결하는 법을 분명히 알려준다.	4.0
8	내가 시스템 이용 시 실수를 범하면, 언제든 빠르고 쉽게 해결할 수 있다.	4.8
9	이 시스템이 제공하는 온톨로지 및 건강정보 콘텐츠와의 연계에 대한 정보들이 확실하다.	4.6
10	내가 필요로 하는 온톨로지 및 건강정보 콘텐츠와의 연계에 대한 정보를 찾는 것이 쉽다.	4.2
11	제공하는 정보가 온톨로지 개발 및 건강정보 콘텐츠 연계 과업을 마치는데 효과적으로 도움을 준다.	4.4
12	시스템 화면에 대한 정보의 나열은 분명하다.	5.0
13	이 시스템의 인터페이스는 흥미로웠다.	5.0
14	나는 이 시스템의 인터페이스를 쓰는 것이 좋다.	5.0
15	이 시스템은 내가 기대한 기능들을 다 갖추고 있다.	4.2
16	전체적으로 나는 이 시스템에 대해 만족스럽다.	4.4
	Total Mean	4.3

3) 평가 결과의 분석 및 활용

가. 휴리스틱 평가 결과의 분석 및 활용

시스템 기능에 대한 도움말이 부족하다는 의견을 반영하여 시스템 사용자를 위한 화면 중심의 설명서를 작성하였다. 오류 예방 및 인식에서 용어 추출 시 기 존재하는 용어와 중복되는 경우에 대한 개선으로는, 용어 추출 후 기존 개발된 온톨로지의 용어와 비교하여 상위로 표시하여 구분될 수 있도록 하고, 추출된 용어를 온톨로지에만 추가하거나, 이미 존재하는 경우 추출된 건강정보 콘텐츠에 연계만 하거나 선택할 수 있는 기능을 추가로 개발하였다.

나. PSSUQ 평가 결과의 분석 및 활용

온톨로지 개발 시 평가할 수 있는 기능이 있으면 좋겠다는 의견에 대해, 온톨로지 개념의 계층 관계 및 속성 관계의 값을 결과로 보여주는 쿼리 활용 온톨로지 평가 기능을 추가로 개발하였다.

제 2 절 건강정보 온톨로지 개발 및 평가 결과

1. 건강정보 온톨로지 개발 결과

개발된 건강정보 온톨로지에는 총 3,823개의 개념 및 12,082개 동의어, 개념 간 계층 관계 18,237쌍과 속성 관계 15,362쌍이 포함되어 있다. 건강정보 온톨로지의 최상위 클래스인 슈퍼클래스는 Figure 16과 같이 구성되어 있으며. 최상위 클래스별로 포함된 개념 및 동의어 수는 Table 9와 같다.

Figure 16. Super classes of Health Information Ontology

Table 9. Number of Ontology Concepts and Synonyms by Superclass

Superclass	No. of concepts	No. of synonyms
Diseases and Disorders	1,683	5,450
Health Problems and Conditions	588	1,757
Assessments and Evaluations	525	1,532
Treatments	642	2,122
Lifestyle Managements	138	427
Self-care	55	234
Risk factors	179	525
Support systems	13	41
Total	3,823	12,082

각 슈퍼클래스에 대해서는 해당 개념에 속하는 개념들을 정의하고 수준에 따라 계층 관계를 맺어 Figure 17과 같이 개발하였다. 예를 들어, 질병 및 장애 슈퍼클래스의 하위 수준 개념으로는 심혈관계 질병 및 장애, 호흡기계 질병 및 장애, 내분비계 질병 및 장애, 비뇨기계 질병 및 장애 등이 포함될 수 있다.

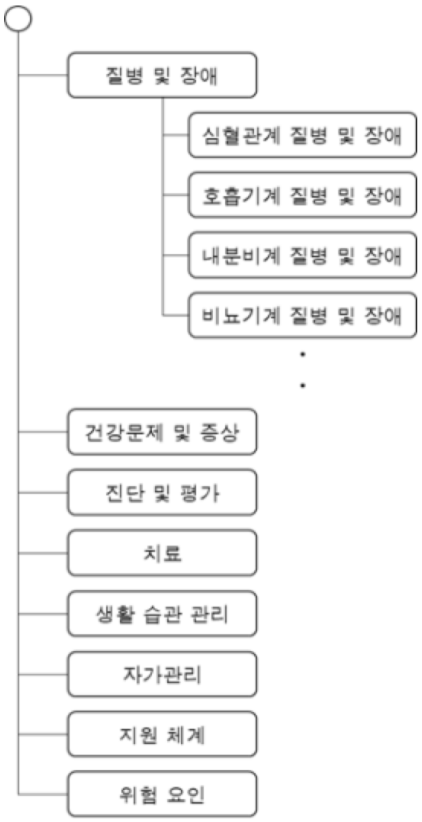


Figure 17. Examples of the Second-Level Classes of Health Information Ontology

개발된 건강정보 온톨로지에는 41개의 속성 관계 유형이 포함되어 있다. 속성 관계 유형 및 속성 관계가 가능한 영역, 범위(Domain, Range) 개념은 제한되어 있으며, 하위 개념은 해당 개념이 속한 상위 개념의 제한 범위를 따라간다. Table 10은 일부 속성 관계 유형 및 속성 관계가 가능한 영역과 범위 개념의 사례이다.

Table 10. Examples of Domain-Range Superclasses by Relation Type

Domain	Relation type	Range
Health Problems and Conditions	hasAssociatedDisease	Disease and Disorders
Diseases and disorders	hasComplication	Disease and Disorders /Health Problems and Conditions
Disease and Disorders /Health Problems and Conditions	hasDrugTherapy	Pharmacological treatment
Disease and Disorders /Health Problems and Conditions	hasEvaluationMethod	Assessment and Evaluation
Disease and Disorders /Health Problems and Conditions	hasRiskFactor	Risk factors
Disease and Disorders /Health Problems and Conditions	hasSelfcare	Self-care
Treatment/ Assessment and Evaluation	hasIndication	Disease and Disorders /Health Problems and Conditions
Diseases and disorders	hasSignandSymptom	Health Problems and Conditions
Disease and Disorders /Health Problems and Conditions/ Self-care	hasSupportSystem	Support system
Treatment	hasTargetDisease	Disease and Disorders
Disease and Disorders /Health Problems and Conditions	hasTreatmentMethod	Treatment

2. 건강정보 온톨로지 평가 결과

1) 평가자 특성

온톨로지 평가는 총 5명의 간호정보학 및 의료정보학 전문가(박사 수료이상)가 시행하였다. 간호정보학 전공자 4명, 의료정보학 전공자 1명은 모두 온톨로지 개발 경험이 있는 전문가이다.

2) 평가 결과

온톨로지 구조 평가 결과로는, 온톨로지 구조의 비일관성을 평가하는 항목 '온톨로지 내에서 클래스 A가 클래스 A의 하위클래스로 설정되어 있는 오류가 있는가? '와 온톨로지의 불완전성을 평가하는 '온톨로지 내

특정 클래스가 부적절한 곳에 위치하고 있는 오류가 있는가?', '온톨로지에서 특정 영역이나 개념이 빠져있는 오류가 있는가?'의 총 세 가지 항목에 대해 모든 평가자로 부터 오류가 없는 것으로 평가받았다. 온톨로지 구조에 대한 서술형 의견에서는 '온톨로지 슈퍼클래스에 따라 상세화 수준이 차이가 있어 추후 개발 시 고려해야할 것임', '새로운 개념이나 지식이 추가되기 용이한 구조로 되어있으며, 사용자의 목적에 따라 온톨로지를 변경/적용하여 사용할 수 있을 것으로 판단됨'이 있었다.

