# 효율적인 암정보 제공을 위한
# 온톨로지 기반 대화시스템 개발

- 일반인 대상 위암정보제공을 중심으로 -

연세대학교 보건대학원

보건정보관리학과

이 수 경

# 효율적인 암정보 제공을 위한 온톨로지 기반 대화시스템 개발

- 일반인 대상 위암정보제공을 중심으로 -

지도 채 영 문 교수

이 논문을 보건학석사 학위논문으로 제출함

**2005**년 **6**월 일

연세대학교 보건대학원

보건정보관리학과

이 수 경

이수경의 보건학석사 학위논문을 인준함

심사위원_____인

심사위원_____인

심사위원_____인

연세대학교 보건대학원

2005년 6월 일

# 감사의 글

10년만에 다시 시작한 학업으로 기뻐했던 첫 시작이 벌써 그 끝이 되었습니다. 대학원에서의 2년 반이라는 짧고도 긴 시간들은 제 인생에 있어 매우 소중한 기억으로 남겨질 것입니다. 논문을 끝까지 쓸 수 있도록 부족한 저를 도와주신 모든 분들께 진심으로 감사의 말씀을 드리고 싶습니다.

좋은 결실을 맺을 수 있도록 논문의 처음부터 끝까지 최선을 다해 지도해주시고, 포기하지 않도록 이끌어 주셨던 채영문 교수님께 깊은 감사를 드립니다. 또한 바쁘신 중에도 세심하게 지도해주시고 프로그램 평가까지도 직접 해주신 강건욱 선생님과 어려웠던 연구 방법 및 평가에 대해 함께 고민해주시고 격려해 주셨던 박경모 교수님께 감사를 드립니다.

서로 의지하며 함께 논문을 준비했던 동기들, 힘찬 응원과 더불어 설문에 참여해주셨던 선후배 여러분들, 특히 시스템 개발을 위해 애써 주셨던 고명자 선생님과 박찬 선생님에게 고마운 마음을 전합니다.

바쁜 업무 중에도 학업을 무사히 마칠 수 있도록 배려해주신 박현문 전무님과 주연훈 상무님, 박인선 차장님, 이미준 과장님, 이영주 대리님 그리고 함께 일했던 의료실손의 팀원 모든 분들께 감사드립니다.

마지막으로 부족한 저에게 변함없는 지지와 후원을 해준 사랑하는 가족들 모두에게 감사의 마음을 전합니다.

2005년 6월

이 수 경 올림

# 차 례

**국 문 요 약** ····· viii

**I. 서 론** ····· 1

1. 연구의 배경 및 필요성 ····· 1

2. 연구목적 ····· 5

**II. 이론적 배경** ····· 6

1. 암 정보 제공 ····· 6

2. 온톨로지 ····· 8

가. 온톨로지의 정의 ····· 8

나. 온톨로지의 특징과 역할 ····· 9

다. 온톨로지의 종류 ····· 10

라. 온톨로지 구성요소 ····· 13

마. 온톨로지 방법론 ····· 14

3. 온톨로지 기반의 활용 시스템 ····· 17

가. 지식표현과 온톨로지 ····· 17

나. 의료분야에서의 온톨로지 구축사례 ····· 19

다. 대화형 에이전트 ····· 21

4. 온톨로지 시스템 평가 및 활용 ····· 22

가. W3C OWL Test Cases ····· 22

나. 실사용 평가 ····· 23

다. 다면 평가 ····· 25

라. 유용성 평가 ····· 25

마. Recall & Precision Test ····· 26

바. F - measure ····· 28

사. 시스템 활용 ····· 30

**III. 연구방법** ····· 31

1. 연구 대상 및 범위 ····· 31

2. 연구의 틀 ····· 32

3. 연구 분석 및 방법 ····· 33

가. 위암정보 FAQ 자료 수집과 분석 ····· 33

나. Protege 이용 질문 용어에 대한 온톨로지 구축 ····· 33

다. 온톨로지 기반의 암정보 제공 대화시스템 개발 ····· 34

라. 시스템 평가 및 활용 방안 제시 ····· 34

**IV. 연구결과** ····· 35

1. 위암 정보 FAQ 자료 수집과 분석 ····· 35

가. 위암 정보 FAQ 자료 수집 ····· 35

나. 위암 정보 FAQ 자료 분석 ····· 37

다. 자료 분석을 통한 중요 용어 목록 열거 ····· 38

2. Protege 이용 질문 용어에 대한 온톨로지 구축 ····· 39

가. 클래스와 클래스 계층 정의 ····· 39

나. 클래스, 서브클래스, 슬롯, 인스턴스 입력 ····· 44

다. 구축 온톨로지의 XML 전환 ····· 47

3. 온톨로지 기반 암정보 제공 대화시스템 개발 ····· 48

가. 시스템 구성 ····· 48

나. 시스템 흐름도 ····· 49

다. 유즈케이스 다이어그램 ····· 50

라. ERD (Entity Relationship Diagram) ····· 52

마. 시스템 화면 ····· 54

4. 시스템 평가 및 활용방안 제시 ····· 59

가. 시스템 성능 평가 ····· 59

나. 시스템 사용성 평가 ····· 62

다. 시스템 활용방안 제시 ····· 68

**V. 고 찰** ····· 69

**VI. 결 론** ····· 72

참고문헌 ····· 75

부 록 ····· 83

영문초록 ····· 86

# 표 차 례

표 1. 유형별 온톨로지 분류 ····· 12
표 2. 온톨로지 구성요소 ····· 13
표 3. 대화형 에이전트에 사용되는 방법 ····· 21
표 4. W3C Entailment Test 결과 비교 ····· 23
표 5. 재현율과 정확률 공식 ····· 27
표 6. 조사 대상 웹사이트 ····· 35
표 7. 질문과 답변 수집자료 ····· 36
표 8. 질문 키워드 정리 ····· 37
표 9. 중요 용어와 정의 ····· 38
표 10. Class 결정 ····· 39
표 11. 질문 Class 구분 ····· 39
표 12. 테이블 명세서 ····· 53
표 13. Recall과 Precision 값 측정 ····· 59
표 14. Recall, Precision, F-measure 측정시 참고사항 ····· 60
표 15. F-measure 값 측정 ····· 61
표 16. 평가 문항과 측정방법 ····· 62
표 17. 일반적 특성 비교 ····· 63

표 18. 위암 관련 질문 유형 ····· 64

표 19. 검색소요시간 및 검색단계 비교 ····· 65

표 20. 직종별 결과 만족도 비교 ····· 66

표 21. 검색 결과 만족도 비교 ····· 67

# 그 림 차 례

그림 1. 10년전과 주요 사망원인별 사망률 비교 ····· 6

그림 2. ON9 top-level ontology library ····· 11

그림 3. 온톨로지 설계 절차 ····· 16

그림 4. Knowledge Meta Process ····· 18

그림 5. 모호 집단의 F-measure ····· 29

그림 6. 온톨로지 지식표현시스템의 정보 공유와 재사용에서의 역할 ····· 30

그림 7. 연구의 틀 ····· 32

그림 8. 의료환경 모델링 ····· 40

그림 9. 질문상황 클래스 계층 ····· 41

그림 10. 질문주제 클래스 계층 ····· 42

그림 11. 질문유형 클래스 계층 ····· 43

그림 12. 클래스와 서브클래스 입력화면 ····· 44

그림 13. 슬롯 입력 화면 ····· 45
그림 14. 인스턴스 입력 화면 ····· 45
그림 15. 슬롯 계층 ····· 46
그림 16. XML 전환 화면 ····· 47
그림 17. 시스템 구성도 ····· 48
그림 18. 시스템 흐름도 ····· 49
그림 19. Use Case Diagram ····· 50
그림 20. ERD(개체관계도) ····· 52
그림 21. 메인 화면 ····· 54
그림 22. 질문 입력 화면 ····· 55
그림 23. 질문 재입력 화면 ····· 56
그림 24. 답변 출력 화면 ····· 57
그림 25. 시스템 관리자 화면 ····· 58
그림 26. 국가암정보센터 웹사이트에 온톨로지 시스템 추가 화면 ····· 68

# 부 록 차 례

부록 1. 위암정보제공을 위한 온톨로지 기반 대화시스템 평가서 ····· 83

# 국 문 요 약

국내 사망원인 1위인 암에 대한 일반인들의 알고자 하는 요구 및 관리에 대한 중요성이 높아지고 있는 현실 속에서 암관련 정보들이 다양하게 제공되고 있지만 개개인의 요구에 부합되는 맞춤형 정보를 찾기는 어렵다. 따라서 이 연구는 일반인을 대상으로 요구에 맞는 보다 효율적인 위암관련 정보를 제공하기 위한 온톨로지 기반의 대화시스템을 개발하고 평가를 통해 활용방안을 제시하고자 하였다.

이 연구는 암관련 정보 제공 웹사이트의 FAQ자료를 대상으로 하였으며, 2005년 4월 시점에서 국내의 병원, 협회, 업체에서 운영하고 있는 암정보 관련 웹사이트를 조사하였다. 7개 웹사이트에서 일반인을 대상으로 제공되는 위암정보 FAQ 자료를 총 140개 수집하였으며 최종 126개의 질문과 답변으로 정리하였다. 질문에 대해서 키워드를 추출하여 총 286개로 정리하였으며 그 중 95개 용어에 대한 정의를 용어사전에서 발췌하였다.

Protege 2000 Ver 3.0을 이용하여 질문 자료에 대한 온톨로지를 구축하였으며, 해당 질문과 대답에 대한 R-DB와 질문 온톨로지는 대화 에이전트를 통해 서로 연결되어 있는 구조를 가지고 있다. 사용자는 인터페이스를 통해 질문을 입력하여 대화 에이전트에게 답변을 요구하며, 찾은 답변을 보게 된다. 인터페이스는 Visual Studio.NET 2003을 사용하여 프로그래밍하였다. 기존의 대화형 에이전트를 참조하여 인터넷을 기반으로 동작할 수 있게 ASP.NET를 사용하였고, 비하인드 코드로서 C#을 사용하였다. DBMS는 MS-SQL Server를 사용하였으며, 비연결성을 보장하기 위하여

ADO.NET으로 DBMS에 접근하였다. 개발된 CancerQ시스템은 직접 웹에서 구현되었다(URL : http://cancerq.ezzin.com).

온톨로지 시스템의 검색 성능을 평가하는 시스템 성능 평가와 사용성 평가를 시행하였다. 시스템 성능 평가로써 검색시스템의 Recall, Precision, F-measure를 암관련 전문의사인 전문가와 연구자가 **2**회 검색을 실시하여 측정한 결과 기존 웹사이트 검색에 비해 온톨로지 시스템에서의 검색 성능이 더 뛰어났다. 사용성 평가는 일반인과 의료인 총 **36**명을 대상으로 기존 웹사이트와 온톨로지 대화시스템에서 각각 위암관련 정보검색을 실시하여 두 가지 시스템에서의 검색 소요 시간, 검색 단계, 검색 결과 만족도 항목을 비교하였다. 질병에 대한 지식을 기본적으로 가지고 있는 의료인과는 달리 일반인의 경우 온톨로지 시스템에 비해 기존 웹사이트에서 원하는 정보를 찾는데 시간이 더 소요되고, 더 많은 횟수를 검색하였으며, 만족도는 온톨로지 시스템에서 더 높았다. 따라서 온톨로지 시스템의 경우 전문가보다는 일반인을 대상으로 전문 지식을 제공할 때 더 효과적이라고 할 수 있다.

국가암정보센터 웹사이트에서 키워드 검색, 내용 검색, 자료별 검색, 용어 사전 등 다양한 방법들을 통해 암정보를 제공하고 있다. 따라서 이 연구의 제한점이 보완되고 보다 발전된 온톨로지 기반의 대화형 에이전트를 국가암정보센터 웹사이트에 추가하여 활용한다면 일반인들에게 보다 친숙하고 빠르게 암정보를 제공할 수 있을 것이다.

---

핵심되는 말: 온톨로지, 대화시스템, 암정보

# I. 서론

## 1. 연구의 배경 및 필요성

국내 사망 원인 1위인 암은 현대사회에서 살아가는 일반인들의 건강을 위협하는 존재이다. 따라서 암에 대해 알고자 하는 높은 요구 및 암 예방과 관리에 대한 중요성이 높아지고 있는 것이 현실이다. 다양한 매체를 통해 암관련 정보들이 제공되고 있지만 개개인의 요구에 부합하는 맞춤형의 정보를 찾기는 매우 어려운 것 또한 사실이다. 다양한 정보를 쉽게 제공받을 수 있는 웹을 통해서 요구에 맞는 효율적인 정보를 제공하였을 때 이와 같은 노력은 궁극적으로 암에 대한 인식과 행동 변화를 통하여 암 예방 및 관리에 기여하고, 암 발생 및 사망률을 감소시켜 삶의 질 향상에 도움이 될 것이다.

암질환은 질환의 규모와 양상으로 볼 때 '공중의 이익과 편익 그리고 필요성'이라는 사회 기여적 측면을 가지고 있어 공공데이터베이스로 구축・활용될 필요성이 크다. 외국의 경우 연구자 중심의 정보망 및 데이터베이스 구축으로 암 연구 및 정책 결정에 필요한 정보를 제공하는 한편, 이를 가공하여 환자 및 일반인들이 필요로 하는 정보 컨텐츠 제공에 활용하고 있다. 암관련 자료의 정보화 및 정보 공동 활용에 있어 다양한 통신방식과 기술응용으로 개발된 시스템 및 소프트웨어간의 이식성, 확장성, 상호운용성의 문제가 발생하는 경우 정보의 보급과 이용에 제약을 주게 되는데 이러한 일련의 문제를 해결하고 암정보산업을 보호하여 국제화에 대비할 수 있도록 하는 것이 필요하다(강건욱, 2003).

최근 지식의 체계화와 활용에 대한 관심이 높아지면서 각 분야에서 온톨로지의 발전에 많은 노력을 기울이고 있다. 정보과학에서 한정된 영역에서 인간과 기계가 읽고 해석할 수 있는 공통된 개념과 관계를 형식적으로 정의하여 사람과 시스템과 같은 에이전트 사이의 의사소통을 가능케 하는 것이다(Gruber, 1993). 또한, 최신의 상호 운용적인 시스템을 위한 기본적인 하부구조로서 시맨틱 웹을 실현하기 위한 기초 작업이며, 이러한 상호운용성은 다양하고 신중하게 설계된 기본 온톨로지에 의존한다(Pisanellia et al, 2004).

온톨로지를 사용하는 경우 대용량, 복잡한 지식의 체계적 구축이 용이하고, 의미 있는 정보로 저장이 가능하다. 전문가의 개입과 판단에 의해서 재해석이 요구되었던 과거 시스템과는 달리 서로 다른 정보 시스템간의 통합과 상호 운용성을 향상시킴으로써 연계 시스템 개발을 위해 투입되어야 하는 비용을 절감할 수 있는 장점이 있다. 또한 복잡한 기존 의학용어 체계들 간 통합을 위한 도구로 매우 유용하며, 궁극적으로 자료의 재사용과 공유를 활성화함으로써 환자 중심의 의료서비스 제공이 가능하게 된다. 즉, 의사가 아닌 환자 자신이 의료서비스가 필요한 바로 그 시점에서 서비스를 받을 수 있다는 것을 의미한다.

최근에 개최한 Semantic Technologies for E-Government (2004년 9월)의 발표 자료에 의하면 시맨틱 웹 기술의 시장규모가 2003년 20억불에서 2010년에 630억불에 달할 것으로 전망하고 있으며, 맥킨지 보고서에 따르면 온톨로지 기술을 이용한 지식관리 분야는 2006년까지 폭발적인 수요 증가가 예상된다(류광택, 2004). 따라서 온톨로지에 관한 다양한 연구들은 사회적, 기술적, 경제적인 모든 측면에서 매우 중요한 의의를 가진다고 할 수 있다.

정보 검색 시스템은 사용자가 원하는 지식 정보를 얼마나 정확하고 빠르게 검색하여 의미 있는 지식 정보를 제공할 수 있는가에 따라 시스템의 성능과 평가가 좌우된다고 할 수 있다(최호섭, 2004). 웹에서 저장된 정보를 검색하는 방법은 주로 단순 키워드 검색방법을 사용하는데, 이는 검색자가 의도한 정보 이외에도 많은 불필요한 정보를 보여주며, 원하는 정보를 찾을 때까지 다시 검색함으로 많은 시간과 비용을 소요한다. 정보가 의미적으로 연결되어 있지 못하고 단순한 텍스트 형식으로 저장되어 있어서 컴퓨터가 이를 의미적으로 처리를 하기에는 불가능한 형태로 되어 있기 때문이다(이진중, 2003). 온톨로지 기반 정보검색 기술은 중요한 정보가 있는 자원을 빠르게 찾아 사용할 수 있다는 점에서 중요한 기술로 자리잡아 가고 있다. 검색 엔진이 온톨로지에 정의된 개념과 규칙을 활용하면서 사용자의 질의의 의미를 분석하여 그와 관련된 정보를 온톨로지에 표현된 관계에 따라 다시 질의를 적절하게 바꿀 수도 있게 한다(양정진, 2003).

대화형 에어전트는 자연어로 해당 분야의 전문 지식을 알려주는 메신저 기반의 대화형 전문가 시스템이다. 특히 대화를 통해서 사용자 의도를 파악하여 해당 동작을 수행하여 작업에 도움을 준다. 자연 언어 대화는 유연성, 명료성, 표현력 면에서 뛰어난 장점을 가지고 있기 때문에 인간과 컴퓨터의 의사소통 수단으로 사용하기에 적합하다(김경민 등, 2002). 우리의 일상생활 중 인터넷 쇼핑몰 등에서 사이버 안내 도우미와 웹상에서 대화를 함으로써 필요한 문제를 해결했던 경험이 있을 것이다. 더욱 복잡한 문제를 다루는 의료 분야에서 온톨로지 기반의 대화시스템을 통해 필요한 정보를 더욱 효율적으로 제공받을 수 있을 것이다.

차세대 인터넷 기술로 여겨지고 있는 시맨틱 웹 기술에서 핵심 역할을 수행하는 온톨로지 기반의 연구가 매우 활발하게 진행되고 있다. 온톨로지 기반 추론을 이용한 시맨틱 검색 시스템에 관한 연구로는 제주도 숙박과 관광정보를 위한 온톨로지 설계(정은경 등, **2003**), 쇼핑 사이트 고객을 위한 검색 시스템(구미숙 등, **2004**), 영화와 클래식 음악 공연에 관한 검색 시스템(하상범, **2005**) 등이 있다. 유비쿼터스 시스템을 위한 시맨틱 다중 에이전트에 관한 연구(최정화, **2005**)도 최근 진행되었다.

온톨로지 형태의 재사용 가능한 의학체계는 여러 다양한 의료정보 시스템들 간의 통합과 상호 운용성 향상에 절대적으로 필요하다. 이의 활용 사례로는 전자건강기록(EHR), 병원정보시스템, 의학 학술 연구, 지능형 자료 검색 등이 있으며, 의학 온톨로지 기반의 지능형 웹 서비스 기술은 환자 중심의 통합 의료서비스를 제공할 수 있다(이재일, **2004**).

의료분야의 온톨로지에 관한 다양한 연구들이 국내에서 최근 활발하게 시행되고 있지만, 주로 의사 및 연구자 등의 전문가들을 대상으로 한 연구가 대부분이며, 온톨로지 기반의 에이전트 시스템 개발에 관한 연구도 매우 드문 편이다.

이 연구는 현재 웹상에서 일반인들에게 제공되고 있는 위암 관련 정보 중에 FAQ 자료를 바탕으로 온톨로지를 구축하고, 이를 기반으로 한 대화형 에이전트를 개발하여 평가하고자 한다. 따라서 일반인들을 대상으로 온톨로지 기반의 암정보를 제공하는 대화시스템 개발에 관한 연구로는 처음 시도되는 연구이므로 의의가 있다고 할 수 있다.

## 2. 연구목적

이 연구는 일반인을 대상으로 요구에 맞는 보다 효율적인 암관련 정보를 제공하기 위한 온톨로지 기반의 대화시스템을 개발하고 평가하는데 그 목적이 있다. 그 구체적인 목적은 다음과 같다.

첫째, 일반인 대상 위암에 관한 정보를 수집하여 분석한다.

둘째, Protege를 이용하여 온톨로지를 구축한다.

셋째, 온톨로지 기반의 암정보 제공 대화시스템을 개발한다.

넷째, 시스템 평가 및 향후 활용방안을 제시한다.

# II. 이론적 배경

## 1. 암 정보 제공

통계청(2004)에 의하면 2003년 한해동안 우리나라 전체사망자(24만6천명)의 25.9%인 6만4천명(10만명당 사망률 131.8명)이 암(악성신생물)으로 사망하여 사망원인 순위 1위였으며, 그 다음은 뇌혈관질환, 심장질환, 당뇨병, 자살 순으로 나타났다(그림 1).

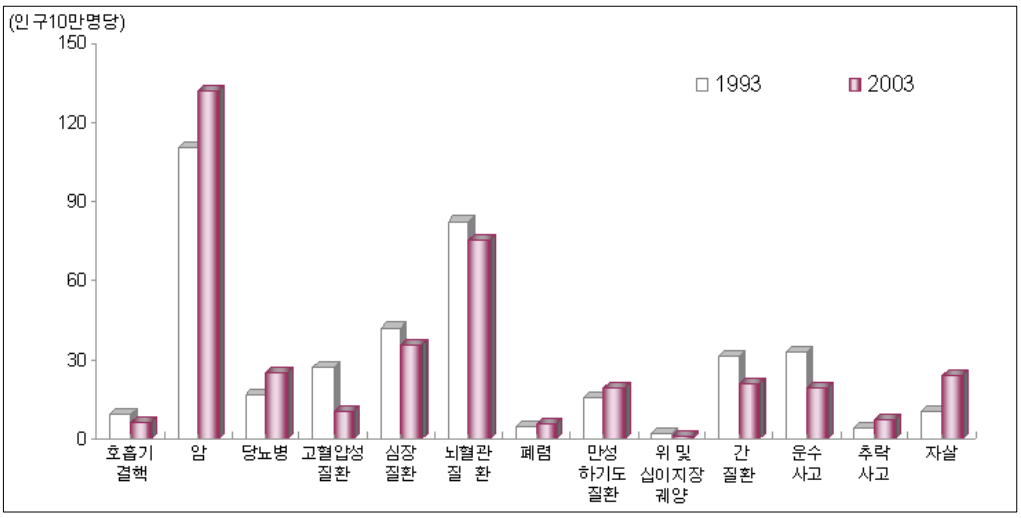


그림 1. 10년전과 주요 사망원인별 사망률 비교 (통계청, 2004)

사망원인 1순위인 암정보에 대한 대중들의 요구는 점점 높아져가고, 이에 대한 질 높은 컨텐츠 뿐만 아니라, 필요한 정보를 효율적으로 제공받고자 하는 요구도 증대되고 있지만 현실은 이러한 요구들을 충족시키지 못하고 있다. 실제로

국내 웹사이트에서 암관련 정보를 찾다보면 기존 제공 자료들이 많이 있기는 하나, 제공자 입장에서 정보들을 제공하기 있기 때문에 정보 검색이 용이하지 않아 결국 사용자들은 상담을 통해 개별적인 답변을 받고 있다.

웹사이트를 통해 조사된 국내 제공 암관련 상담 내용은 정보 사용자적 입장에서 실제로 필요로 하는 내용을 개인별 특성에 맞추어 제공할 수 있다고 판단되는 Q&A형태보다는 정해진 영역 내에서 일반적이고 비특이적인 형태로 상담해 주는 형식인 FAQ의 형태로 제공되고 있어 정보 사용자의 요구를 충분히 만족시키지 못하고 있다. 이와 같은 정보 제공 형태와 내용의 문제는 국내에서 활용할 수 있는 근거에 기반한 암정보의 부족때문이기도 하나 암정보를 비롯한 건강관련 전문상담을 위한 인력과 제공체계의 부재도 중요 원인으로 작용한다고 할 수 있다. 그러므로 개인단위의 맞춤형 정보를 실시간으로 제공하고 있는 선진 외국의 정보 제공체계를 모델로 한 사용자 입장의 암정보 제공을 위한 전문가와 관련 기관의 구체적인 노력이 이루어져야 할 것이다(김추봉 등, **2002**).

선진 외국의 경우 근거에 기반한 표준화된 내용과 형식으로 암정보 제공 시스템을 통해 사용자 친화적으로 제공하고 있다. 미국은 국립암연구소(NCI, National Cancer Center Institute)에서 국가적인 차원에서 암정보 관리를 하고 있으며, CancerNet이라는 포탈사이트를 운영하여 전문가나 일반인들을 위한 광범위한 과학적 지식에 근거한 표준화된 암관련 정보를 제공하고 있다. 캐나다와 영국의 경우도 미국 NCI의 자료를 활용하여 암정보를 국가적으로 제공・관리하고 있으며, 일본의 경우 의학영상참고자료센터를 구축하여 표준화된 암관련 영상 정보를 제공하고 있다(강건욱, **2003**).

## 2. 온톨로지

### 가. 온톨로지의 정의

온톨로지는 지식을 컴퓨터가 처리할 수 있는 상징체계로 만들고 인간이 그 체계를 이해할 수 있도록 구현한 커뮤니케이션 기술이다(김은경, **2004**).

학문으로서의 온톨로지는 19세기 독일 철학자들에 의한 형이상학(metaphysics)의 한 분야로 세계를 구성하는 모든 종류의 대상들의 존재를 탐구하는 철학이며, 아리스토텔레스의 Categories에서부터 비롯되었다. 온톨로지의 의미는 Guarino(**2002**)에 따르면 철학의 존재론 의미에서부터 논리 이론에서 단어라는 의미까지 다양한 의미를 가지지만 정보기술에서의 온톨로지는 지식의 어떤 특정 영역 내에 있는 실체 및 상호작용의 작업 모델을 의미한다.

Sowa(**2000**)가 지식공학적 관점에서 정의한 내용을 보면 온톨로지 연구는 어떤 영역내 존재하거나 존재할 수 있는 모든 대상의 범주화에 대한 연구며, 이러한 관심 분야내에 존재하리라 가정된 대상들의 목록인 결과물을 온톨로지(an ontology)라고 정의하였다. Gruber(1993)는 온톨로지는 "개념화 명세서(specification of a conceptualization)" 혹은 "특정 영역의 용어들과 그들간의 관계를 나타낸 명시적이고 형식적인 명세"라고 정의하였다. Stuber등(1998)은 "온톨로지는 공유된 개념화의 형식적, 명시적인 명세(A formal, explicit specification of a shared conceptualization)"로 정의하였는데, 여기서 공유의 의미는 합의된 지식을 말하며 개념화란 세상의 어떠한 현상에 대한 추상적인 모델을 의미한다. 형식적이란 기계가 해석할 수 있음을 의미한다. 명시적인 명세는 개념, 속성, 관계, 함수, 공리가 형식적으로 정의되는 것을 말한다.

### 나. 온톨로지의 특징과 역할

온톨로지는 인간들 혹은 소프트웨어 에이젼트들 사이의 정보의 구조의 일반적 이해를 공유하고, 재사용하고, 도메인의 가정들을 명백히 하기 위해 사용된다. 공유된다는 뜻은 합의된 지식을 말하며, 이러한 합의는 또한 사용가능한 지식으로부터 도메인 지식을 분리하고, 분석하기 위해 필요하다. 온톨로지의 개념적 접근과 개념간의 관계에 대한 정의의 명확성과 다양성은 정보검색에 있어서 재현율과 정확율을 높이고 질의 형성을 용이하게 한다(McGuiness, 1998).

온톨로지의 역할은 다음과 같다. 첫째, 의미있는 주석은 지식의 식별과 수집 및 개발에 있어서 전문가는 물론 일반인에게도 지식의 습득과 생성의 수단이 될 수 있다. 둘째, 지식통합에 있어서 개념적 일반화와 표현적 형식화를 제공한다는 것이다. 셋째, 온톨로지를 이용하여 개념화한 형식적 구조는 지식보급을 위한 표준화되고 집중적인 접근체계를 갖출 수 있어 문제해결을 위한 실시간 학습환경을 제공한다(Fisher, 2001).

개념을 구조화한 온톨로지는 인간의 의사소통을 도울 수 있다. 또한 온톨로지는 다른 모델링 방법, 패러다임, 언어, 소프트웨어를 사용하는 컴퓨터 시스템 사이에 상호 운용성을 갖게 한다. 이 경우 온톨로지 구축은 해당 영역지식에 대한 공유된 개념을 재사용한다는 점에서 중요하다. 온톨로지는 어떤 주제 영역의 중요한 개념, 관계, 속성을 공식적인 기호로 표현하기 때문에 컴퓨터가 인식할 수 있는 기호화된 표현은 소프트웨어 시스템에서 재사용 가능한 공유요소가 된다. 온톨로지에서 구축된 지식을 지식기반 시스템에 입수하여 개발 속도를 향상시킬 수 있으며 온톨로지에서 입수된 지식은 신뢰도가 높을 것이다(김은경, 2004)

### 다. 온톨로지의 종류

온톨로지는 철학적인 이론으로부터 현실의 문제해결에 관계하는 구체적인 과제까지 넓은 분야에서 다루고 있기 때문에 온톨로지 자신에 관한 정확한 분류와 이해가 혼란스러운 것이 사실이다. Mizoquchi(1997)는 온톨로지를 사용필요성에 따라 컨텐트 온톨로지(Content ontology), 커뮤니케이션 온톨로지(Communication ontology), 인덱싱 온톨로지(Indexing ontology), 메타온톨로지(Meta ontology) 4가지로 분류하였다. 컨텐트 온톨로지는 온톨로지를 설계할 때의 입장을 강조한 재이용 가능한 지식 베이스를 구축하는 것을 고려할 때 필요한 온톨로지이다. 커뮤니케이션 온톨로지는 에이젼트들 간의 의사전달을 위한 최소한의 온톨로지로 데이터베이스의 개념스키마의 유사성과 온톨로지간의 이식성에 관여하는 온톨로지이다. 인덱싱 온톨로지는 지식이 아닌 사례를 대상으로 한 사례베이스 구축을 목적으로 한다. 검색이 목적으로 상세한 기술을 보장하는 개념화는 요구되지 않는다. 메타 온톨로지는 온톨로지 기술언어에서 제공하고 있는 온톨로지 기술을 위한 온톨로지로 기술용의 미디어이다. Ontolingua에서의 프레임 온톨로지(Frame ontology)가 그 예이다.