온톨로지 내용 평가 결과는 Table 11과 같이, 온톨로지 내용의 일관성, 완성도, 간결성, 확대 가능성, 민감성에 해당하는 항목들에서 모두 3점 이상을 받았다. 그중 온톨로지 내용의 일관성을 평가하는 '온톨로지 내 개념의 형식적 정의가 실제 세계와 일치하는가?'와 '온톨로지 내 모든 정의와 공리에서 모순이 발견되지 않는가?'는 각각 평균 4.6점과 4점을 받았다. 온톨로지의 완성도를 평가하는 '온톨로지에 포함되어야 하는 개념이 빠짐없이 포함되어있는가?','온톨로지는 모든 영역이 균형있게 개발되었는가?', '온톨로지 내 개념, 관계 이름은 애매모호하지 않고 명확하게 이해되는가?' 항목은 각각 평균 3.8점, 3.8점, 4.6점을 받았다. 온톨로지의 간결성을 평가하는 '온톨로지에 불필요한 개념은 없는가?', '온톨로지에 불필요한 관계는 없는가?', '온톨로지 내 클래스는 중복된 의미를 갖지 않고 명확하게 정의되었는가?' 항목은 각각 평균 4.8, 4.4, 5.0점을 받았다. 온톨로지의 확대 가능성을 평가하는 '향후 온톨로지 내 새로운 개념이나 지식이 추가될 때 이미 온톨로지에 정의된 개념이나 지식이 변경 없이 확장될 수 있겠는가?', '향후 온톨로지가 다른 목적에 따라 확장 또는 세분화하는데 어렵지 않겠는가?' 항목은 각각 평균 4.6점, 4.4점을 받았고 마지막으로 온톨로지의 민감성을 평가하는 '정의에 작은 변화가 생기더라도 온톨로지의 구조가 견고하게 유지되는가?' 항목은 평균 4.0점을 받았다.

온톨로지 내용에 대한 서술형 의견에서는 '온톨로지 규모가 커서 모든 개념에 대해 평가하기에는 어려움이 있으나 상위 수준 개념들이 잘 개발되어 확장 개발이나 세분화가 가능할 것으로 보임'이라는 의견이 있었다.

Table 11. Evaluation Result of Health Information Ontology

Category	Item	Mean Score
일관성 (Consistency)	온톨로지 내 개념의 형식적 정의가 실제 세계와 일치하는가?	4.6
	온톨로지 내 모든 정의와 공리에서 모순이 발견되지 않는가?	4.0
완성도 (Completeness)	온톨로지에 포함되어야 하는 개념이 빠짐없이 포함되어 있는가?	3.8
	온톨로지는 모든 영역이 균형있게 개발되었는가?	3.8
	온톨로지 내 개념, 관계 이름은 애매모호하지 않고 명확하게 이해되는가?	4.6
간결성 (Conciseness)	온톨로지에 불필요한 개념은 없는가?	4.8
	온톨로지에 불필요한 관계는 없는가?	4.4
	온톨로지 내 클래스는 중복된 의미를 갖지 않고 명확하게 정의되었는가?	5.0
확대 가능성 (Expandability)	향후 온톨로지 내 새로운 개념이나 지식이 추가될 때 이미 온톨로지에 정의된 개념이나 지식이 변경없이 확장될 수 있겠는가?	4.6
	향후 온톨로지가 다른 목적에 따라 확장 또는 세분화하는데 어렵지 않겠는가?	4.4
민감성 (Sensitiveness)	정의에 작은 변화가 생기더라도 온톨로지의 구조가 견고하게 유지되는가?	4.0

3) 평가 결과의 분석 및 활용

온톨로지 내용의 일관성을 평가하는 '온톨로지는 모든 공리와 정의에서 모순이 발견되지 않는다' 항목에 대해, 각 슈퍼클래스 내에서 수준이 비슷하게 구성되도록 하고 같은 수준 내에 이질적인 개념이 포함되어 있지 않도록 각 슈퍼클래스별 개념을 트리구조로 재확인하였으며, 검토 과정에 수준이 맞지 않는 개념이 존재하는 경우 다른 수준으로 관계를 변경하였다. 예를 들어, 질병 및 장애(1 수준)-악성 종양(2 수준) 하위 수준 3에 '폐암'과 '직업성 폐암'이 모두 존재하여, '직업성 폐암'을 수준 4로 조정하였다.

또한 기타 평가 의견 중 '온톨로지 슈퍼클래스에 따라 상세화 수준이 차이가 있어 추후 개발 시 고려해야할 것임'에 대해, 먼저 슈퍼클래스 별

로 상세 수준(level)과 포함된 개념 수를 비교하였다. 질병 및 장애 4수준, 건강문제 및 상태 4수준, 진단 및 평가 6수준, 생활습관 관리 5수준, 자가 관리 4수준, 지원체계 3수준, 치료 5수준으로 개념이 개발되어 있었다. 각 슈퍼클래스 별 수준의 차이가 발생하는 것은 사용자들이 검색할 만한 개념들의 수준이 영역별로 다르고 그에 따라 필요한 그룹화 개념이 다르기 때문인 것으로 보인다. 예를 들어 진단 및 평가에 대해서는 어떤 검사에 대해 목적, 방법, 평가 내용, 검체에 대한 분류가 필요하며 이에 따른 중간 그룹화 개념도 필요하므로 슈퍼클래스 별로 수준이 다른 것이 적절하다고 판단되어 별도로 수정 과정을 거치지는 않았다.

제 3 절 건강정보 콘텐츠에 건강정보 온톨로지 개념 연계 및 적합성 평가 결과

1. 건강정보 콘텐츠에 건강정보 온톨로지 개념 연계 결과

건강정보 콘텐츠는 2021년 신규 개발이 완료된 콘텐츠를 대상으로 하여 주제 수준 316개, 목차 단위 콘텐츠 3,241개에 대해 연계한 온톨로지 개념 수는 Table 12와 같다. 건강정보 콘텐츠 316개에 대해서 총 14,740개의 온톨로지 개념이 연계되어 하나의 건강정보 콘텐츠 당 평균 46.6개의 개념이 연계되었다. 목차 단위 텍스트 콘텐츠는 하나당 평균 3.6개, 이미지 콘텐츠는 평균 6.5개, 동영상 콘텐츠는 평균 3.1개의 온톨로지 개념이 각각 연계되었다.

Table 12. Result of Health Information Content and Health Information Ontology Concept Annotation

No. of content	No. of segmented content/ No. of media content		No. of annotated ontology concept
316	Segmented text content	2,136	7,587
	Image content	1,031	6,706
	Video Content	74	227
Total		3,241	14,740

2. 건강정보 콘텐츠에 건강정보 온톨로지 개념 연계 적합성 평가 결과

1) 평가자 특성

건강정보 콘텐츠에 건강정보 온톨로지 개념을 연계한 적합성 평가는 총 5명의 간호정보학 및 의료정보학 전문가(박사 이상)가 시행하였다. 간호정보학 전공자 4명, 의료정보학 전공자 1명은 모두 온톨로지 개발 및 온톨로지 개념의 건강정보 콘텐츠 연계 경험이 있는 전문가이다.

2) 건강정보 콘텐츠에 건강정보 온톨로지 개념 연계 적합성 평가 결과

건강정보 콘텐츠에 건강정보 온톨로지 개념을 연계하는 원칙에 따라 연계가 잘 되었는지 시스템의 evaluation 기능을 활용하여 평가한 결과 총 50개의 검색 키워드가 329개의 목차단위 텍스트 콘텐츠 혹은 미디어 건강정보 콘텐츠에 연계되어 키워드 하나당 평균 연계 목차단위 건강정보 콘텐츠 혹은 미디어 콘텐츠는 6.58개였으며 범위는 1~43개였다. 검색 결과는 주제 단위에서는 함께 묶여서 조회되므로 주제 수준에서 묶으면 키워드 하나 당 1~8개의 건강정보 콘텐츠가 결과로 조회되었다.

평가 결과에 대한 사례는 Table 13과 같다. 검색 시 사용한 검색 키워드와, 해당 키워드의 검색 결과 콘텐츠 명이 제시되어 있으며 미디어의 경우는 미디어 콘텐츠의 제목이 표기되어 있다. 50개의 검색 키워드 중 '양수검사', '무월경' '프로게스테론' 3개 키워드에 대해 각각 하나의 콘텐츠 연계가 부적합하다는 평가를 받았다. 먼저 '양수검사'와 '태아 DNA 선별검사 (NIPT 검사) [검사의 개요]' 콘텐츠의 연계는, 양수검사 키워드의 검색 결과로 다른 종류의 검사 개요가 등장하는 사례가 부적합하다는 평가를 받았다. 또한 '프로게스테론'과 '폐경기 [건강문제의 약물치료]' 콘텐츠의 연계는 프로게스테론이 호르몬치료에서 다른 호르몬과 함께 적용되기는 하나 주 치료 성분이 아니라 부적합한 연계라는 평가를 받았다. 마지막으로 '무월경'과 '유루증 [질병의 증상]' 콘텐츠의 연계는 무월경과 유루증이 동반하여 나타나는 경우가 많기는 하나 유루증의 증상이 무월경인 것은 아니므로 부적절한 연계라는 평가를 받았다.