Guarino(**2002**)는 온톨로지의 종류를 개념화 주제에 따라 Application ontologies, Domain ontologies, Generic ontologies, Representation ontologies 4가지로 분류하였다. Representation ontologies 온톨로지의 예로 Meta-level ontology(Ontolingua's frame ontology)를 들었다. 온톨로지 구축시에 일어나는 어느 세계에서나 일어날 수 있는 시간, 공간, 대상, 사건 등과 같은 개념들의 필요성을 Top-level ontology의 필요성을 강조하였다. 상위 온톨로지는 지식의 공유와 재사용을 위하여 서로 다른 목적에서 같은 대상을 서로 다르게 분류하기

때문에 더 일반적인 수준의 상위 온톨로지가 필요하다는 것이다. ONIONS(ONtological Integration Of Naive Sources)는 개념분석과 온톨로지적 통합을 위한 방법론이다. 어떠한 상호 운용성을 지지하기 위해 제공되는 생산물이다. 만약 적합한 시스템에 사용된다면 말이다. ONIONS 방법론은 의료용어체계를 위한 대규모 온톨로지 라이브러리를 개발하는데 전념한 결과이다. 개발방법은 라이브러리내 모듈을 위한 기술 논리-기반 설계를 개척했으며 일반적인 이론들의 사용을 확장시켰다. ON9 top-level 의학 온톨로지 라이브러리(그림 **2**)는 UMLS, GRAIL-Core, SNOMED, ICD, GABRIELI **5**개의 주요 의학용어체계를 통합하여 1,**500**개 개념에 대한 **50**개 온톨로지의 일련의 셋을 포함한다. 이 모듈은 의료 용어체계들의 통합에 사용된 일반적인 이론들을 포함하며, 합의 결과인 중재 온톨로지들을 포함한다(Pisanellia et al, **2004**).

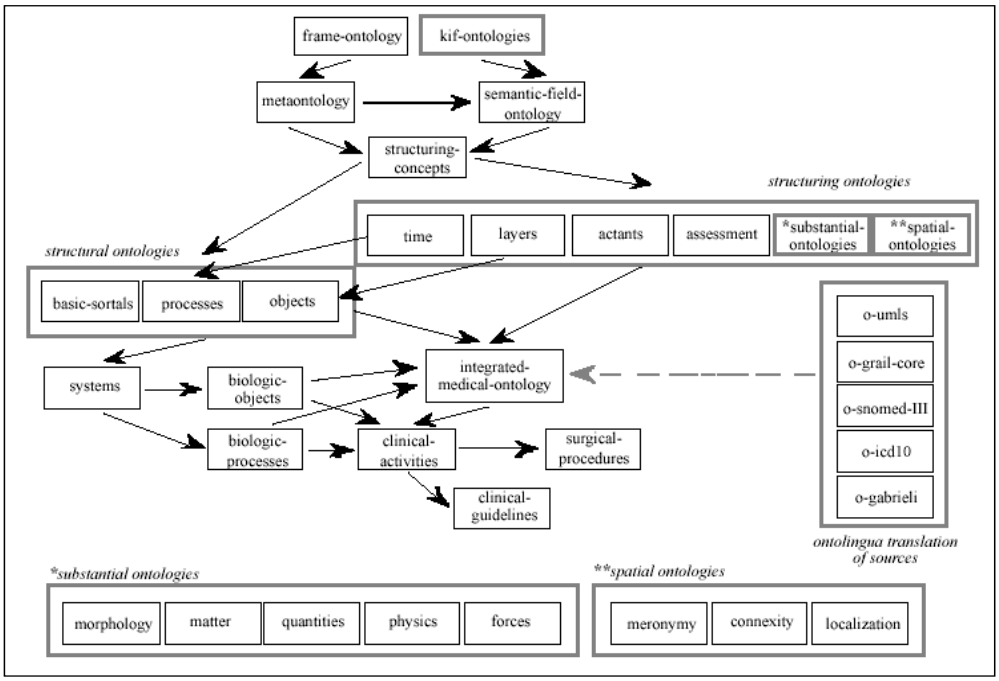


그림 **2**. ON9 top-level ontology library (Pisanellia et al, **2004**)

Guarino와 Van Heijst, Pisanellia 등 여러 온톨로지 연구자들의 유형별 온톨로지의 분류를 정리하면 표 1과 같다(김미정, 2004).

표 1. 유형별 온톨로지 분류

| 유형 | 특징 | 구축 사례 |
|---|---|---|
| Generic Ontology | 1) 공통적인 상위개념을 표현<br>2) 온톨로지를 이용한 자연언어 표현의 분석과 온톨로지 생성에 중요. 재사용에 유용<br>3) 시간, 공간, 프로세스, 인과관계, 부분/전체 관계, 상위/하위 등의 보편성이 높은 기초 개념에 관한 온톨로지의 연구. | DMOZ<br>WordNet<br>Lexcical ontology |
| Representation ontology | 특정분야의 개념표현이 아닌 지식표현형식에 근거한 개념화 | Ontolingua의 Frame ontology |
| Intermediate ontology | 도메인의 일반적 개념과 관계맺는 온톨로지 Domain ontology와 Generic ontology 사이에서 현실적 설계내의 인터페이스 기능을 한다. 많은 중재 온톨로지들이 간단히 "비-일반적인"영역의 상위레벨처럼 행동함. | |
| Top-level ontology | 특정 분야나 문제와는 독립된 어느 세계에서나 적용될 수 있는 기본적인 개념.<br>기존의 시간, 공간, 사건, 대상 등 | UMLS<br>Semantic Net |
| Domain ontology | 영역에 대한 특성화/일반화한 온톨로지<br>의료, 철강등 특정 분야의 온톨로지 개발. 현실문제와의 접점으로서의 기능을 담당 | UMLS,<br>GO(GeneOntology)<br>GLIF<br>HL7 |
| Task Ontology | 특정 업무에 대한 온톨로지<br>인간의 문제 해결 과정을 기술하기 위한 온톨로지의 개발. | 암환자 치료를 위한 가이드라인<br>병리업무 온톨로지 |

자료 : 김미정. 한글 주소(Chief Complaint) 표현에 관한 연구. 2004

### 라. 온톨로지 구성요소

온톨로지는 클래스(Class), 관계(Relations), 함수(Fuction), 공리(Axiom), 인스턴스(**I**nstance) **5**가지 요소를 이용하여 형식화 된다(Corcho, **2000**). 클래스는 추출된 개념에 해당하며, 분류된 개념의 단위를 말한다. 관계는 개념들 사이의 상호관계의 유형이다. 함수는 관계의 특별한 경우로 관계가 값을 가질 때 성립한다. 공리는 적절한 논리 언어로 표시된다, 일차논리(First order logic) 이 예이다. 인스턴스는 특정요소의 실례를 표현한다(표 **2**).

표 **2**. 온톨로지 구성요소

| 요소 | 내 용 | 구 분 | 예시 |
|---|---|---|---|
| 클래스<br>(Class) | - 컨셉, 도메인에서 개념을 설명 | 클래스<br>하위클래스 | 직업<br>교사, 예술가 |
| 인스턴스<br>(**I**nstance) | - 실제 가지는 값 | 클래스<br>인스턴스 | 사람<br>홍길동 |
| 관계<br>(Relations) | - 클래스와 클래스간, 클래스와 인스턴스간 의미적 연관성 정의<br>- 시간 : 순서 의미 필요한 경우<br>- 의존 : 영향의 결과 관계 정의 | 부분과전체<br>계층관계<br>시간관계<br>의존관계 | 손과 팔<br>사람, Staff<br>고장 후 교체 |
| 함수<br>(Fuction) | - 사용자가 미리 정의한 기능적 특성을 가진 함수 | 가정<br>적용 | equal (A, B)<br>equal (x, y) |
| 공리<br>(Axiom) | - **I**S-A 관계에 있는 클래스 구조<br>- 어떠한 논리적 체계를 구성하기 위해 기본 명제들을 증명 없이 받아들이기로 함 | 대칭관계<br>이행관계<br>역관계 | 철수/영희 결혼<br>서울,일본,로마<br>아버지와아들 |

자료 : 이진중. 온톨로지 기반 지능형 검색시스템-역사유물 검색 모델링 중심으로. **2003**

### 마. 온톨로지 방법론

형식적 온톨로지 방법론의 핵심 원리는 다음 5가지로 요약할 수 있다 (Pisanellia et al, 2004).

1) 논리적 일관성(Logical consistency) : 온톨로지는 명시적이고, 형식적인 의미와 논리적 언어를 표현해야 한다.
2) 의미 적용범위(Semantic coverage) : 도메인으로부터 모든 실체가 적용된다는 가정하에 외부 범위부터 그 범위가 가질 수 있는 실체의 형태들도 포함될 수 있도록 노력해야 한다.
3) 설계 명료성(Modeling precision) : 온톨로지는 해당영역에 단지 의도했던 부분의 설계 전부를 포함하여야 한다.
4) 강한 모듈성(Strong modularity) : 다른 온톨로지나 기술 및 상황으로부터 도메인 이론을 조직화하여 가능한 모든 도메인의 개념적 공간을 모듈화해야 한다.
5) 확장성(Scalability) : 표현하는 언어들은 완성하기 위해 도메인과 과업에 따르는 의도된 의미들과 자세한 설명들은 재현될 수 있다. 참조 온톨로지나 시멘틱 웹을 통한 질의 검색 서비스 등의 확장성이 중요하다.

온톨로지 기반 정보검색시스템을 개발하기 위해서는 일련의 기술 정보를 마련해야 하는데, 'OntoWeb' 프로젝트에서 온톨로지 기반 정보검색 애플리케이션 개발과 활용을 위한 가이드라인 항목과 정보검색시스템 설계 단계에 필요한 주요 사항을 나타낸 것이다(최호섭, 2004).

1. 온톨로지를 이용한 애플리케이션 성질 결정

2. 사용자와 관련된 사항을 확인

3. 특정 조직의 정보과 지식에 대한 분석

4. 소유권 문제

5. 특정 조직 내 시스템 조사

6. 추론 처리 결정

7. 평가 기준과 측정 기준 결정

8. 활용 가능한 범위 결정

9. Data noise 고려

10. 존재하는 온톨로지들에 대한 연구

11. 온톨로지 관리 도구와 처리 절차 분석

12. 외부 온톨로지의 유지 관찰

13. 공급자와 소비자 어플리케이션의 통합 고려

온톨로지 개발과정은 일반적인 온톨로지를 개발할 때의 절차를 보여주고 있고, 개발 목적이나 도메인에 따라 다양하게 적용된 방법론들이 존재한다(Noy, 2001). 이렇게 광범위한 범주로 구분된 개발 과정은 세부적인 작업 공정으로 다시 나눌 수 있다(그림 3). 이 모델은 8단계의 작업 공정으로 온톨로지 제작 절차를 구분하고 있다(이진중, 2003).

1단계 온톨로지 영역과 범위 결정

2단계 온톨로지 재사용 고려

3단계 중요 용어 목록 열거

4단계 클래스와 클래스 계층 정의

**5**단계 클래스의 속성-슬롯정의(Property정의)

**6**단계 슬롯의 파셋 정의(Constraints 정의)

7단계 인스턴스 생성

8단계 온톨로지 평가

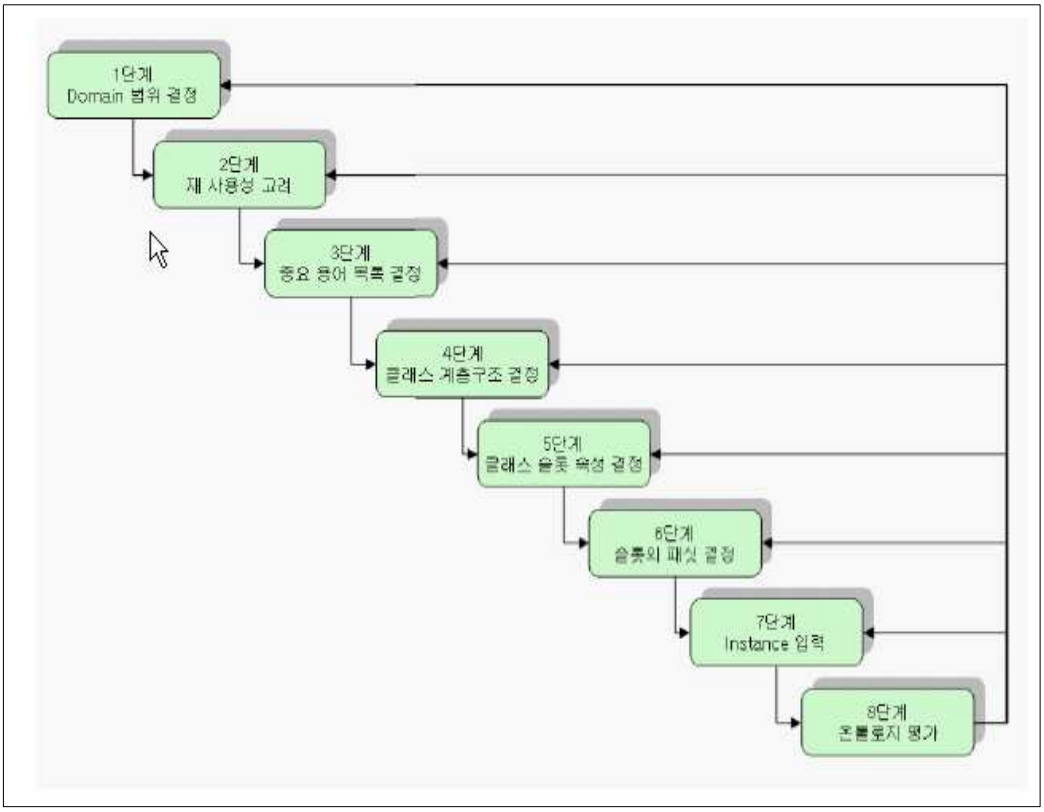


그림 3. 온톨로지 설계 절차 (이진중, **200**3)

## 3. 온톨로지 기반의 활용 시스템

### 가. 지식표현과 온톨로지

지식표현이란 주어진 문제를 해결하기 위하여 사실과 관계성 등을 코드화하여 지식베이스에 저장하는 것을 말하며, 온톨로지는 이러한 지식표현을 위한 한 방법이다. 지식표현은 어떠한 문제를 정확하게 해결하기 위해서는 방대한 양의 지식이 필요한데 그러한 지식을 컴퓨터에 효과적으로 저장하기 위하여 지식표현의 중요성이 제기된다. 지식 표현 시 고려해야 할 사항은 선택된 지식은 일반적이어야 하며, 표현된 지식은 다른 사람이 쉽게 이해할 수 있어야 하며, 표현된 지식에 지식을 삽입·삭제 혹은 수정 시 기존의 지식 표현을 크게 수정함이 없이 작업할 수 있어야 하고, 가능한 추론이 될 수 있도록 지식을 표현하여야 한다. 논리기반의 지식표현 중 일차술어논리(First-order predicate logic)는 적은 수의 원시기호만을 포함하면서도 가장 강력한 표현력을 제공한다. 최근 온톨로지 언어로 가장 주목 박고 있는 것이 기술논리(Descriptive logics)이며, 이는 어휘사용의 의미를 표현하고, 용어의 개념과 기술을 통해 지식을 서술한다. 온톨로지에서의 의미 표현은 "관계"에 의해서다. is-a, instance-of, subset-of, member-of 등의 관계는 가장 많이 사용되는 기본적인 관계이다. 지식표현은 3가지 다른 분야로부터 논리와 기술을 응용하는 다각적인 주제이다. 논리(Logic)은 형식적 구조와 추론의 규칙을 제공하며, 온톨로지(Ontology)는 응용분야에 존재하는 사물의 종류를 정의하는 것이다. 계산(Computation)은 순수철학으로부터 지식표현을 구별하는 어플리케이션을 지원한다(Sowa, **2000**).

Knowledge Meta Process(그림 4)는 주요한 **5**단계로 구성된다. 주요한 흐름은 온톨로지 기반하의 KM 어플리케이션에 이르는 단계들을 보여준다. 타당성 조사(Feasibility study), 시작(Kickoff), 정밀화(Refinement), 평가(Evaluation), 적용 및 개선(Application & Evolution)이다. 단계 하단에 보여지는 Box는 각 스텝의 주요한 하부 스텝 목록이다. 이와 같은 어플리케이션의 개발은 또한 다른 프로세스(소프트웨어 공학과 인간 문제)에 의해 진행된다는 것에 주시해야 한다(Sure, **2002**)

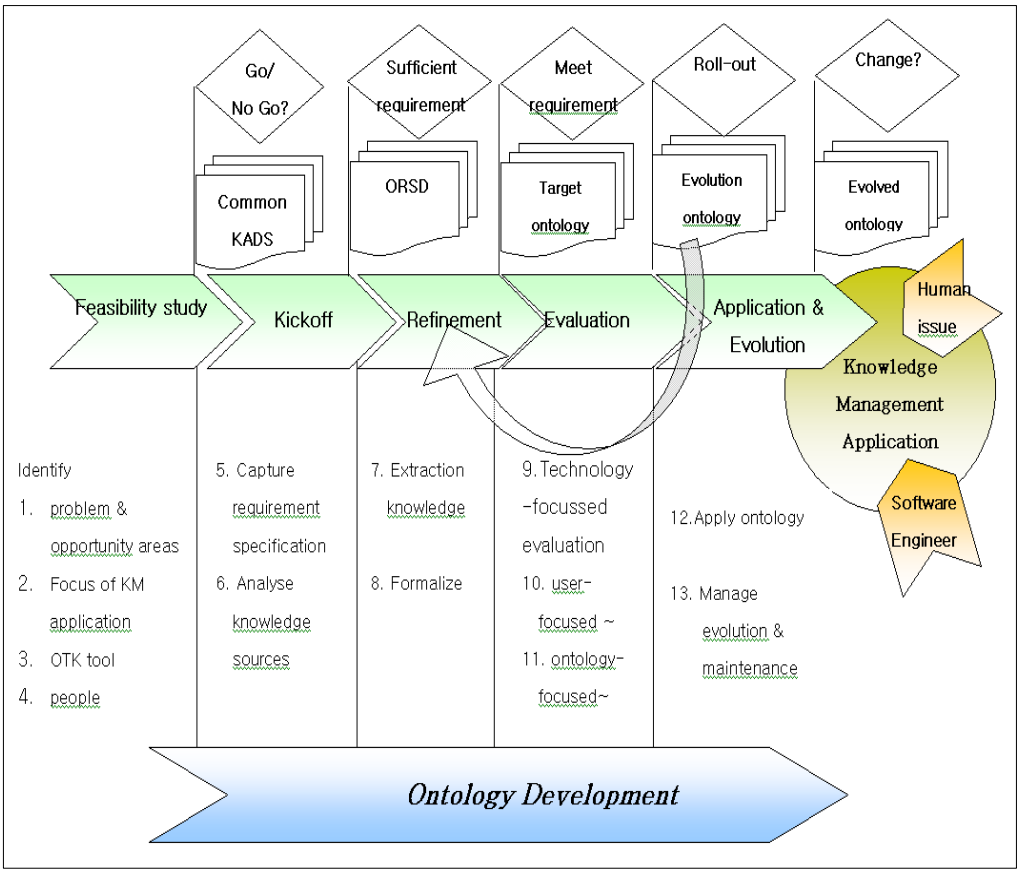


그림 4. Knowledge Meta Process (Sure, **2002**)

### 나. 의료분야에서의 온톨로지 구축사례

의료정보처리를 위한 구조화된 지식처리의 연구는 용어와 분류체계, 지능적 사용자 인터페이스, 의사결정지원, 의미론적 정보처리, 시스템 통합 분야 등에서 활발히 이루어지고 있으며, 이러한 기반을 효과적으로 지원하기 위해서 해당 분야의 적절한 온톨로지의 구축이 핵심적으로 요구된다(Rector, 1999). 최근 대규모 도메인 온톨로지의 발전은 특히, 의학과 생명과학분야에서 많은 노력이 있어왔다(Milward, 2003).

의료분야에서 주목할 만한 온톨로지 구축사례는 다음과 같다(김미정, 2004).

① GALEN

맨체스터 대학을 중심으로 유럽연합에서 개발된 의학용어시스템이다. GALEN의 CORE Model은 해부학, 생리학, 병리학에 있어서의 가장 핵심이 되는 의학개념을 포함하고 있고, 온톨로지 표현언어인 GRAIL (GALEN Representation and Integration Language)은 기술논리기반의 규칙을 제공하여 풍부한 개념표현 방법을 제공한다.

② Gene Ontology Consortium(GOC)

생물학 용어들 간의 관계를 정의하는 온톨로지로 Gene Ontology (GO)를 제안하였다. Gene Ontology는 크게 분자기능, 생물학적 프로세스, 세포요소 세 가지 온톨로지로 구성되어 있다.

③ UMLS(Unified Medical Language System)

미국국립의학도서관(NLM)에서 15개국의 약 100여개의 의료분야에 관련된 용어체계를 통합하여 용어 코드의 각 개념을 연결하여 구축한 광범위

한 통합의학용어체계이다. 어휘 총목록과 의미론적 그물망을 제공하여 의무기록에 사용되는 기초 어휘 목록을 제공한다. UMLS는 4가지의 지식자원으로 구성되어 있다. 의료분야의 다양한 용어, 개념, 명명을 다루는 메타시소러스(Metathesaurus)와 개념을 분류하고, 이들 개념간의 관계를 구축하는 의미망(Semantic network), 자연어 처리를 위해 개발된 특수목록 어휘집(Specialist Lexicon)과 정보자원지도(Information Source Map)로 구성되어 있다.

④ HL-7 RIM (Health Level-7 Reference Information Model)

듀크 대학을 중심으로 진행된 서로 다른 병원 정보시스템간의 정보교환을 촉진할 목적으로 시작되었으며, 많은 제한점에도 불구하고 널리 사용되는 이기종 시스템간의 메시지 전달 및 의무기록에 사용되고 있다. 현재 대부분의 용어의 메시지 표준으로 자리를 잡아가고 있다. HL-7내 표준화 위원회에서는 설계할 가이드라인들의 종류를 결정하고, HL-7 RIM은 다른 온톨로지들과의 조정을 하기 위한 모델이다.

⑤ 그 외에 워싱턴대학의 해부학 정보의 지식기반인 Digital Anatomist, 미국가정의학전문의협회 ABFP(American Board of Family Practice), 가이드라인 기반의 진료(Guideline-Based Medical Care), FMA(Foundational Model of Anatomy) 등이 있다.

### 다. 대화형 에이전트

대화형 에어전트는 자연어로 해당 분야의 전문 지식을 알려주는 메신저 기반의 대화형 전문가 시스템이다. 특히 대화를 통해서 사용자 의도를 파악하여 해당 동작을 수행하여 작업에 도움을 준다. 자연 언어 대화는 유연성, 명료성, 표현력 면에서 뛰어난 장점을 가지고 있기 때문에 인간과 컴퓨터의 의사소통 수단으로 사용하기에 적합하다(김경민 등, **2002**).

질의에 대한 답변을 수행하기 위하여 필요한 지식을 스크립트에 기술한다. 표 3은 대화형 에이전트에 사용되는 지식표현 방법을 보여준다. 표 3의 다양한 기법들을 기초로 대화형 에이전트는 웹페이지 소개 도우미나 프로그램 사용 도우미, 상품구매 도우미, 여행경로 안내 도우미 등으로 연구되고 있다. 이 외에도 상용화된 대화형 에이전트로는 NativeMinds사의 Nicole, Artificial Life사의 SmartBot, **V**irtual Personalities사의 **V**erbot, 국내에서 개발된 MSN messenger상의 심심이, 휴보그 등이 있다(김경민 등, **2002**).

표 3. 대화형 에이전트에 사용되는 방법

| 기술 | 수행 작업 예 | 기술난이도 | |
|---|---|---|---|
| 단순패턴매칭 | 웹검색엔진 | 단순질의응답 | 쉬움 |
| 제한대본기반모델 | 게임 | 단순한 선택을 통한 대화진행 | ↕ |
| 유한상태모델 | 장거리전화 | 상태전환을 통한 제한된 유연성 | |
| 프레임기반모델 | 열차정보검색 | 단순패턴매칭을 이용한 정보기입 | |
| 계획기반모델 | 부엌디자인도우미 | 동적주제변화 | 복잡함 |

## 4. 온톨로지 시스템 평가 및 활용

### 가. W3C OWL Test Cases

국제 웹 표준화 기관인 W3C에서는 시맨틱 웹 및 웹 온톨로지 언어에 관련된 다양한 표준을 제정하고 있을 뿐만 아니라 표준 명세를 구현한 결과물에 대한 시험을 할 수 있는 표준 Test Case들을 마련하고 있다. W3C에서 정의하고 있는 OWL Test Case는 다음과 같이 크게 8가지의 종류로 구성된다.

- Entailment Tests
- Non-Entailment Tests
- OWL for OWL Tests
- Consistency Tests
- Inconsistency Tests
- Import Entailment Tests
- Import Level Tests
- Tests for Incorrect Use of OWL Name space

이중 추론 엔진의 성능을 시험하기 위해 가장 일반적으로 활용될 수 있는 시험은 Entailment Test로서 주어진 전제(Premise)에 대해 올바른 결론(Conclusion)이 도출되는 지를 검사하는 시험이다(표 4). 현재 Entailment Test는 인증된 72항목의 Test Case와 제안된 25개의 항목으로 구성되어 있으며 계속적으로 새로운 Case들이 추가되고 있는 상태이다(하영국, 2003).

표 4. W3C Entailment Test 결과 비교

| Implementations Test | OWLer | Euler | FaCT | FOWL | Hoolet | Jena2 | Pellet | Surnia |
|---|---|---|---|---|---|---|---|---|
| Tests by Function: 44 Approved (# of passed tests) | 88% (39) | 100% (44) | 2% (1) | 63% (28) | 38% (17) | 84% (37) | 100% (44) | 54% (24) |
| Tests by Issue: 15 Approved (# of passed tests) | 66% (10) | 100% (15) | 6% (1) | 13% (2) | 40% (6) | 80% (12) | 80% (12) | 0% (0) |
| Description Logic Tests: 13 Approved (# of passed tests) | 53% (7) | 100% (13) | 15% (2) | 0% (0) | 92% (12) | 23% (3) | 61% (8) | 0% (0) |
| Total: 72 Approved (# of passed tests) | 77% (56) | 100% (72) | 5% (4) | 41% (30) | 48% (35) | 72% (52) | 88% (64) | 33% (24) |

자료 : OWL Web Ontology Language Test Cases. Available from URL: http://www.w3.org/TR/2004/REC-owl-test-20040210/#testEntailment Accessed May 6, 2005

### 나. 실사용 평가

Achour(2001)의 연구에서는 실사용 평가 방법을 사용하였다. 첫번째 평가는 CDSS 결합력과 정확성을 검증하기 위해 수행하였는데, 30개 처방에서 결정된 수혈 처방의 신뢰성을 조사(적혈구 수혈 97%, 혈소판 수혈 98%)하였다. 수혈 제품의 자격에 대한 시스템과 의사의 불일치는 30사례 중 18사례(60%)였으며, 불일치는 환자 상태에 대해 더 많은 정보를 더하고 지식 베이스에 규칙을 수정할 필요가 있다는 것을 보여준다. 새로운 20개

의 처방은 수정된 CDSS 사용 평가를 위해 사용되었다. 몇몇의 불일치는 의사의 처방 순서와 CDSS 결론(40%) 사이에 지속적이었다. 그러나 CDSS와 그 순서를 재검토한 전문가 사이의 동의는 100% 의사가 현재의 주입 지침을 따르지 않은 사실에 의해 설명되었다. 운영 위원회는 새로운 주입 지침을 고수하는 현재 실행을 수정하기 위해 지명되었다. 원형은 Henri Mondor 병원에서 현재 평가가 진행 중이다.

박경모(2002)의 연구에서는 GOMS(Model of Goals, Operators, Methods, and Selection rules)방식으로 사용성 테스트를 수행하였다. 텍스트 입력방식에 비해 2배정도의 입력속도의 증가를 가져왔으며, 추후 입력의 정확성 등을 포함한 추가 사용성 테스트를 수행할 예정이다. GOMS란 목표, 조작자, 방식, 선택 규칙 모델을 이용한 인터페이스 정량 평가는 키보드에서 키를 누르는데 걸리는 시간, 화면상의 특정한 지점을 가리키는데 걸리는 시간, 사용자의 손을 키보드에서 옮기는데 걸리는 시간, 다음 단계로 이행하기 위해 정신적으로 준비하는데 걸리는 시간, 사용자가 입력 후 컴퓨터의 반응을 기다리는 시간을 측정하여 분석하는 방법이다.

김경민(2002)의 연구에서는 6개 주제별 60여개 토픽으로 구성된 스크립트 구축 후 대화를 실시하고, 분석을 통한 높은 점수의 매칭 결과가 답변으로 출력되는 지 여부를 확인하였으며, 2차 실험은 인터페이스 성능 비교를 위한 사용성 테스트 실시하였다. 실험 대상자는 대화형 에이전트 테스트 경험이 없는 대학생 12명으로 구성하여 10가지 토픽에 해당하는 스크립트를 입력하는데 걸리는 시간을 각각 측정하는 방법을 사용하였다.

### 다. 다면 평가

Sure(**2002**)의 연구에서는 3개의 다른 유형의 평가방법으로 분류하였다. 즉 기술 중심, 사용자 중심, 온톨로지 중심의 3가지 유형이 해당된다.

① 기술 중심의 평가

평가를 위한 프레임웍은 두가지 측면으로 구성된다. 이는 툴에 의해 생성된 온톨로지들의 속성에 대한 평가(예를 들어 언어 유사성, 일관성)를 하는 것과 기술적 속성(즉, 평가툴 속성 자체에 대한 평가까지 포함된 툴 또는 어플리케이션에 대한 평가, 예를 들면 상호 운용성, 규모 등)을 포함한다.

② 사용자 중심의 평가

평가에서의 주안점은 사용자들이 KM 어플리케이션에 만족하는지에 대한 평가, 유사한 업무를 해결하는데 기존에 존재하는 어플리케이션보다는 최소한 더 나아야 한다는 것이다.