Table 13. Examples of Health Information Content and Health Information Ontology Concept Annotation

검색 키워드	검색 결과
담낭 절제술	담낭 용종 [질병의 치료] [질병의 비약물 치료] [질병 관련 자주하는 질문]
	담낭염 [질병의 치료] (이미지: 담낭절제수술) [질병의 합병증]

	[질병 관련 자주하는 질문]
경식도 심초음파	심초음파검사 [검사 장비] (이미지: 경흉부 심초음파와 경식도 심초음파) [검사 절차] [검사 관련 자주하는 질문]
	대동맥 박리 [질병의 진단 및 검사]
갑상선 열결절	갑상선 스캔 [검사 결과 해석] (이미지: 냉결절과 열결절)
경막 외 마취	자연분만과정 [건강문제 관련 자주하는 질문] (이미지: 경막외 마취)
	전신마취 [비침습적 치료의 적응증] (이미지: 마취의 종류)
근육 이완제	얼굴떨림(얼굴연축_눈꺼풀연축) [건강문제의 약물 치료]
치실	잇몸병(치주질환) [질병의 자가관리]
금주	알코올성 간질환 [질병의 치료]
	음주 [생환습관의 개요] [생활습관 관련 일반적 실천 방법]
	위험음주! 알려드리겠습니다! [생활습관 실천 방법]
갑상선 호르몬 보충 치료	갑상선기능저하증 [질병의 치료]
구개 편도	편도염 [질병의 원인]
행동 치료	자폐증 [정신건강 관련 질병의 치료]
	아동기 품행장애(품행장애) [질병의 치료]
CES-D	우울감 [정신건강 관련 질병의 자가 평가]
틱 장애	말더듬 [정신건강 관련 질병 관련 자주하는 질문]
	얼굴떨림(얼굴연축_눈꺼풀연축) [건강문제의 원인]
저포드맙 식이요법	과민성장증후군 [질병의 비약물 치료]
	복부 팽만 [건강문제의 비약물 치료]
귀지	외이도염 [질병 관련 자주하는 질문]
산전 검사	정상임신관리(임신의 진단과 관리) [생환습관의 평가 및 검사]
	산전 기형아 검사 [검사 목적]
	양수 검사 [검사 적응증 및 금기증]
프로게스테론	자궁내막증 [질병의 약물 치료]
	자궁외 임신 [질병의 진단 및 검사]
	~~폐경기 [건강문제의 약물 치료]~~
언어 치료	자폐증 [정신건강 관련 질병의 치료] [정신건강 관련 질병의 비약물 치료]
	말더듬 [정신건강 관련 질병의 비약물 치료]
	뇌졸중 환자의 재활 [비침습적 치료 방법] (이미지: 뇌졸중 재활치료 팀: 물리치료, 작업치료, 언어치료)
홀터 검사	부정맥 [질병의 진단 및 검사]
	심전도검사 [검사 장비] (활동중 심전도: 홀터기록)

3) 건강정보 콘텐츠와 건강정보 온톨로지 개념 연계 평가 결과의 분석 및 활용

평가를 통해 지적받은 부적합한 연계 결과에 대해서는 재검토를 거쳐 검색 키워드 연계를 삭제하였다.

검색 결과 사례를 살펴보면, 하나의 검색어를 통해 다양한 콘텐츠에 등장하는 목차별 콘텐츠 및 이미지 콘텐츠를 조회할 수 있다. 예를 들어, ‘경막 외 마취’ 검색 키워드를 통해 ‘자연 분만 과정’ 과 ‘전신 마취’에서 해당 관련 목차 콘텐츠를 조회할 수 있고, ‘행동 치료’ 검색 키워드를 통해 ‘자폐증’, ‘아동기 품행장애’에서 관련 목차 콘텐츠를 조회할 수 있어 의미론적 검색이 가능하다. 이는 단순한 콘텐츠명, 목차명 또는 이미지명을 바탕으로한 검색이 아니라 검색 키워드의 적절한 연계(annotation)을 통해 구현된 결과라고 볼 수 있다.

제 5 장 논의

제 1 절 건강정보 온톨로지 개발 및 건강정보 콘텐츠 연계 시스템

건강정보 온톨로지 개발 및 건강정보 콘텐츠 연계 시스템의 의의는 기존에 개발된 온톨로지 개발 및 관리 시스템의 기능들을 포함하면서 동시에 온톨로지 개발 목적에 맞게 건강정보 콘텐츠에 온톨로지 개념을 연계하는 기능을 포함하도록 필수 기능을 정의하여 개발했다는 점에 있다.

이전의 온톨로지 개발 및 관리 시스템은 초점을 온톨로지의 개념 및 관계, 동의어를 생성 및 수정하고 시각화하는 데 맞춘다(Stanford University, 2022; McGuinnes, 2000; Bechhofer, 2001; Domingue, 1998). 그러한 기존 시스템은 온톨로지의 다양한 활용 중 하나가 데이터 검색, 즉 온톨로지 기반의 검색어로 인덱싱된 데이터 간 의미론적 유사성을 기반으로 한 데이터 검색(Asim, 2019)이라는 점에 비추어 볼 때 한계가 있다. 이에 비해 본 연구는 시스템을 활용하여 온톨로지를 개발하는 범위에서 더 나아가 콘텐츠에 온톨로지 개념을 연계하여 해당 결과물을 온톨로지 기반의 의미론적 정보검색에 활용할 수 있도록 개발하였다.

또한 기존의 온톨로지 개발 과정이 수작업으로 진행되어 시간이 많이 소요되는 점을 개선하기 위해 텍스트 콘텐츠의 자연어처리 기능을 구현해 온톨로지 개발에 활용할 수 있도록 했다. 여기에는 자연어처리 기반 온톨로지 개발 과정을 시스템으로 개발한 온톨로지 자동 구축 연구들(심준 & 이홍철, 2009; Ruiz-Martínez et al., 2011; Cahyani, & Wasito, 2017)을 참고하였다. 또한 이미지 콘텐츠의 광학 문자 인식 기능까지 구현해 해당 용어도 자연어처리 결과와 마찬가지로 온톨로지 개발에 활용할 수 있도록 하였다. 기존의 건강정보 온톨로지에 포함되어있는 용어인지 아닌지, 의학용어사전에 포함되어있는 용어인지 아닌지 여부에 따라 우선순위를 표시해 자연어처리 결과가 온톨로지 개발에 더 의미 있게 활

용될 수 있도록 구현하였다.

다만 자연어처리 및 광학문자인식 성능은 꾸준히 개선되어야 할 부분으로 보인다. 자연어처리를 통한 자동 키워드 연계에 대한 이전 연구(Takata, 2014)에서는 단순 용어 추출이 아닌 다양한 자연어처리 학습 모델을 이용하여 정보검색 결과를 비교 분석하였으며 앞으로 가중치 조절을 통한 검색 결과의 최적화가 이루어져야 한다고 제언하였다. 광학문자인식에 대해서도 현재 이미지 품질 등의 문제로 오류가 발생하는 부분을 고도화할 필요가 있다.

제 2 절 건강정보 온톨로지

본 연구의 온톨로지 개발 목적은 국가건강정보포털의 콘텐츠 연계를 통한 검색기능 개선이다. 따라서 건강정보 온톨로지가 건강정보 영역 내 정확성을 확보하면서 동시에 개념의 동의어로 사용자 용어를 포함할 수 있도록 개발하였다.

개발 과정에 온톨로지 개념이 의미론적으로 오류가 없도록 하기 위해 다양한 보건의료 온톨로지를 참고하였으며, 그중 SNOMED CT 계층구조 및 MeSH를 주로 차용하였다. 특히 콘텐츠에서 추출된 각각의 개념이 국가건강정보포털의 소비자들이 검색할 만한 용어로 잘 그룹화 될 수 있도록 상위 수준 개념과 가장 하위 수준 개념 사이에 적절한 중간 개념을 생성할 때는, 참고한 온톨로지에서 유사 개념의 주변 개념을 살펴 필요한 개념을 건강정보 온톨로지에 추가하였다. 이 때 SNOMED CT의 개념 정의를 통해 건강정보 온톨로지의 개념명 및 개념 간 관계를 재 정의하기도 하였다.