③ 프로세스 및 실용중심의 평가

평가 방법 외에 정식으로 온톨로지를 평가할 필요가 있다. 가장 나은 접근방법중 하나는 ontoClean 접근법으로 이는 철학적인 개념에 의거 좀 더 정확한 온톨로지의 계층 구조를 유도한다.

### 라. 유용성 평가

Campbell(**2001**)의 연구는 임상진단의 인코딩을 위한 자연어 인터페이스를 지원하는 SNOMED RT의 유용성(Utility)을 평가하는 것이다. UNMC

어휘집에서 전체 언어 Set의 1%에 해당하는 128개의 임상진술을 랜덤 샘플링한다. 용어는 치료전반에 걸쳐 표현되며, 완전히 임상적이다. 용어의 전체 샘플내의 자유도를 합하면, 통계의 각각의 부분율을 계산하는 것이 가능하다.

① 자유도(DOF ; Degree of freedom)

임상 진술내 의미의 복잡성의 수치적 평가로 표준적 인코딩 안의 RT개념들을 모두 세어서 계산한다.

② 어휘적 신호(Lexical Cue)

RT 용어 Set 내의 용어 혹은 동의어로 나타낸 형태소, 단어 혹은 구문이 RT 개념과 관계된 어휘적 신호이다.

예) myocardial infarction은 3개의 자유도에 따르는 어휘적 신호임

- myocardial infarction, myocardium, infarct 연결 가능
- cardi, infarct, heart, myocard 형태소는 구성요소로 연결 가능

### 마. Recall & Precision Test

재현율(Recall)과 정확률(Precision)은 검색 효과의 평가를 위해 일반적으로 사용되는 척도이다. 재현율은 전체 적합 문서에 대한 검색된 적합 문서의 비율을 나타내고 정확률은 전체 검색 문서에 대한 검색된 적합 문서의 비율이다. 즉, 재현율은 문서 집합에서 사용자가 원하는 문서를 어느 정도 검색하였는가를 나타내고, 정확률은 검색된 문서들 중에서 사용자가 원하는 문서가 얼마나 포함되어 있는가를 나타낸다. 이때 검색 효과 평가 척도 재현율과 정확률은 표 **5**와 같이 표현될 수 있다(Salton et al, 1983).

표 **5**. 재현율과 정확률 공식

$$\text{재현율} = \frac{\text{검색된 적합 문서 수}}{\text{적합 문서 수}} = \frac{\text{RETREL}}{\text{RETREL} + \text{NRETREL}}$$

$$\text{정확률} = \frac{\text{검색된 적합 문서 수}}{\text{검색된 문서 수}} = \frac{\text{RETREL}}{\text{RETREL} + \text{RETNREL}}$$

자료 : Salton et al, Extended Boolean Information Retrieval. 1983

김민경(**2005**)의 연구에서는 유방종양영상 검색 시스템의 성능평가 척도로써 재현율과 정확률을 사용하였다. 내용기반 영상 검색의 효율성을 부여하기 위해 일반적으로 정확한 매칭이 아닌 유사 매칭을 수행하는 내용기반 검색에 많이 사용되고 있는 Recall과 Precision을 사용하였다. Recall은 영상 데이터베이스에서 질의와 관련된 영상 중 검색된 영상의 비율이고, Precision은 검색된 영상 중에서 질의와 관련된 영상의 비율을 나타낸다. Recall이 크다는 것은 올바른 영상을 놓치지 않는다는 것을 뜻하며, Precision이 높다는 것은 잘못 검출된 영상이 많지 않다는 것을 뜻한다.

미국 국립의학도서관의 MEDLARS(Medical Literature Analysis and Retrieval System)의 성능 테스트는 Mesh에 제공된 색인 엔트리와 함께 생체임상의학 기사들의 DB, 의학 주제명, 시소러스, 그리고 MEDLARS의 운영 DB에서 이루어졌다. MEDLARS 테스트의 목적은 현 시스템을 평가하고 개선방안을 찾기 위함이다. 탐색 샘플이 완성되고 적합성 평가를 위해 이용자들에게 보내지면 이용자들은 아래와 같은 범위 내에서 검색된 항목들을 표시해야 했다(이두영, **200**3).

재현율과 정확률은 아래의 공식으로 수치화했다.

( H1 : 매우 가치 있음, H2 : 가치 적음, W1 : 가치 없음, W2 : 매우 가치 없음)

$$\text{Recall ration} = \frac{H1 + H2}{T} \times 100$$ (T는 적합한 항목의 총수)

$$\text{Precision ration} = \frac{H1 + H2}{L} \times 100$$ (L은 검색된 샘플 항목의 수)

**바. F - measure**

문서 분류 결과에 대한 성능 평가를 위한 척도로 Recall, Precision, F - measure 값을 주로 사용한다(Buckland, 1994). Recall 값은 카테고리 내의 전체 문서 중에서 정확하게 분류된 문서의 분류 비율을 의미한다. Precision 값은 분류된 문서 중에서 정확하게 분류된 문서의 비율을 의미한다. 다음 식은 Recall과 Precision의 정의식이다.

$$\text{Recall} = \frac{\text{정확하게 분류된 문서수}}{\text{해당 카테고리의 문서수}} \times 100$$

$$\text{Precision} = \frac{\text{정확하게 분류된 문서수}}{\text{분류된 전체 문서수}} \times 100$$

이러한 Recall과 Precision 값은 서로 반비례 관계에 있으므로 적절한 조정 과정이 필요하다. Lewis 등(1994)은 Recall과 Precision의 문제점을 보완하기 위하여 Recall과 Precision을 결합한 F - measure 개념을 제안하였다. 다음 식은 F - measure 에 대한 정의식이다.

$$F_{\beta} = \frac{(\beta^2 + 1) \times \text{Precision} \times \text{Recall}}{\beta^2 \times \text{Precision} + \text{Recall}}$$

$\beta$는 Recall과 Precision 값의 중요도에 따른 가중치를 나타낸다. 즉 $\beta = 0$ 이면 F 값은 Precision 값과 동일하고, $\beta = \infty$ 이면 F 값은 Recall 값과 동일하다. $\beta = 1$ 이면 Recall 값과 Precision 값에 동일한 가중치를 적용하여 F 값을 계산한다. 그리고 $\beta = 0.5$ 이면 Precision 값에 2배의 가중치를 적용하여 Precision 값의 중요도를 높여서 계산한다. 그러므로 Recall과 Precision 값의 중요도에 따라 $\beta$의 가중치를 선택적으로 조정할 수 있다.

김지숙(2001)의 연구에서도 제안된 대표 색인어 추출 기법의 정확성을 검증하기 위하여 Recall 값과 Precision 값에 동일한 가중치를 적용하여 F-measure 값을 비교함으로써 분류 성능을 평가하였으며, 김병찬(2001)의 연구에서 제안한 추천 시스템 모델에 대한 효율성 검증을 위하여 Recall, Precision, F-measure 측정 기법에 의해 비교 실험하였다(그림 5).

| 평가방법 | Recall | | Precision | | F-Measure | |
|---|---|---|---|---|---|---|
| 횟수 | 1 | 2 | 1 | 2 | 1 | 2 |
| 제안기법 | 0.8 | 0.71 | 0.87 | 0.57 | 0.73 | 0.63 |
| 협동적필터링 | 0.4 | 0.2 | 0.87 | 0.25 | 0.5 | 0.22 |

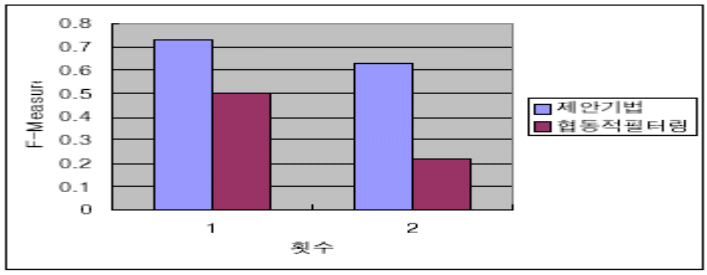


그림 5. 모호 집단의 F-measure(김병찬, 2001)

### 사. 시스템 활용

온톨로지 형태의 재사용 가능한 의학체계는 여러 다양한 의료정보 시스템들 간의 통합과 상호 운용성 향상에 절대적으로 필요하다(그림 **6**). 이의 활용 사례로는 전자건강기록(EHR), 병원정보시스템, 의학 학술 연구, 지능형 자료 검색 등이 있으며, 의학 온톨로지 기반의 지능형 웹 서비스 기술은 환자 중심의 통합 의료서비스를 제공할 수 있다. 기대 효과로는 의료기관들의 각종 정보시스템 개발 비용을 절감할 수 있으며, 수집된 의료정보를 연구나 교육을 위해 재활용이 가능하다. 또한 의료기관간 원활한 정보 공유를 통해 양질의 진료에 기여할 수 있고, 의사결정지원시스템 등 전문가 시스템 등을 활용하여 진료 오류 감소에 기여할 수 있으며, 온톨로지 기반의 다양한 지능형 의료 서비스의 개발이 가능하다(이재일, **2004**).

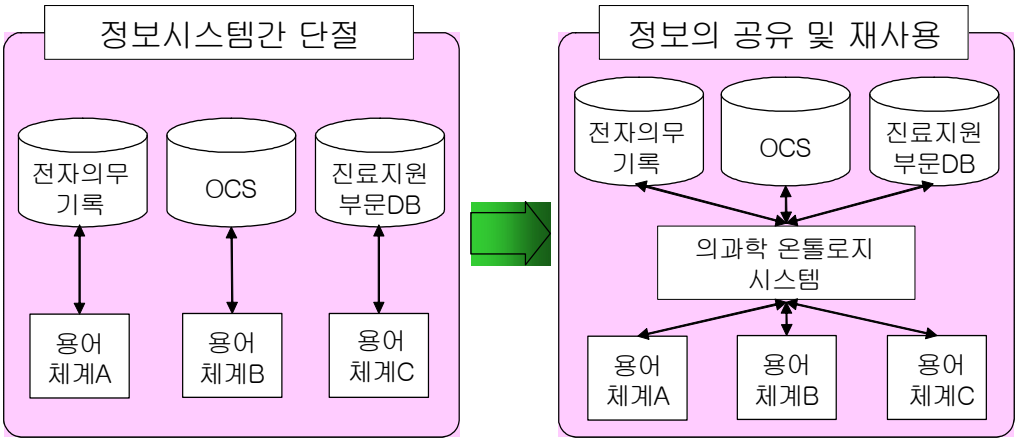


그림 **6**. 온톨로지 지식표현시스템의 정보공유와 재사용에서의 역할(이재일, **2004**)

# III. 연구방법

## 1. 연구대상 및 범위

이 연구는 암관련 정보 제공 웹사이트의 FAQ 자료를 대상으로 하였다. 2005년 4월 시점에서 병원, 협회, 업체 등에서 운영중인 국내 Web site를 조사하였으며, 일반인을 대상으로 제공되는 FAQ 자료 중에 위암 관련 정보로 제한하였다.

해당 웹사이트의 선정시 제공 정보에 대한 신뢰성을 확보하기 위해 선행 연구로써 시행되었던 대국민 암 관리 정보 Call-center 구축을 위한 암정보 표준화 개발 연구(강건욱, 2003)의 조사 대상에 포함시켰던 웹사이트와 국립암센터 웹사이트로 국한하였다. 또한 온톨로지 구축 시 제한적 영역에서 규칙성을 확보하기 위해서 너무 구체적이고 개별적인 개인 상담 내용은 포함시키지 않았으며, 자주 묻는 질문과 답변으로 일반화되어 있는 FAQ 자료 중에서 위암 관련 자료를 선택하였다.

## 2. 연구의 틀

**일반인 대상 위암에 관한 정보를 수집 및 분석**

□ 암정보 제공 Web site FAQ 자료 수집 및 분석
- 국내 병원, 협회, 업체 등에서 운영중인 Web site 대상

⇩

**Protege를 이용하여 질문용어에 대한 온톨로지 구축**

□ 질문 분석을 통한 분류 체계를 바탕으로 한 온톨로지 구축
- 영역 결정, 재사용성 고려, 용어 목록, 계층 정의, Protege 입력

⇩

**온톨로지 기반의 암정보 제공 대화시스템 개발**

□ 온톨로지를 이용한 대화형 에이전트 개발
- Protege Data를 XML로 전환하여 대화형 에이전트와 연결

⇩

**시스템 평가 및 향후 활용방안을 제시**

□ 시스템 평가 및 활용 방안 제시
- 시스템 성능 평가와 사용성 평가 시행 및 활용 방안 제시

그림 7 . 연구의 틀

## 3. 연구 분석 및 방법

### 가. 위암 정보 FAQ 자료 수집과 분석

암정보 제공 웹사이트의 위암관련 FAQ 자료를 질문과 답변으로 구분하여 엑셀 파일에 직접 입력하여 수집하였다. 입력한 질문과 답변 내용 중에서 중복되는 질문과 답변은 제외하였으며, 2가지 이상의 질문과 답변이 있는 경우는 1가지로 각각 분리하여 정리하였다.

질문에 대한 키워드를 추출하였으며, 해당 질문에 대해 중요하다고 생각되는 우선순위별로 키워드를 정리하였다. 유사하거나 불필요한 키워드는 합치거나 제거하였다.

### 나. Protege 이용 질문 용어에 대한 온톨로지 구축

수집된 질문을 분석하여 질문 온톨로지를 구축하였다. 온톨로지 구축 개발도구로는 Protege 2000을 사용하였다. Protege 2000은 스탠포드 대학의 의학정보 그룹인 SMI(Stanford Medical Informatics)에서 사용자 그룹과 공동으로 개발해 나가는 오픈-자원(Open-source) 온톨로지 구축 툴로서 지식관리 구축을 위한 통합환경을 제공하며, 풍부한 지식모델을 제공한다. 또한, XML 형태로 변환하는 기능이 있어서 온톨로지 문서를 작성할 수 있다.

### 다. 온톨로지 기반의 암정보 제공 대화시스템 개발

인터페이스는 **V**isual Studio.NET **200**3(Microsoft corporation)을 사용하여 프로그래밍 하였다. 기존의 대화형 에이전트를 참조하여 인터넷을 기반으로 동작할 수 있게 ASP.NET를 사용하였고, 비하인드 코드로서 C#을 사용하였다. DBMS는 MS-SQL Server(Microsoft corporation)를 사용하였으며, 비연결성을 보장하기 위하여 ADO.NET으로 DBMS에 접근하였다. 질문DB와 대답DB는 R-DB 형태이며, 질문용어 온톨로지DB는 프로티지DB와 R-DB를 사용하여 구축하였다.

### 라. 시스템 평가 및 활용 방안 제시

온톨로지 시스템의 검색 성능을 평가하는 시스템 성능 평가와 사용성 평가를 시행하였다. 시스템 성능 평가로써 검색시스템의 재현율과 정확률, F-measure를 암관련 전문의사인 전문가가 평가함으로써 타당성을 확보하였다. 시스템 사용성 평가는 기존 웹사이트와 온톨로지 대화시스템에서 각각 위암관련 정보검색을 실시하여 두 가지 시스템에서의 검색 소요 시간, 검색 단계, 검색 결과 만족도 항목을 비교하였다. 이 연구에서 사용된 모든 자료 분석의 통계 처리는 SAS program 8.1 version을 이용하여 $X^2$- test와 t - test를 시행하였다.