본 연구에서 개발된 건강정보 온톨로지를 유사한 목적으로 개발된 MedlinePlus 온톨로지(BioPortal, 2022)와 비교하면 온톨로지의 규모 및 온톨로지의 상세성 수준에서 차이를 발견할 수 있다. 본 연구의 건강정보 온톨로지의 개념은 총 3,823개이고 MedlinePlus 온톨로지의 2022년 버전 개념은 총 2,254개로, MedlinePlus 사이트가 제공하는 건강정보가

1천여 건으로 국가 건강정보 포털의 700여 건에 비해 많다는 여건을 감안하면 본 연구의 건강정보 온톨로지 규모가 콘텐츠 수에 비해 큰 것으로 나타났다. 이렇게 된 요인은 앞서 언급한 중간 개념의 생성이라고 추정된다. MedlinePlus 온톨로지는 최대 수준이 2 수준에 그치는 반면 본 연구에서는 중간 개념 검색을 고려해 계층 구조를 상세히 개발했기 때문에 건강정보 온톨로지가 최대 8 수준으로 개발되었다.

또한 건강정보 온톨로지는 사용자의 검색 활용성을 고려해야 하므로 동의어를 최대한 많이 수집하였다. 참여형 온라인 사전인 우리말샘을 참고했으며 그 외에도 온라인 커뮤니티 등에서 사용되는 용어들을 검색해 추가하였다. 그 결과 개념 하나당 평균 3.5개의 동의어를 생성하였다. 의학용어의 경우에는 최대한 이해가 쉬운 용어로 변환하여 동의어로 추가하는 과정을 거쳐 국가건강정보포털의 분류별 검색 및 화면에 적용할 수 있도록 하였다. 예를 들어, 신체 계통 분류에서는 '근 골격계 질환'은 '뼈 근육 질환', '이비인후과 질환'은 '귀 코 목 질환'으로 한글 대표용어를 개선하였다.

이러한 온톨로지 구조 및 사용자 용어를 통해 국가건강정보포털 검색 시 건강정보 콘텐츠에 대한 접근성이 강화될 것으로 기대된다.

제 3 절 건강정보 콘텐츠에 건강정보 온톨로지 개념 연계

본 연구에서 건강정보 콘텐츠에 건강정보 온톨로지를 연계하는 과정의 의의는 의미론적 연계를 수작업에 그치지 않고 건강정보 온톨로지 개발 및 건강정보 콘텐츠 연계 시스템을 통해 자동으로 수행한 것에 있다. 특히 미디어 콘텐츠 검색은 용어 기반으로 할 수 없기 때문에 개념을 연계하는 방식이 더 중요했다. 이 때 시스템의 광학문자인식 기능은 연구 수행의 효율성 측면에서 매우 유용하였다. 또한 건강정보 콘텐츠의 표준화된 구조를 이용한 자동 키워드 생성 기능 역시, 연구자가 일일이 연계하지 않아도 되도록 해주는 편리한 기능으로 구현되었다.

건강정보 콘텐츠에 건강정보 온톨로지를 연계한 결과에서는 용어 기반이 아닌 의미 기반의 검색을 통해 여러 콘텐츠에 산재된 관련 콘텐츠를 검색 결과로 얻을 수 있음을 확인하였다. 이 방식은 텍스트 콘텐츠뿐 아니라 미디어 콘텐츠에도 적용되어, 이전에 이미지나 동영상이 제목으로만 검색되던 단계를 크게 개선하였다.

건강정보 콘텐츠에 건강정보 온톨로지를 연계한 결과에 대한 평가는 어떤 결과가 정답이라는 기준(golden standard)이 없다는 점이 한계가 될 수 있다. 의미론적 인덱싱을 통한 온톨로지 기반 검색 시스템을 개발한 이전 연구(Kara, 2012)에 대해서는 기존의 검색 시스템 대비 검색 정확도를 비교한 사례가 있다. 그러나 이는 검색 결과로 나타나는 값에 정답이 있는 경우였다.

본 연구에서 적합성 평가 대상 키워드 중 3개의 키워드는 검색 결과로 나온 건강정보 콘텐츠가 부적합하다는 전문가 의견을 받았는데, 본 연구의 경우 이러한 적합 및 부적합에 대한 의견은 사용자에 따라 달라질 수 있다. 왜냐하면 온톨로지 기반 건강정보 콘텐츠 검색 기능의 장점들에 대한 의견을 제시할 때, 사람에 따라 가중치가 다르기 때문이다. 장점 중에는 텍스트로 드러나지 않는 온톨로지 개념 및 동의어의 연계를 통해 검색되지 않던 건강정보 콘텐츠를 조회할 수 있다는 것이 있다. 또는 수많은 텍스트 중 연계된 개념을 통한 검색 결과가 더 우선적으로 나타나게끔 한다는 것도 있다. 연관성이 조금이라도 있는 콘텐츠가 많이 나오는 결과와 검색 연관성이 높은 콘텐츠가 적게 나오는 결과 중 어느 쪽을 선호하는지는 사용자에 따라 다를 수 있는 것이다.

이에 대해서는 사용자가 검색행위를 한 후 해당 검색 결과에서 어떤 경로를 통해 이동하고 어떤 페이지에서 머물렀는지 등의 사용 기록이나 사용자 만족도 등을 통해서 고도화해나갈 필요가 있다. 이를 위해서는 개념 연계 결과에 대한 산출물이 검색엔진에 적용하는 추가 연구도 필요하다.

제 6 장 결론 및 제언

본 연구는 국가건강정보포털의 건강정보 콘텐츠 검색 기능 개선을 위해 온톨로지 기반의 검색을 제공하는 시스템을 개발하고 평가하였다. 본 연구에서 개발한 건강정보 온톨로지 개발 및 건강정보 콘텐츠 연계 시스템의 의의는 기존 온톨로지 개발 및 관리 프로그램이 온톨로지에 국한되고 온톨로지가 활용될 수 있는 정보와 직접 연계하지 못한다는 점을 보완하여 건강정보 콘텐츠와 자동으로 연계할 수 있도록 개발하였다는 것에 있다. 또한 온톨로지 개발 과정에 자연어처리 및 광학문자인식 기능을 구현하여 온톨로지 개발이 효율적으로 이루어지도록 구현하였다. 개발된 시스템은 온톨로지 개발 및 온톨로지와 건강정보 콘텐츠 연계에 대한 사용성 평가에서 전문가들에게 높은 점수를 받았다.

본 시스템을 다른 영역 및 시스템에도 잘 활용하기 위해서는, 온톨로지와 지식베이스를 연계하는 다른 시스템에도 모듈과 같이 적용할 수 있도록 개발 내용을 공개하고 관리하며 발전시켜 나갈 필요가 있다. 본 연구는 건강정보 온톨로지를 개발하고 이를 건강정보 콘텐츠와 연계하였으나 아직 질병관리청의 건강정보 CMS (Content Management System) 내에서 작동하지 않는다. 현재 운영은 개발된 온톨로지 및 연계 결과를 export하여 다시 건강정보포털 운영 서버에 입력하는 방식으로 이루어지고 있다. 이러한 작업의 비효율성을 개선하기 위해서는 본 연구의 건강정보 온톨로지 개발 및 건강정보 콘텐츠 연계 시스템이 CMS 내에서 동작하도록 하는 고도화가 필요하다.

해당 시스템을 통해 개발된 건강정보 온톨로지 역시 구조와 내용 두 측면에서 모두 긍정적인 평가를 받았다. 건강정보 콘텐츠 연계의 적합성은 건강정보 온톨로지와 건강정보 콘텐츠 연계 원칙에 따라 이루어졌으며 전문가들에게 평가받은 결과에 따라 일부 결과를 개선할 수 있었다. 그러나 본 연구에서는 온톨로지와 건강정보 콘텐츠 연계에 대한 평가가 전문가 평가에 그쳐, 앞으로 실제 사용자들을 대상으로 한 사용성 및 검색 개선 평가가 이루어져야 할 필요가 있다.