온톨로지 활용 시스템에 관한 평가를 시행한 결과를 바탕으로 향후 활용 방안을 수립하였다.

# IV. 연구결과

## 1. 위암 정보 FAQ 자료 수집과 분석

### 가. 위암 정보 FAQ 자료 수집

총 11개의 웹사이트를 조사하였으며, 현재 운영이 되고 있지 않거나 FAQ 자료를 제공하지 않는 웹사이트는 제외하였다(표 **6**).

표 **6**. 조사 대상 웹사이트

| 웹사이트명 | URL |
|---|---|
| 국립암센터 암정보서비스 | http://www.ncc.re.kr/ |
| 서울대학교 혈액종양내과 암종합정보망 | http://hmo.snu.ac.kr/ |
| 연세대학교 세브란스병원 암센터 | http://www.severance.or.kr/cancer |
| 고려대학교의료원 종양혈액내과 | http://www.korea.ac.kr/~kuho/ |
| 사단법인 대한암협회 | http://www.kocancer.org |
| 암닥터 | http://www.amdoctor.com/ |
| Hi-doc 사이버종합병원클리닉 | http://www.hidoc.co.kr/ |
| 암이란닷컴 | http://www.am**2**ran.com/ |
| 코어메드 | http://www.coremed.co.kr/ |
| 에버케어 | http://www.evercare.co.kr/ |
| 원자력병원 암유전상담 클리닉 | http://www.geneclinic.re.kr/ |

위 표의 상위 7개 웹사이트에서 일반인을 대상으로 제공되는 FAQ 자료 중에 위암 관련 정보를 수집하였다. 너무 구체적이고 개별적인 개인 상담 내용은 포함시키지 않았으며, 자주 묻는 질문과 답변으로 일반화되어 있는 FAQ 자료 중에서 140개의 위암 관련 자료를 선택하였다(표 7).

표 7. 질문과 답변 수집자료(예시)

| 질문 | 대답 |
|---|---|
| 위암이란 무엇인가요? | 위에서 발생하는 모든 암을 위암이라고 합니다. 종류로는 점막에서 발생한 위선암, 신경 또는 근육에서 발생한 기질성암, 림프조직에서 발생한 위림프종 등이 있습니다. 그러나 위선암이 대부분을 차지하기 때문에 위암이라고 하면 일반적으로 위선암을 지칭한다고 할 수 있습니다. |
| 위염이나 위궤양이 오래되면 위암이 되나요? | 위점막 병변 중 만성 위축성 위염은 전암성 병변입니다. 위축성 위염은 위점막이 위축되어 위산분비가 감소하는 병으로 위 속이 저산증이 되면 세균증식이 일어나게 됩니다. 위 속에 세균이 증식하면 섭취한 식품 속의 질산염과 아민이 결합되어 발암성이 강한 니트로조아민이 발생하는 것입니다. 그러나 일반적인 급성위염이나 만성위염은 위암과의 연관성이 없습니다. |
| 위암의 재발 및 전이 징후는 어떠한가요? | 암에 의한 특이한 증상은 없으며, 암이 진행함에 따라 동반되는 상복부 불쾌감, 팽만감, 동통, 소화불량, 식욕부진, 체중감소, 빈혈 등의 진행성 전신증상이 있을 수 있으나, 어느 부분으로 전이되느냐에 따라 징후는 달라질 수 있으므로 정기적인 진료를 통한 꾸준한 관찰이 필요합니다. |

자료 : 국립암센터 암정보서비스 FAQ. Available from URL: http://www.ncc.re.kr. Accessed May **6, 2005**

### 나. 위암 정보 FAQ 자료 분석

초기에 수집된 위암관련 FAQ 자료는 140개의 질문과 답으로 구성되었으나 수집된 질문과 답변 내용을 분석하여 중복건은 제외하고, 복합건은 분리하여 총 126개의 질문과 답변으로 최종 정리되었다.

126개 각각의 질문에 대한 키워드를 추출하였으며, 해당 질문에 대해 중요하다고 생각되는 우선순위별로 키워드를 정리하였다. 유사하거나 불필요한 키워드는 합치거나 제거하여서 총 286개로 정리되었다(표 8).

표 8. 질문 키워드 정리 (예시)

| 질문 | 키워드 |
|---|---|
| 위염이나 위궤양이 오래되면 위암이 되나요? | 위암, 예방, 위험요인, 원인, 위염, 위궤양 |
| 위암치료 후 급격한 체중감량이 있습니다. 식이요법으로 어떤 것이 좋을까요? | 위암, 생활, 식생활, 식이요법, 체중감량 |
| 위암을 진단 받고 얼마나 빨리 수술을 받아야 하나요? | 위암, 치료방법, 수술, 진행속도, 수술시기 |
| 위암의 항암화학요법에 사용되는 약은 어떠한 것이 있나요? | 위암, 치료방법, 항암화학요법, 항암제, 약물치료 |

### 다. 자료 분석을 통한 중요 용어 목록 열거

질문 자료 분석을 통해 추출된 총 **286**개의 키워드 중에서 국립암센터 암정보서비스 용어사전에서 찾을 수 있었던 **95**개의 용어에 대한 정의를 발췌하였다(표 9).

표 9. 중요 용어와 정의 (예시)

| 중요 용어 | 정의 |
|---|---|
| 위궤양 | 위의 점막과 점막하층을 지나 근육층까지 손상을 입은 상태입니다. |
| 항암화학요법 | 경구용 또는 주사용 약물을 사용하여 암을 치료하는 방법입니다. 수술이나 방사선 치료와는 달리 국소적인 치료뿐만 아니라 전신적인 요법으로 시행됩니다. 이는 단독 사용보다는 수술이나 방사선 치료와 병행하여 사용됩니다. |
| 생존율 | 보통 생존율이라 함은 암과 같은 특정질병에 걸린 사람 중에서 특정기간 동안 생존하는 비율을 말합니다. |
| 위내시경 | 식도를 통해 위를 들여다 볼 수 있는 거울입니다. 가느다랗고 긴 부드러운 관끝에는 렌즈와 불빛이 달려있어 위와 식도부분을 육안으로 보는 것처럼 볼 수 있습니다. 검사를 위한 용도로도 사용하지만 간단한 용종의 제거 등의 수술적 용도로도 사용됩니다. |
| 식이요법 | 질병을 치료하기 위하여 식사를 조절하는 보조 의료법입니다. 감식요법 또는 식사요법이라고도 합니다. 어떤 종류의 질병(특히 당뇨병·신질환·위장질환·간질환 등)을 적극적으로 치료하기 위해 의사의 지시에 따라 정상 식사를 수정함으로써 소화·영양흡수를 가능하게 하고, 동시에 병상을 호전시키는 중요한 보조의료 방법입니다. |

자료 : 국립암센터 암정보서비스 용어사전. Available from URL: http://www.ncc.re.kr. Accessed May **6, 2005**

## 2. Protege 이용 질문 용어에 대한 온톨로지 구축

### 가. 클래스와 클래스 계층 정의

온톨로지 구축 개발도구로는 Protege **2000 V**er 3.**0**을 사용하였다. 우선 질문의 클래스를 3가지로 결정하였으며(표 **10**), 질문을 클래스 단위로 구분하였다(표 11).

표 **10**. Class 결정

| **Class** | | **Comment** |
|---|---|---|
| 질문 상황 | Sub Domain | 질문을 하고자 하는 상황<br>즉, 증상 또는 질병의 상태 등을 표현함 |
| 질문 주제 | Subject | 질문의 주제<br>즉, 어떤 부분을 알고 싶은지 구체적으로 표현함 |
| 질문 유형 | Lexicon | 질문의 유형<br>즉, 상황과 주제에 따른 질문의 형태로 의문문을 나타냄 |

표 11. 질문 Class 구분(예시)

| **구분** | **질문 상황** | **질문 주제** | **질문 유형** |
|---|---|---|---|
| 질문 1 | 위암을 | 예방할 수 있는 방법은 | 어떤 것이 있나요? |
| 질문 **2** | 위암의 | 발생 원인은 | 무엇인가요? |
| 질문 3 | 위염이나 위궤양이 | 오래되면 위암이 | 되나요? |
| 질문 4 | 위암 수술 후 | 항암화학요법의 주기는 | 얼마나 되나요? |

Hajdukiewicz(**2001**)는 수술실이라는 특정분야에서의 의료문제해결에 관한 온톨로지 연구를 다루면서 전통적 접근법인 생물의학적 지식모델링에서 추론계층 분석, 작업분야모델 구축, 인간-컴퓨터 Display를 위한 통합의료정보 설계로 진행하였으며, 이 연구에서도 위암정보라는 특정 도메인에서의 지식들을 조직화하는데 온톨로지를 활용하였다(그림 8).

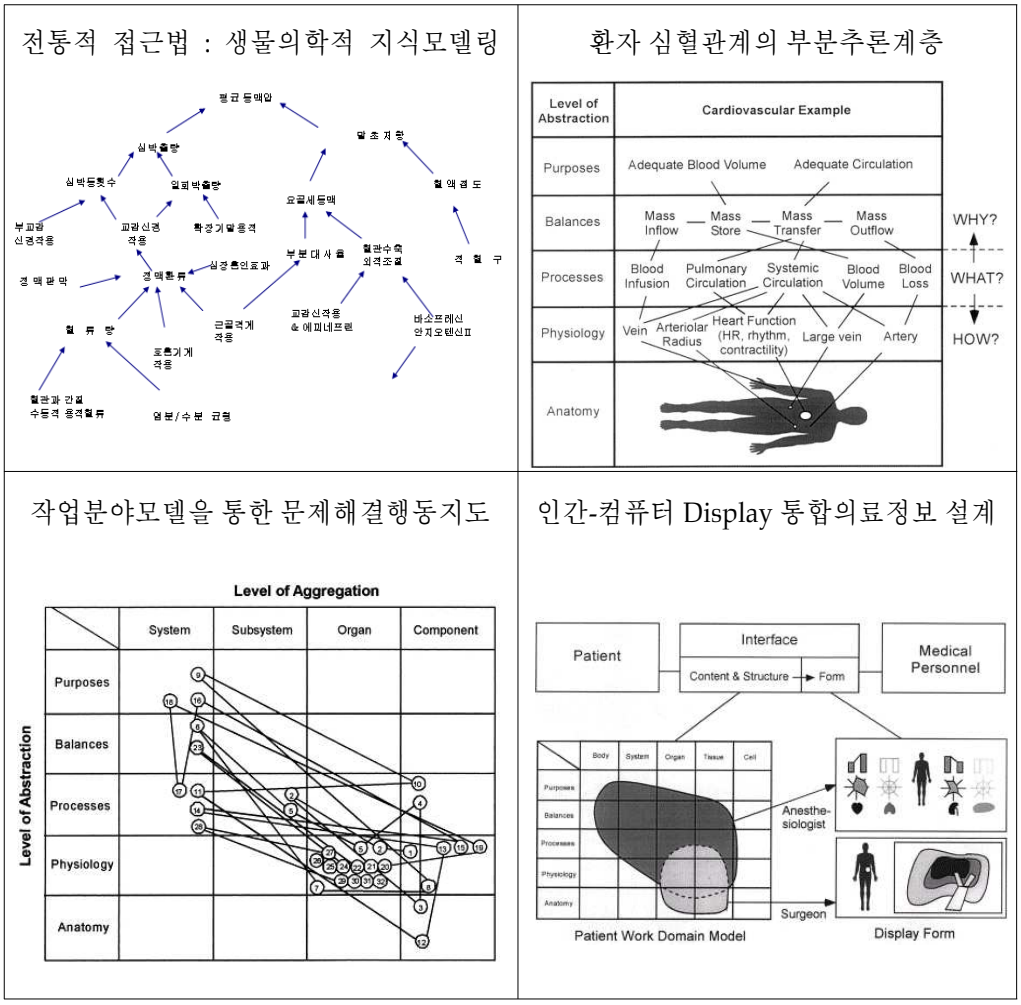


그림 8. 의료환경 모델링 : 통합의료정보 설계를 위한 온톨로지

질문을 클래스 단위로 구분하여 3가지 대표 클래스에 대하여 각각의 계층을 정의하여 대표 클래스에 따른 서브클래스를 결정하였다(그림 9-11).

① 질문 상황

그림 9. 질문상황 클래스 계층

② 질문 주제

그림 10. 질문주제 클래스 계층

③ 질문유형

그림 11. 질문유형 클래스 계층

### 나. 클래스, 서브클래스, 슬롯, 인스턴스 입력

대표 클래스 3가지에 대하여 각각의 서브클래스를 Protege에 입력하였다. 아래 그림은 클래스와 서브클래스를 입력하는 화면이다(그림 12).

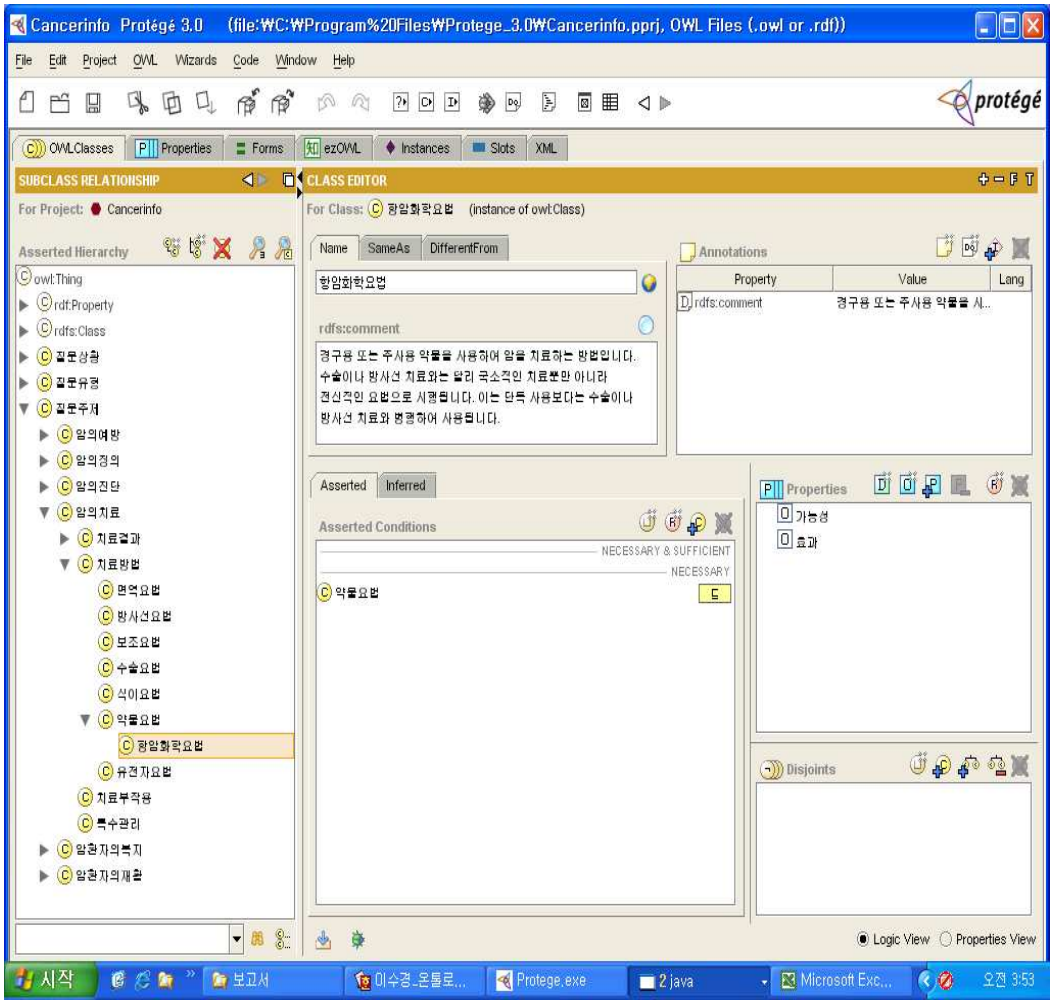


그림 12. 클래스와 서브클래스 입력 화면

클래스 입력과 동시에 필요로 되는 슬롯과 해당 인스턴스를 추가로 Protege에 입력하였다. 아래 그림 13과 그림 14는 슬롯과 인스턴스를 입력하는 화면이다.

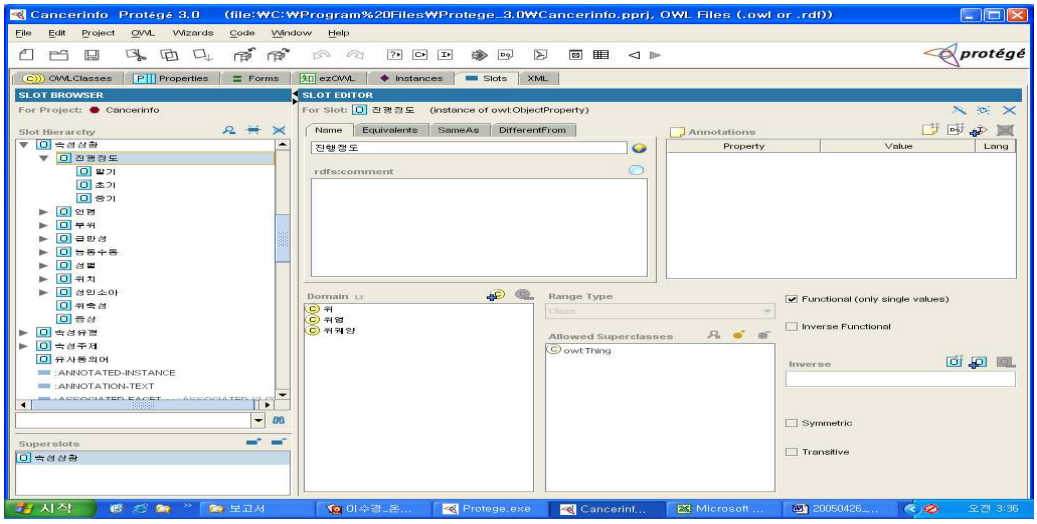


그림 13. 슬롯 입력 화면

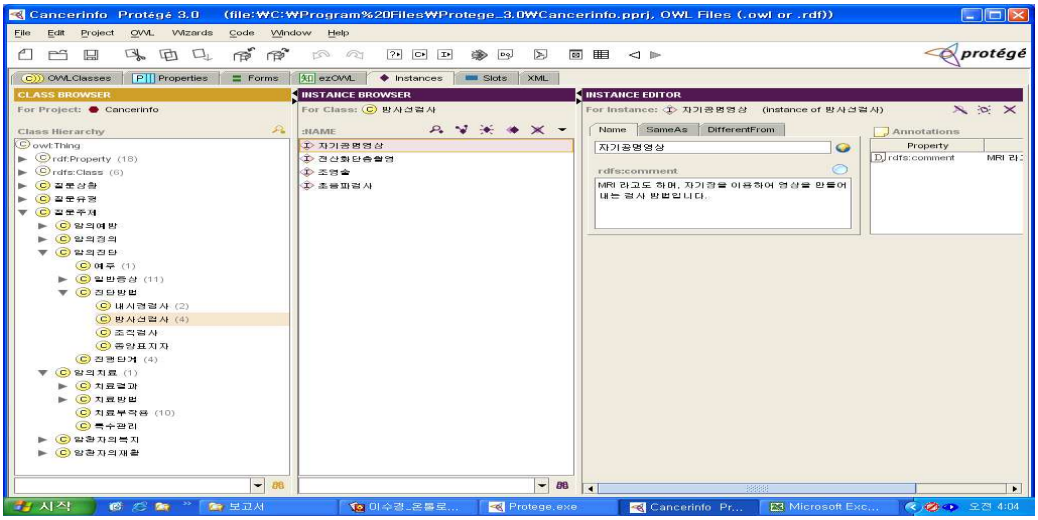


그림 14. 인스턴스 입력 화면

입력된 슬롯의 계층은 아래 그림 15와 같다.

그림 15. 슬롯 계층

### 다. 구축 온톨로지의 XML 전환

모든 항목을 입력 후에 XML Sheet로 이동하여 3가지 클래스와 인스턴스에 해당하는 총 4개의 XML 파일을 생성하였다. 아래 그림 16은 생성된 XML 화면이다.

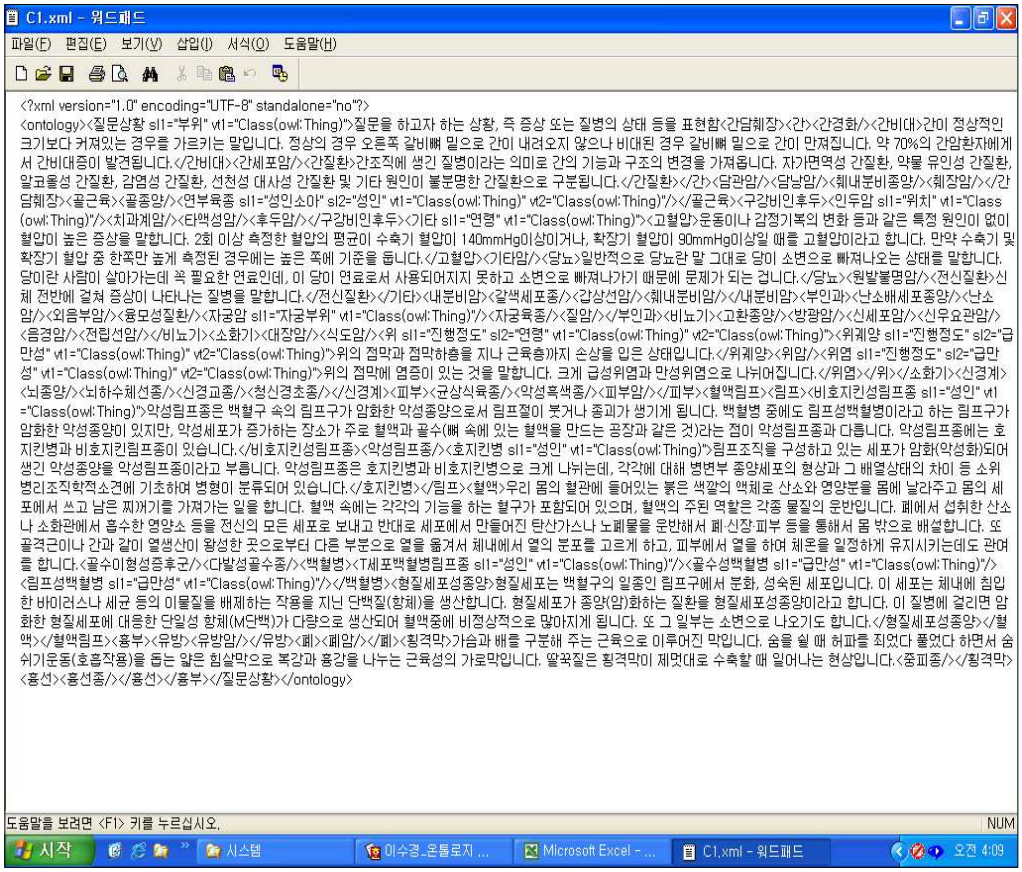


그림 16. XML 전환 화면

## 3. 온톨로지 기반 암정보 제공 대화시스템 개발

### 가. 시스템 구성

위암정보 제공을 위해 온톨로지 기반 대화시스템을 만들기 위해 구축된 해당 질문과 대답에 대한 R-DB와 질문 온톨로지는 대화 에이전트를 통해 서로 연결되어 있는 구조를 가지고 있다. 사용자는 사용자 인터페이스를 통해 질문을 입력하여 대화 에이전트에게 답변을 요구하며, 답변을 찾은 후에는 인터페이스를 통해 해당 답변이 사용자에게 보여지게 된다(그림 17).

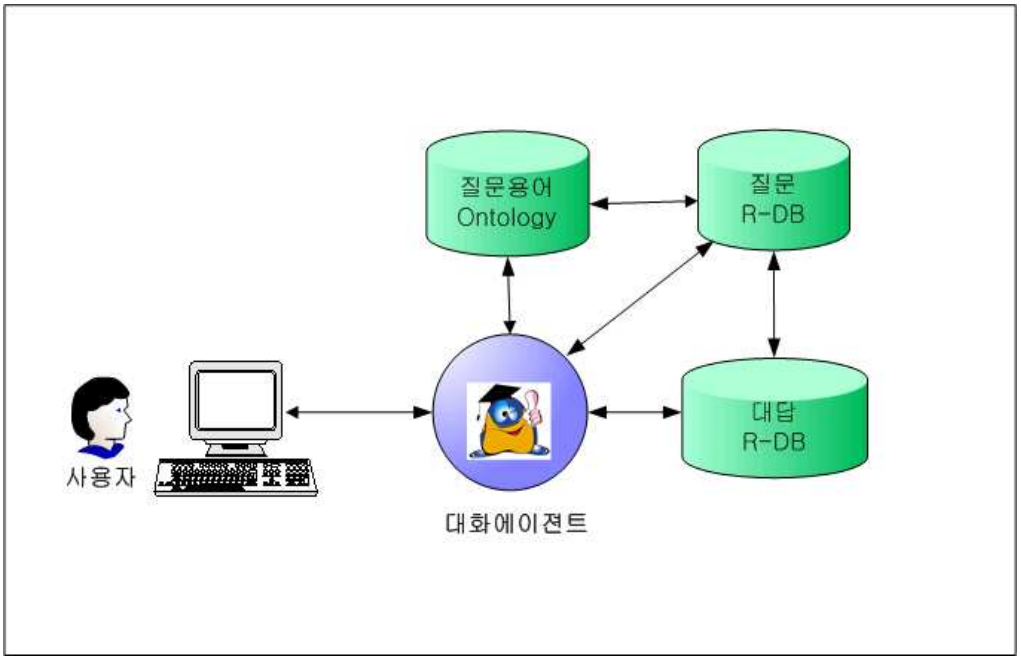


그림 17. 시스템 구성도

## 나. 시스템 흐름도

먼저 사용자가 질문을 입력하면 질문 DB에서 동일한 질문을 찾게 된다. 질문 DB에서 동일한 질문이 없다면 질문 온톨로지를 이용하여 유사한 질문을 찾는다. 출력된 유사한 질문에 대해 사용자가 답변을 보고 싶다고 선택하면 해당 답변이 출력되고, 답변을 보고 싶지 않다고 선택하면 질문을 구체적으로 재입력하여 유사 질문을 찾는 과정을 반복하게 된다(그림 18).

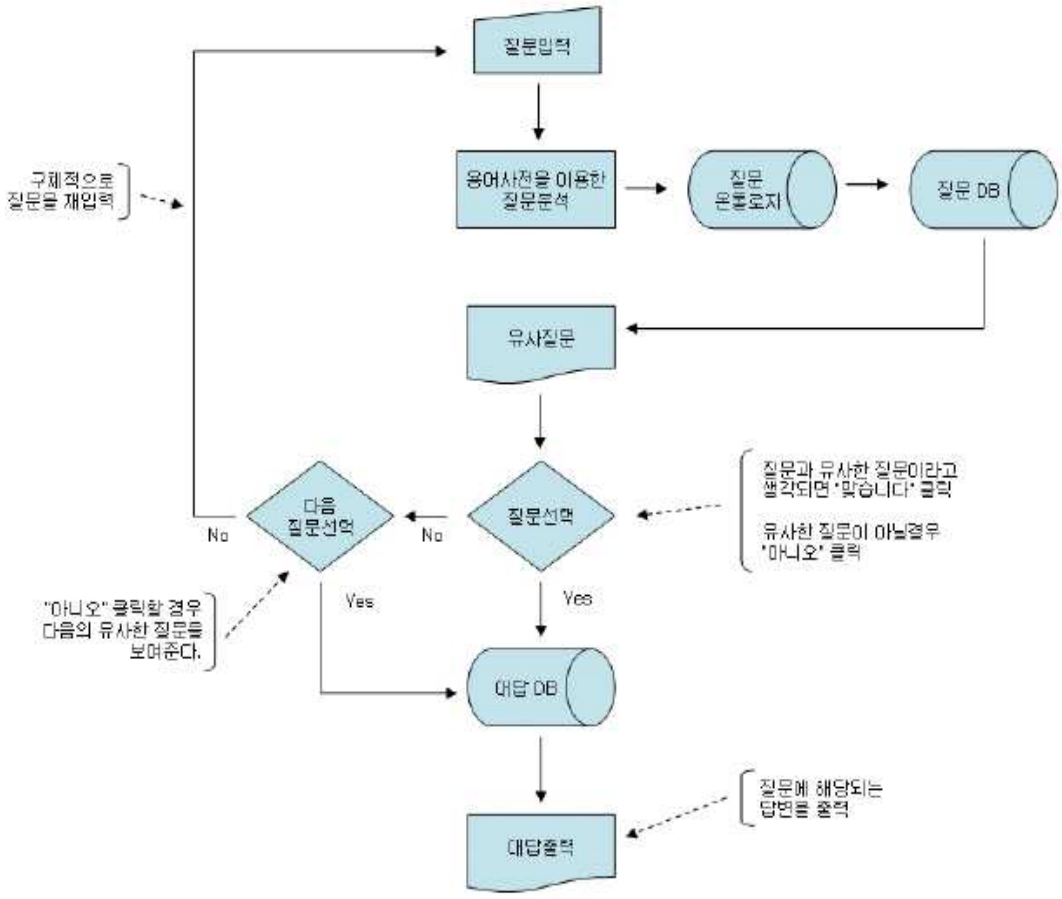


그림 18. 시스템 흐름도

### 다. 유즈케이스 다이어그램

사용자 관점에서 논리적인 시스템의 기능을 정의한 유즈케이스 다이어그램은 아래와 같다(그림 19).