또한 현재 건강정보 콘텐츠에 연계되어 있는 건강정보 온톨로지 개념과 연계 결과의 활용 범위를 확대하려면, 온톨로지 개념과 보건의료정보 온톨로지인 SNOMED CT 개념을 연계할 수 있다. 이를 통해 추후 PHR(Patient Health Record)에 입력되는 환자 정보에 따라 맞춤형 건강정보 콘텐츠를 제공할 수 있을 것으로 기대된다. 예를 들어, PHR에 환자가 진단받은 진단명이나 환자의 증상이나 복용 약물, 치료 및 검사에 대해 SNOMED CT 개념을 가지고 있으면 건강정보 콘텐츠에 있는 연계 개념과 연결하여 환자 상황에 맞는 데이터 기반의 맞춤형 교육자료 등 적절한 건강정보를 제공할 수 있게 될 것이다.

더불어 검색 기능의 확대를 위해서는 이와 같은 온톨로지 기반의 검색을 제공하면서 온톨로지 개념에 대한 사용자 용어의 수집을 확대하는 작업이 필요하다. 본 연구에서는 사용자 용어의 수집을 연구자가 직접 하였는데, 이를 자동화하여 소셜미디어 등에서 더 많은 실제 사용자 용어를 수집할 수 있는 방안을 마련할 필요가 있다.

앞으로는 온톨로지 기반 검색에 더불어 사용자의 검색 행태 기록 등을 활용하여 사용자가 원하는 검색어를 자동 완성으로 제공해주거나 사용자가 찾으리라고 예측되는 콘텐츠를 상위에 노출해 주는 등의 방법으로 보다 개인화된 건강정보 콘텐츠를 제공할 수 있을 것이다. 구글에서는 사용자가 입력한 검색어 및 검색 결과 중 선택한 페이지, 해당 페이지에 머문 시간 등 사용 기록을 바탕으로 맞춤형 정보를 제공할 뿐 아니라 해당 정보를 바탕으로 다양한 추천을 제시하기도 한다. 이러한 방법을 적절한 개인 정보 보안 방침 하에 적용하여 환자의 생활습관 관리, 질병 예방, 치료 및 치료 후 관리에 이르는 과정에 맞추어 개인화된 건강정보 콘텐츠를 제공할 필요가 있다.

참 고 문 헌

Asim, M. N., Wasim, M., Khan, M. U. G., Mahmood, N., & Mahmood, W. (2019). The use of ontology in retrieval: a study on textual, multilingual, and multimedia retrieval. *IEEE Access, 7*, 21662-21686.

Bechhofer, S., Horrocks, I., Goble, C., & Stevens, R. (2001, September). OilEd: a reason-able ontology editor for the semantic web. In Annual Conference on Artificial Intelligence(pp. 396-408). Springer, Berlin, Heidelberg.

Bioportal. (2022). https://bioportal.bioontology.org/ontologies/MEDLINEPLUS

Cahyani, D. E., & Wasito, I. (2017). Automatic ontology construction using text corpora and ontology design patterns (ODPs) in Alzheimer's disease. *Jurnal Ilmu Komputer dan Informasi, 10(2)*, 59-66.

Chin, J. P., Diehl, V. A., & Norman, K. L. (1988, May). Development of an instrument measuring user satisfaction of the human-computer interface. *In Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 213-218).

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS quarterly*, 319-340.

Domingue, J. (1998). TADZEBAO AND WEBONTO: DISCUSSING,

BROWSING, AND EDITING ONTOLOGIES.

Dung, T. Q., & Kameyama, W. (2007, 5-9 March 2007). A Proposal of Ontology-based Health Care Information Extraction System: VnHIES. *Paper presented at the 2007 IEEE International Conference on Research, Innovation and Vision for the Future*.

Gomez-Perez, A. (2004). Ontology evaluation. *In Handbook on ontologies* (pp. 251-273). Springer, Berlin, Heidelberg.

Gruber, T. (2018). *Ontology*.

Jung, H., Park, H.-A., & Song, T.-M. (2016). Development and Evaluation of an Adolescents' Depression Ontology for Analyzing Social Data. *Studies in health technology and informatics, 225*, 442-446.

Jung, H., Park, H.-A., & Song, T.-M. (2017). Ontology-based approach to social data sentiment analysis: detection of adolescent depression signals. *Journal of medical Internet research, 19(7)*, e259.

Kara, S., Alan, Ö., Sabuncu, O., Akpınar, S., Cicekli, N. K., & Alpaslan, F. N. (2012). An ontology-based retrieval system using semantic indexing. *Information Systems,37*(4), 294-305.

Kim, A. R., Park, H., & Song, T.-M. (2017). Development and Evaluation of an Obesity Ontology for Social Big Data Analysis. *Healthcare informatics research, 23(3)*, 159-168.

Kim, H., Mentzer, J., & Taira, R. (2019). Developing a physical

activity ontology to support the interoperability of physical activity data. *Journal of medical Internet research, 21(4)*, e12776.

Kim, H. Y., Park, H. A., Min, Y. H., & Jeon, E. (2013). Development of an obesity management ontology based on the nursing process for the mobile-device domain. Journal of medical Internet research, 15(6), e2512.

Ko, G. W., Kim, S. W., Park, S. J., No, Y. J., & Choi, S. P. (2019). Implementation of ontology-based service by exploiting massive crime investigation records: focusing on intrusion theft. Journal of the *Korean Society for Library and Information Science, 53(1)*, 57-81.

Lewis, J. R. (1992, October). Psychometric evaluation of the post-study system usability questionnaire: The PSSUQ. *In Proceedings of the human factors society annual meeting (Vol. 36*, No. 16, pp. 1259-1260). Sage CA: Los Angeles, CA: Sage Publications.

Lewis, J. R. (2002). Psychometric evaluation of the PSSUQ using data from five years of usability studies. *International Journal of Human-Computer Interaction, 14(3-4)*, 463-488.

McGuinness, D. L., Fikes, R., Rice, J., & Wilder, S. (2000). The chimaera ontology environment. AAAI/IAAI, 2000, 1123-1124.

Mizuno, S., Ogishima, S., Nishigori, H., Jamieson, D. G., Verspoor, K., Tanaka, H., ... & Nakaya, J. (2016). The pre-eclampsia ontology: a disease ontology representing the domain knowledge specific to

pre-eclampsia. *PloS one, 11(10)*, e0162828.

Nielsen, J. (1995). 10 usability heuristics for user interface design. *Nielsen Norman Group, 1(1)*.

Noy, N. F., & McGuinness, D. L. (2001). Ontology development 101: A guide to creating your first ontology. Stanford Knowledge Systems Laboratory Technical Report KSL-01-05 and Stanford Medical Informatics Technical Report SMI-2001-0880.

Mendonça, R., Rosa, A. F., Oliveira, J. L., & Teixeira, A. (2012, June). Towards ontology based health information search in Portuguese—A case study in neurologic diseases. In 7th Iberian Conference on Information Systems and Technologies (CISTI 2012) (pp. 1-4). IEEE.

Ruiz-Martínez, J. M., Valencia-García, R., Fernández-Breis, J. T., García-Sánchez, F., & Martínez-Béjar, R. (2011). Ontology learning from biomedical natural language documents using UMLS. *Expert Systems with Applications, 38(10)*, 12365-12378.

Sayed, A., & Al Muqrishi, A. (2017). IBRI-CASONTO: Ontology-based semantic search engine. *Egyptian Informatics Journal, 18(3)*, 181-192.

Skopeteas, S., Fiedler, I., Hellmuth, S., Schwarz, A., Stoel, R., Fanselow, G., ... & Krifka, M. (2006). Questionnaire on information structure (QUIS): reference manual (Vol. 4). Universitätsverlag Potsdam.

Stanford Universuty. (2022). https://protege.stanford.edu

Takada, T., Arai, M., & Takagi, T. (2014). Automatic keyword annotation system using newspapers.Journal of *Advanced Computational Intelligence and Intelligent Informatics,18*(3), 340-346.

Tchechmedjiev, A., Abdaoui, A., Emonet, V., Zevio, S., & Jonquet, C. (2018). SIFR annotator: ontology-based semantic annotation of French biomedical text and clinical notes. *BMC bioinformatics, 19(1)*, 1-26.

Yu, B. (2019). Research on information retrieval model based on ontology. *EURASIP Journal on Wireless Communications and Networking, 2019(1)*, 30. doi:10.1186/s13638-019-1354-z

국민 권익 위원회. (2010). https://blog.daum.net/loveacrc/2731

김애란. (2021). 근치적 위절제술을 시행한 위암 환자의 정보제공을 위한 위암 질의응답 챗봇 개발 및 평가 (Doctoral dissertation, 서울대학교 대학원).

김성훈, & 오삼균. (2015). 온톨로지 품질평가를 위한 평가항목 추출에 관한 연구. 정보관리학회지, *32(2),* 193-219.

김학수. (2017). 우리말 자연어 처리 기술: 과거와 현재.

박정은, 김화선, 장민정, & 홍해숙. (2014) 항고혈압제-식이 상호작용 관리를 위한 온톨로지 기반의 임상의사결정지원시스템 구현. 대한간호학회지, 44(3), 294-304.

보건복지부. (2016). 소비자 건강정보 제공을 위한 중장기 전략수립에 관한 연구.

보건복지부. (2018). 건강정보 개발 및 국가건강정보포털 리뉴얼.

송태민, 이연희, 이기호, 진달래, 박현애, 변정용. (2015). 소셜 빅데이터 기반 저출산 정책 수요 예측(연구보고서 2015-21-11). 세종특별자치시:한국보건사회연구원.

송태민, 이기호, 진달래, 천미경, 서동철, 박현애. (2016). 2016년 소셜 빅데이터 기반 보건복지 이슈 동향 분석(정책자료 2016-02). 세종특별자치시:한국보건사회연구원.

심준, 이홍철. (2009). 검색 키워드 확장을 이용한 온톨로지 자동 생성 시스템 개발. 한국산학기술학회논문지, *10(6)*, 1220-1228.

오영삼, & 조영은. (2019). 온라인 건강정보 활용의 한계와 발전방향 모색: 무지의 틀을 이용한 전문가 지식 분석을 중심으로. 보건사회연구, *39(2)*, 358-393.

임수연, 박성배, & 이상조. (2005). 의미관계 정보를 이용한 약품 온톨로지의 구축과 활용. 정보과학회논문지*:* 소프트웨어 및 응용, *32(5)*, 428-437.

장인호. (2011). 온톨로지 기반 법률 검색시스템의구축 및 평가에 관한 연구. 한국문헌정보학회지, 45(2), 345-366.

질병관리청. (2021). 건강정보 개발 및 국가건강정보포털 운영 최종 보고서.

허원실. (2008). 시스템 분석과 설계*:*효과적인 비즈니스 정보시스템 개발 (3rd ed.). 서울. 한빛미디어.

건강정보 온톨로지 개발 및 건강정보 콘텐츠 연계 시스템 전문가 평가지

연구 제목: 건강정보 온톨로지 기반 국가 건강정보 콘텐츠 연계 시스템의 개발 및 평가

본 연구에서는 국가 건강정보 포털의 온톨로지 기반 검색을 지원하기 위해, 건강정보 온톨로지 개념과 건강정보 콘텐츠를 연계하고자 합니다. 이를 위해 온톨로지를 개발 및 관리하고, 국가 건강정보 콘텐츠와 연계할 수 있는 "건강정보 온톨로지 개발 및 건강정보 콘텐츠 연계" 시스템을 개발하였습니다. 함께 제공되는 시스템 링크를 사용해 시스템을 사용한 뒤, 휴리스틱 평가 및 PSSUQ 도구를 사용해 평가해 주십시오.

연구 참여의 중단을 원하시면 언제든지 참여를 중단하실 수 있습니다.

본 평가와 관련한 문의사항은 010-6425-4215 (강한나, 간호 정보학 박사과정)로 주시기 바랍니다. 감사합니다.

1. 휴리스틱 평가

시스템을 사용 후 다음 휴리스틱 평가 항목을 심각성 수준 척도에 따라 평가해주시기 바랍니다.

*심각성 수준 척도

0: 사용성 문제 없음

1: 사용성 문제가 있으나 반드시 수정할 필요는 없음

2: 사용성 문제가 있으며 수정이 필요하지만 우선순위가 상대적으로 낮음

3: 사용성 문제가 있으며 수정이 필요하고 우선순위가 상대적으로 높음

4: 사용성 문제가 있으며 반드시 수정되어야 함

No	Item	Score
1	시스템 상태의 시각화	
2	시스템과 실 세계의 일치	
3	사용자 제어 및 자율성	
4	일관성 및 표준	
5	오류 예방	
6	기억보다는 인식	
7	사용의 유연성 및 효율성	
8	공학적으로 간소화된 디자인	
9	오류 인식, 진단 및 복구 지원	
10	도움말 및 설명서	

2. PSSUQ (Post-Study System Usability Questionnaire)

시스템을 사용 후 시스템 유용성에 대해 다음과 같은 항목에 대해 5점 리커트 척도로 평가해주시기 바랍니다.

(1점: 매우 그렇지 않다 2점: 약간 그렇지 않다 3점: 보통이다 4점: 약간 그렇다 5점: 매우 그렇다)

영역	No	Item	Score
시스템 유용성	1	전반적으로 나는 시스템의 쉬운 사용방법에 만족한다.	
	2	이 시스템은 사용하기에 간단하다.	
	3	나는 이 시스템으로 과업을 빨리 끝마칠 수 있다.	
	4	나는 이 시스템을 사용하는 것이 편하다.	
	5	이 시스템은 배우기가 쉽다.	
	6	나는 이 시스템을 이용해서 과업을 잘 할 수 있다고 생각한다.	
정보의 질	7	에러 메시지는 문제를 해결하는 법을 분명히 알려준다.	
	8	내가 시스템 이용 시 실수를 범하면, 언제든 빠르고 쉽게 해결할 수 있다.	
	9	이 시스템이 제공하는 정보들이 확실하다.	
	10	내가 필요로 하는 정보를 찾는 것이 쉽다.	
	11	제공하는 정보가 과업을 마치는데 효과적으로 도움을 준다.	
	12	시스템 화면에 대한 정보의 나열은 분명하다.	
인터페이스 질	13	이 시스템의 인터페이스는 흥미로웠다.	
	14	나는 이 시스템의 인터페이스를 쓰는 것이 좋다.	
	15	이 시스템은 내가 기대한 기능들을 다 갖추고 있다.	
총평	16	전체적으로 나는 이 시스템에 대해 만족스럽다.	

3. 시스템 전반에 대한 조언

시스템 전반에 대한 조언을 자유롭게 기술해주시기 바랍니다.

온톨로지 전문가 평가지

연구 제목: 건강정보 온톨로지 개발 및 국가 건강정보 콘텐츠 연계 시스템의 개발 및 평가

본 연구에서는 국가건강정보포털에서 제공되는 건강정보의 검색 기능을 향상하고 소비자들의 정보 활용성 증가를 위해 건강정보 온톨로지를 개발하였습니다. 온톨로지 개발과 건강정보 콘텐츠 연계에 경험이 있는 전문가를 대상으로 온톨로지의 정확성과 타당성을 평가하고자 합니다.

먼저 온톨로지 평가는 Gomez-Perez(2010)의 연구를 참고한 두가지 주제로 구성되어 있습니다. 먼저 온톨로지의 개념 분류 체계(taxonomy)의 비 일관성과 불완전성에 대해 오류 있음/ 없음으로 평가하고, 다음으로 온톨로지 내용의 일관성, 완성도, 간결성, 확대 가능성, 민감성에 대해 5점 리커트 척도로 평가해주십시오. 마지막으로 온톨로지 개발 전반에 대한 조언을 자유롭게 기술해주시면 됩니다.

다음으로 건강정보 콘텐츠와 온톨로지 개념 연계 평가는 제공된 키워드로 검색 시 조회되는 연계 콘텐츠가 적절한지, 콘텐츠 별로 연계된 키워드가 적절한지 두가지로 평가해주시면 됩니다.

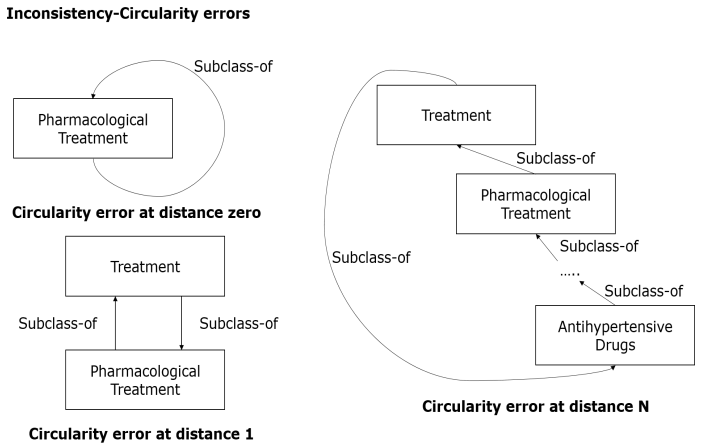
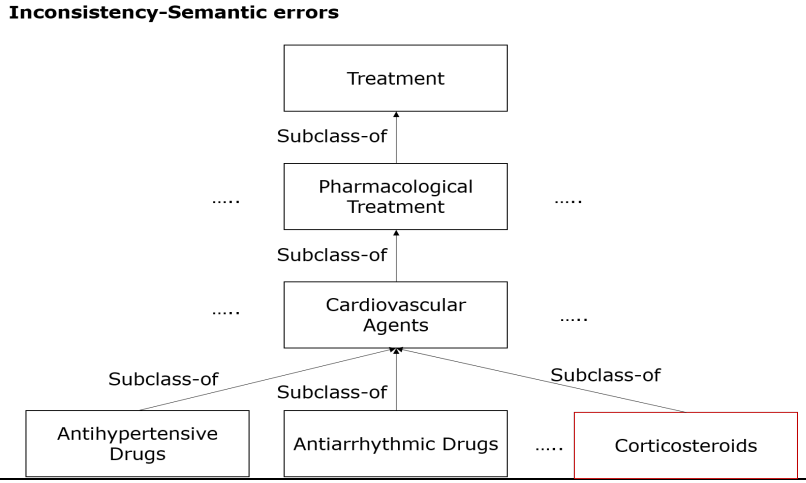
함께 제공되는 건강정보 온톨로지의 접속 링크를 참고해 주십시오.

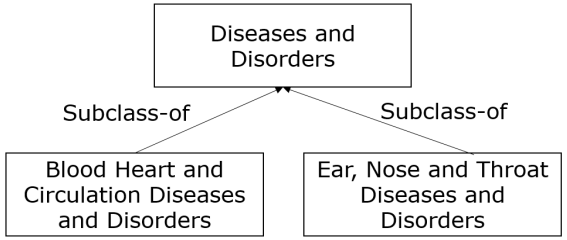
본 평가와 관련한 문의사항은 010-6425-4215 (강한나, 간호 정보학 박사과정)로 주시기 바랍니다. 감사합니다.

〈온톨로지 평가〉

1. Taxonomy evaluation

온톨로지의 개념 분류 체계(taxonomy)의 비 일관성과 불완전성에 대해 오류 있음/ 없음으로 평가해 주십시오.

분류 1	분류2	분류3	평가 문항	예시	오류여부
Taxonomy evaluation	비일관성 Inconsistency	Circulatory errors	온톨로지 내에서 클래스 A가 클래스 A의 하위클래스로 설정되어있는 오류가 있는가?		□있음 □없음
	불완전성 Incompleteness	Semantic errors	온톨로지 내 특정 클래스가 부적절한 곳에 위치하고 있는 오류가 있는가?		□있음 □없음

		Incomplete concept classification	온톨로지에서 특정 영역이나 개념이 빠져있는 오류가 있는가?	**Incompleteness-Incomplete concept classification**  Diseases and Disorders의 하위 클래스로 Blood Heart and Circulation Diseases and Disorders와 Ear, Nose and Throat Diseases and Disorders만 있는 경우	□있음 □없음

2. Content Evaluation

온톨로지 내용의 일관성, 완성도, 간결성, 확대 가능성, 민감성에 대해 5점 리커트 척도로 평가해주십시오.

(1점: 매우 그렇지 않다 2점: 약간 그렇지 않다 3점: 보통이다 4점: 약간 그렇다 5점: 매우 그렇다)

분류 1	분류2	평가 문항	점수
Content evaluation	일관성 (Consistency)	온톨로지 내 개념의 형식적 정의가 실제 세계와 일치하는가?	□1점 □2점 □3점 □4점 □5점
		온톨로지 내 모든 정의와 공리에서 모순이 발견되지 않는가?	□1점 □2점 □3점 □4점 □5점
	완성도 (Completeness)	온톨로지에 포함되어야 하는 개념이 빠짐없이 포함되어 있는가?	□1점 □2점 □3점 □4점 □5점
		온톨로지는 모든 영역이 균형있게 개발되었는가?	□1점 □2점 □3점 □4점 □5점
		온톨로지 내 클래스, 관계, 인스턴스 이름은 애매모호하지 않고 명확하게 이해되는가?	□1점 □2점 □3점 □4점 □5점
	간결성 (Conciseness)	온톨로지에 불필요한 개념은 없는가?	□1점 □2점 □3점 □4점 □5점
		온톨로지에 불필요한 관계는 없는가?	□1점 □2점 □3점 □4점 □5점
		온톨로지 내 클래스는 중복된 의미를 갖지 않고 명확하게 정의되었는가?	□1점 □2점 □3점 □4점 □5점

	확대 가능성 (Expandability)	향후 온톨로지 내 새로운 개념이나 지식이 추가될 때 이미 온톨로지에 정의된 개념이나 지식이 변경없이 확장될 수 있겠는가?	□1점 □2점 □3점 □4점 □5점
		향후 온톨로지가 다른 목적에 따라 확장 또는 세분화하는데 어렵지 않겠는가?	□1점 □2점 □3점 □4점 □5점
	민감성 (Sensitiveness)	정의에 작은 변화가 생기더라도 온톨로지의 구조가 견고하게 유지되는가?	□1점 □2점 □3점 □4점 □5점

3. 온톨로지 전반에 대한 조언

<건강정보 콘텐츠와 온톨로지 개념 연계 평가>

다음 각각의 키워드에 대해 조회되는 연계 콘텐츠가 내용상 적절한지 확인하고, 각각의 연계 콘텐츠에 대해 전체 연계된 개념의 수와 내용이 적절한지 평가하여 예 혹은 아니오로 답해주십시오.

검색 키워드	검색 개념에 연계된 콘텐츠가 적절한가	결과로 제시된 콘텐츠에 연계된 전체 개념의 수와 내용이 적절한가
1		
2		
3		
4		
5		
6		
7		
8		
9		
10		

발견된 문제점을 기술해 주십시오.

<전문가 제공 키워드 리스트>

다음 각각의 키워드에 대해 조회되는 연계 콘텐츠가 적합한지 확인하고 적합하지 않은 경우 의견을 작성해 주십시오.

평가자 1	평가자 2	평가자 3	평가자 4	평가자 5
1.황사 2.갑상선의 구조 3.귀두 포피염 4.정관 수술 합병증 5.항문 종양 6.자폐증 7.혈액 수혈 8.담낭 절제술 9.부종 10 MMR 백신	1.갑상선 열결절 2.경막 외 마취 3.근육 이완제 4.치실 5.급성 충수염 6.경식도 심초음파 7.욕창 사정 도구 8.DASH 식이 9.저체온증 10.후두	1.갑상선 호르몬 보충 치료 2.구개 편도 3.급성 염좌 4.항문 통증 5.급성 간 부전 6.행동 치료 7.CES-D 8.다이옥신 9.서혜부 탈장 10.부비동 x-ray	1.말 더듬 2.틱 장애 3.복부 팽만 4.저포드맙 식이요법 5.귀지 6.심전도 검사 7.양수 검사 8.산전 검사 9.프로게스테론 10.이명	1.언어 치료 2.당불내성 3.유산균 4.외이도염 5.급성 심근경색 6.홀터 검사 7.융모양막염 8.무월경 9.금주 10.메니에르병

심의결과 통보서

수신

연구책임자	**이름:** 박현애	**소속:** 간호대학 간호학과	**직위:** 교수
지원기관	정부기관(질병관리청)		

과제정보

승인번호	IRB No. 2207/003-002		
연구과제명	건강정보 온톨로지 기반 국가 건강정보 콘텐츠 연계 시스템의 개발 및 평가		
연구종류	학술연구 , 설문조사		
심의종류	신규		
심의일자	2022-07-14		
심의대상	설명문 및 동의서 또는 서면동의 면제사유서 , 연구결과정리양식 , 생명윤리준수서약서 , 모집문건 , 설문지(면담질문지) , 연구비 수주용 연구계획서 , 연구책임자 경력사항		
심의결과	승인		
승인일자	2022-07-14	**승인유효기간**	2023-07-13
정기보고주기	12개월		
심의의견	1. 심의결과 제출하신 연구계획에 대해 승인합니다. 2. 연구자께서는 승인된 문서를 사용하여 연구를 진행하시기 바라며, 만일 연구진행 과정에서 계획상에 변경사항(연구자 변경, 연구내용 변경 등)이 발생할 경우 본 위원회에 변경신청을 하여 승인 받은 후 연구를 진행하여 주십시오. 3. 유효기간 내 연구가 끝났을 경우 종료 보고서를 제출하여야 하며, 승인유효기간 이후에도 연구를 계속하고자 할 경우, 2023-06-13까지 지속심의를 받도록 하여 주십시오.		
검토의견	**계획서 검토의견**		
	동의서 검토의견	대상자가 전문가들로 구성되어 있고 연구 참여에 따른 위험이 낮긴 하나 참여의 자발성은 반드시 확보해 주시기 바랍니다.	
	기타 검토의견		

2022년 07월 14일

서울대학교 생명윤리위원회 위원장

본 위원회가 승인한 연구를 수행하는 연구자들은 다음의 사항을 준수해야 합니다.

1. 반드시 계획서에 따라 연구를 수행해야 합니다.
2. 위원회의 승인을 받은 연구참여자 동의서를 사용해야 합니다.
3. 모국어가 한국어가 아닌 연구참여자에게는 승인된 동의서를 연구참여자의 모국어로 번역하여 사용해야 하며 번역본은 인증 및 위원회의 승인을 거쳐야 합니다.
4. 연구참여자 보호를 위해 불가피한 경우를 제외하고는 연구 진행중의 변경에 대해서는 위원회의 사전 승인을 받아야 합니다. 연구참여자의 보호를 위해 취해진 응급상황에서의 변경에 대해서는 즉각 위원회에 보고해야 합니다.
5. 위원회에서 승인 받은 계획서에 따라 등록된 연구참여자의 사망, 입원, 심각한 질병에 대하여는 위원회에 서면으로 보고해야 합니다.
6. 임상시험 또는 연구참여자의 안전에 대해 유해한 영향을 미칠 수 있는 새로운 정보는 즉각 위원회에 보고해야 합니다.
7. 위원회의 요구가 있을 때에는 연구의 진행과 관련된 사항에 관하여 위원회에 보고해야 합니다.
8. 연구참여자 모집광고는 사용 전에 위원회로부터 승인을 받아야 합니다.
9. 강제 혹은 부당한 영향력이 없는 상태에서 충분한 설명에 근거하여 연구참여자로부터 동의를 받아야 하며, 잠재적인 연구참여자에 대해서 연구 참여 여부를 숙려할 수 있도록 충분한 기회를 제공해야 합니다.

심의결과 통보서

수신

연구책임자	**이름:** 박현애	**소속:** 간호대학 간호학과	**직위:** 교수
지원기관	정부기관(질병관리청)		

과제정보

승인번호	IRB No. 2207/003-002		
연구과제명	건강정보 온톨로지 기반 국가 건강정보 콘텐츠 연계 시스템의 개발 및 평가		
연구종류	학술연구 , 설문조사		
심의종류	변경		
심의일자	2022-11-07		
심의대상	설명문 및 동의서 또는 서면동의 면제사유서 , 연구결과정리양식 , 변경대조표 , 모집문건 , 설문지(면담질문지)		
심의결과	승인		
승인일자	2022-11-07	**승인유효기간**	2023-07-13
정기보고주기	12개월		
심의의견	1. 심의결과 제출하신 연구계획에 대해 승인합니다. 2. 연구자께서는 승인된 문서를 사용하여 연구를 진행하시기 바라며, 만일 연구진행 과정에서 계획상에 변경사항(연구자 변경, 연구내용 변경 등)이 발생할 경우 본 위원회에 변경신청을 하여 승인 받은 후 연구를 진행하여 주십시오. 3. 유효기간 내 연구가 끝났을 경우 종료 보고서를 제출하여야 하며, 승인유효기간 이후에도 연구를 계속하고자 할 경우, 2023-06-13까지 지속심의를 받도록 하여 주십시오.		
검토의견	**계획서 검토의견**		
	동의서 검토의견		
	기타 검토의견		

2022년 11월 07일

서울대학교 생명윤리위원회 위원장

본 위원회가 승인한 연구를 수행하는 연구자들은 다음의 사항을 준수해야 합니다.

1. 반드시 계획서에 따라 연구를 수행해야 합니다.
2. 위원회의 승인을 받은 연구참여자 동의서를 사용해야 합니다.
3. 모국어가 한국어가 아닌 연구참여자에게는 승인된 동의서를 연구참여자의 모국어로 번역하여 사용해야 하며 번역본은 인증 및 위원회의 승인을 거쳐야 합니다.
4. 연구참여자 보호를 위해 불가피한 경우를 제외하고는 연구 진행중의 변경에 대해서는 위원회의 사전 승인을 받아야 합니다. 연구참여자의 보호를 위해 취해진 응급상황에서의 변경에 대해서는 즉각 위원회에 보고해야 합니다.
5. 위원회에서 승인 받은 계획서에 따라 등록된 연구참여자의 사망, 입원, 심각한 질병에 대하여는 위원회에 서면으로 보고해야 합니다.
6. 임상시험 또는 연구참여자의 안전에 대해 유해한 영향을 미칠 수 있는 새로운 정보는 즉각 위원회에 보고해야 합니다.
7. 위원회의 요구가 있을 때에는 연구의 진행과 관련된 사항에 관하여 위원회에 보고해야 합니다.
8. 연구참여자 모집광고는 사용 전에 위원회로부터 승인을 받아야 합니다.
9. 강제 혹은 부당한 영향력이 없는 상태에서 충분한 설명에 근거하여 연구참여자로부터 동의를 받아야 하며, 잠재적인 연구참여자에 대해서 연구 참여 여부를 숙려할 수 있도록 충분한 기회를 제공해야 합니다.

Abstract

Development and evaluation of ontology-based annotation system for National Health Information Portal contents

Hannah Kang

College of Nursing

The Graduate School

Seoul National University

The purpose of this study is to develop and evaluate a system that annotates health information ontology to health information contents to improve search of the National Health Information Portal, develop a health information ontology utilizing the system, and then annotate health information ontology concept to health information contents and evaluate the suitability of annotations.

This study was categorized into three stages. The first step was to develop a system that creates health information ontology that connects back to health information content. In the second step, a health information ontology was developed using the system. In the last step, health information ontology concepts were annotated to health information contents.

The health information ontology development and health information content annotation system received heuristic evaluation

and PSSUQ evaluation from five experts. Heuristic evaluation disclosed two items with a severity level of three or higher that must be corrected due to usability problems. As a result of the PSSUQ evaluation, the average score for each item was 3.4 to 5.0 points, with the overall average score of 4.3 points. The health information ontology was developed with 3,823 concepts and 12,082 synonyms.

The health information ontology was evaluated as having no errors in its structure and received an average of 4.4 points in content evaluation. Also, the evaluation confirmed that no structural error was indicated in the health information ontology, resulting in an average of 4.4 points. As a result, 14,740 ontology concepts were annotated to 3,241 contents. In the expert evaluation on annotation suitability, three concepts out of fifty concepts were evaluated as inappropriate.

The significance of the health information ontology development and health information content annotation system developed in this study is to automatically annotate ontology concepts to the health information contents, whereas previous ontology development and management system was limited to ontology itself, which cannot be directly linked to the information contents utilizing available ontology information.

The developed health information ontology received positive evaluations in both structure and content. The suitability of health information content annotation was created based on the principles of health information ontology and health information content annotation. Some results were improved according to the results evaluated by experts.

In the future, integration between the ontology development and health information content annotation system with CMS is vital in

order to apply to the content development process. In addition, it is necessary to automate the collection of user terms so that the ontology-based search function can be strengthened. In addition to providing ontology-based search, generating personalized health information contents by exploiting user's search behavior is essential.

Keywords : ontology, ontology based search, health information content, information system development, consumer health information
Student Number : 2018-34470