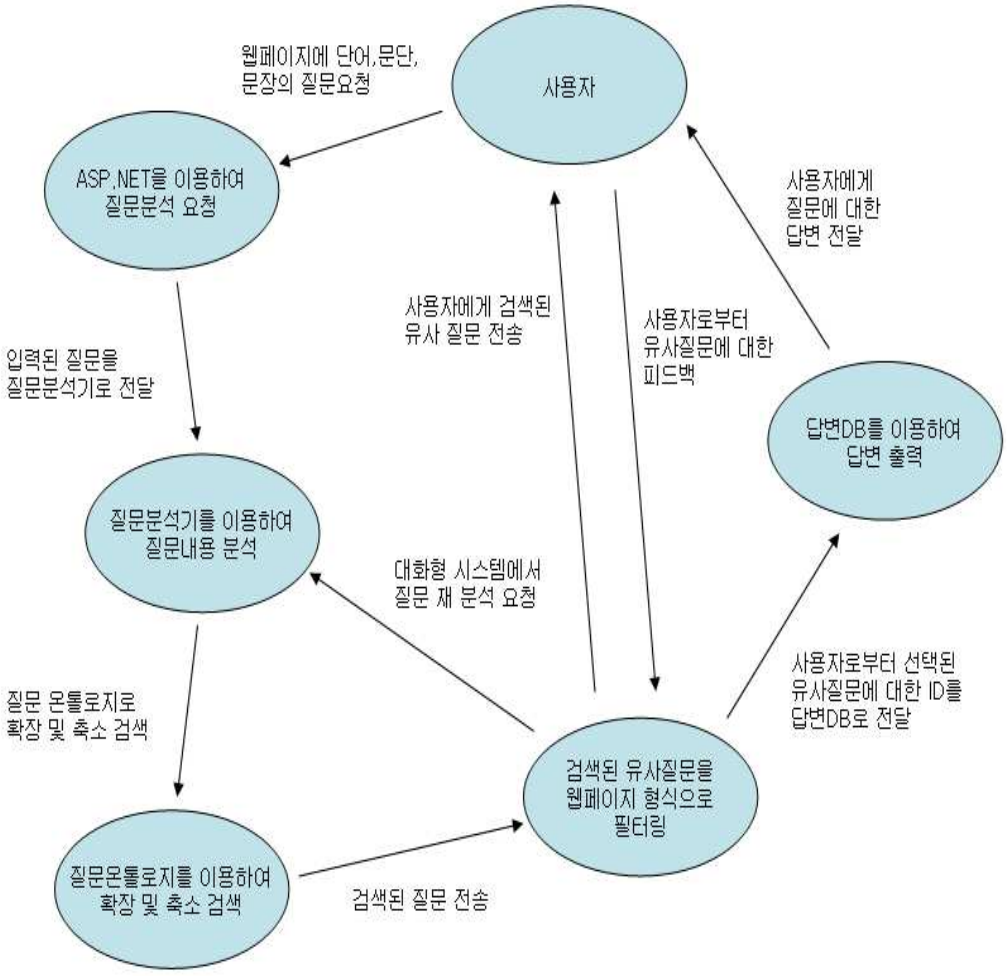


그림 19. Use Case Diagram

① 사용자가 질문을 입력하면, 질문 문장을 용어 사전을 사용하여 단어별로 분석한다.

② 분석한 문장내의 단어를 프로티지 DB를 사용하여, 상위 및 하위 개념과 유사 개념을 포함하여 온톨로지 객체를 만든다.

③ 유사 질문내에 있는 질문들을 위의 1, **2**번과 같은 방법으로 온톨로지 객체를 만든다.

④ 사용자가 입력한 온톨로지 객체와 유사질문에서 생성한 온톨로지 객체를 비교하여 관계도가 가장 높은 유사 질문들을 리스트화 한다.

⑤ 관계도가 가장 높은 유사질문리스트 내에서 프로티지DB를 사용하여 우선순위를 결정하여 사용자 인터페이스에 출력한다.

⑥ 사용자가 "맞습니다"를 클릭하면 선택된 유사질문의 답변 아이디를 사용하여 답변DB에서 답변을 찾아 사용자에게 출력한다.

⑦ 사용자가 "아니요"를 클릭하면 다음 순위의 질문을 출력한다.

⑧ 준비된 유사질문이 모두 출력된 후에도 사용자가 찾는 유사질문이 없을 경우 사용자에게 좀더 구체화된 질문을 입력하도록 유도한다.

### 라. ERD (Entity Relationship Diagram)

시스템에 필요한 주요 엔티티(Entity)를 표시하고 객체가 지닌 속성간의 관계를 그림으로 표현한 ERD는 아래와 같다(그림 **20**).

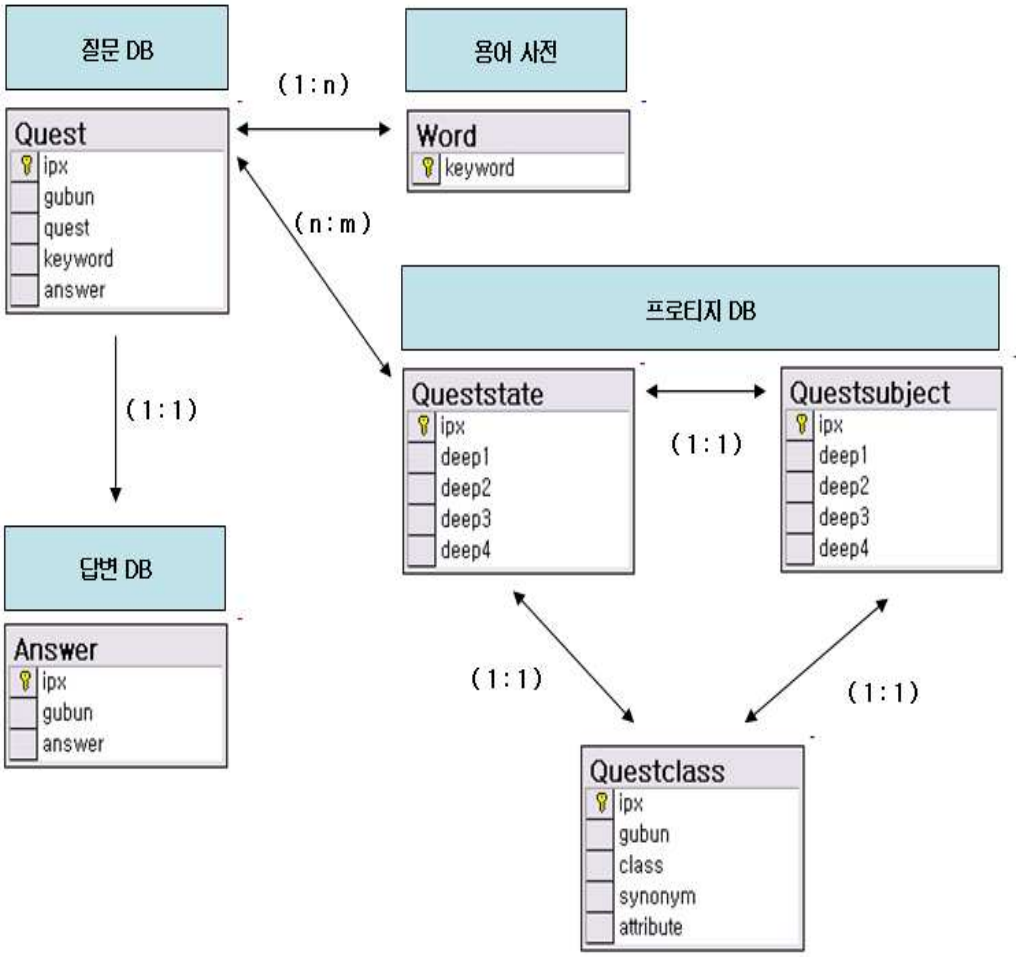


그림 **20**. ERD(개체관계도)

다음은 ERD를 구성하고 있는 테이블의 상세 명세서이다(표 12).

표 12. 테이블 명세서

| DB구분 | 테이블명 | 필드명 | 설명 |
|---|---|---|---|
| 질문 DB | Quest | ipx | 질문 ID |
| | | gubun | 질문 출처 |
| | | quest | 질문 내용 |
| | | key word | 키워드 |
| | | answer | 답변 번호 |
| 답변 DB | Answer | ipx | 답변 ID |
| | | gubun | 답변 출처 |
| | | answer | 답변 내용 |
| 용어사전 | Word | key word | 키워드 |
| 프로티지 | Quest_state | ipx | 질문 상황 ID |
| | (질문 상황) | deep 1 | 클래스 |
| | | deep 2 | 서브클래스 |
| | | deep 3 | 하위 서브클래스 |
| | | deep 4 | 인스턴스 |
| | Quest_subject | ipx | 질문 주제 ID |
| | (질문 주제) | deep 1 | 클래스 |
| | | deep 2 | 서브클래스 |
| | | deep 3 | 하위 서브클래스 |
| | | deep 4 | 인스턴스 |
| | Quest_class | ipx | 질문 유형 ID |
| | (질문 유형) | gubun | 질문 유형 분류 |
| | | class | 클래스 |
| | | synonym | 동의어/유사어 |
| | | attribute | 속성 |

### 마. 시스템 화면

① 메인 화면

다음은 CancerQ시스템의 메인화면으로 직접 웹에서 구현된다(URL : http://cancerq.ezzin.com). 질문 입력에 대한 간단한 설명과 대화창을 컨트롤하기 위한 보내기 버튼과 새로시작 버튼이 있다(그림 21).

그림 21. 메인 화면

② 질문 입력 화면

다음은 사용자가 위암 관련 질문을 입력한 후 보내기 버튼을 누르게 되면 CancerQ가 유사한 질문을 보여주고, 유사한 질문에 대한 대답 출력 여부를 선택하도록 설명하는 화면이다(그림 **22**).

그림 **22**. 질문 입력 화면

③ 질문 재입력 화면

다음은 사용자가 CancerQ가 보여준 유사한 질문에 대한 대답을 보고 싶지 않다고 '아니오'를 클릭하게 되면 유사질문을 계속적으로 보여준다. 결국 사용자가 원하는 유사질문을 찾지 못한 경우에 질문을 구체적으로 재입력하는 화면이다(그림 23).

그림 23. 질문 재입력 화면

④ 답변 출력 화면

다음은 사용자가 CancerQ가 보여준 유사한 질문에 대한 대답을 보겠다고 '맞습니다'를 클릭하게 되면 하이퍼링크 되어 해당 답변이 보여지는 화면이다. 입력한 질문과 선택한 질문, 답변이 함께 보여진다(그림 **24**).

그림 **24**. 답변 출력 화면

⑤ 시스템 관리자 화면

다음은 시스템 관리자가 현재 입력된 질문과 답변 외에 추가적으로 질문과 답변을 등록하는 화면이다. 즉, 관리자 모드를 통해 추가하고자 하는 질문과 답변을 모두 등록한다(그림 **25**).

그림 **25**. 시스템 관리자 화면

## 4. 시스템 평가 및 활용방안 제시

### 가. 시스템 성능 평가

기존 웹사이트와 온톨로지 기반 검색시스템의 Recall과 Precision, F-measure를 전문가인 암관련 전문의사와 본 연구자가 평가를 실시하였다. 기존 웹사이트는 국가암정보센터의 자주묻는질문 화면(http://www.cancer.go.kr)에서, 온톨로지 시스템은 새로 개발한 화면(http://cancerq.ezzin.com)에서 **2**회 검색을 시행하여 Recall과 Precision을 측정하였다(표 13).

표 13. Recall과 Precision 값 측정

| 구 분 | 해당 적합 문서수(A) | 검색된 전체 문서수( B ) | 검색된 적합 문서수( C ) | Recall ( C/A) | Precision ( C/B) |
|---|---|---|---|---|---|
| 암관련 전문의사 | | | | | |
| - 기존 웹사이트 | | | | | |
| 1 회 | **6** | **5** | 3 | **0.50** | **0.60** |
| **2** 회 | **6** | 3 | 3 | **0.50** | 1.00 |
| - 온톨로지 시스템 | | | | | |
| 1 회 | **6** | **6** | **6** | 1.00 | 1.00 |
| **2** 회 | **6** | **6** | **6** | 1.00 | 1.00 |
| 연구자 | | | | | |
| - 기존 웹사이트 | | | | | |
| 1 회 | **2** | **6** | **2** | 1.00 | 0.33 |
| **2** 회 | **2** | 4 | **2** | 1.00 | **0.50** |
| - 온톨로지 시스템 | | | | | |
| 1 회 | **2** | 3 | 1 | **0.50** | 0.33 |
| **2** 회 | 3 | 3 | **2** | **0.67** | **0.67** |

암관련 전문의사의 평가시 온톨로지 시스템의 Recall과 Precision 값이 기존 웹사이트 검색시의 값보다 각각 평균 **0.5**, 평균 **0.2** 높았다. 이는 기존 웹사이트 검색에 비해 온톨로지 시스템이 적합한 자료를 더 많이, 더 정확하게 검색한다는 것을 의미한다. 연구자 평가시 Recall 값은 기존 웹사이트 검색에서 평균 **0.42** 높았으며, Precision 값은 온톨로지 시스템에서 평균 **0.08** 더 높았다. 이는 기존 시스템이 온톨로지 시스템에 비해 적합한 자료를 더 많이 검색해 내지만 부적합한 자료를 검색하지 않는 검색의 정확성은 떨어진다고 할 수 있다.

시스템의 성능 평가를 위해 F-measure 값 측정시 평가자의 주관에 따라 Recall과 Precision 값의 중요도를 고려하여 가중치인 $\beta$ 값을 결정하여야 한다(표 14).

표 14. Recall, Precision, F-measure 측정시 참고사항

| 구 분 | 내 용 |
|---|---|
| $\beta$ | - Recall과 Precision의 중요도에 따른 가중치<br>$\beta$ = 1 일 경우 값의 중요도는 Recall = Precision<br>$\beta$ < 1 일 경우 값의 중요도는 Recall < Precision<br>$\beta$ > 1 일 경우 값의 중요도는 Recall > Precision |
| Recall | - 전체자료 중에서 사용자 질의결과 검색된 자료의 비율<br>- 적합한 자료를 얼마나 많이 검색했느냐를 나타냄<br>- 적합한 자료를 찾아내는 능력 |
| Precision | - 검색한 자료 중에서 적합한 자료의 비율<br>- 검색 정확성을 측정하는 기준<br>- 부적합한 자료를 검색하지 않는 능력 |
| 재현율과 정확률의 관계 | - 재현율과 정확률은 반비례 관계에 있기 때문에 검색한 자료 숫자가 많아지면 많아질수록 적합 문헌은 많이 검색되지만, 반대로 부적합 문헌도 많이 포함되므로 정확률이 떨어지게 된다. |

기존 웹사이트와 온톨로지 대화시스템의 F-measure 측정 결과는 표 15와 같다. 암관련 전문의사의 시스템 성능 평가시 Precision 값의 중요도에 가중치를 두어 $\beta$값을 0.5로 적용하였으며, F-measure 값이 온톨로지 시스템은 1.0으로 기존 웹사이트 검색의 0.71 보다 높았다.

연구자의 시스템 성능 평가시 이 연구에서는 시스템의 정확한 검색이 더 중요하다고 판단되어 $\beta$값을 0.5로 적용한 결과 F-measure 값이 기존 웹사이트 검색은 평균 0.47, 온톨로지 시스템은 평균 0.51로 0.04 더 높았다. 따라서 기존 웹사이트에 비해 온톨로지 시스템에서의 검색 성능이 더 뛰어나다고 할 수 있다.

표 15. F-measure 값 측정

| 구 분 | Recall (R) | Precision (P) | 가중치 ($\beta$) | F - measure $(\beta^2+1)PR/\beta^2P+R$ | 평균 |
|---|---|---|---|---|---|
| 암관련 전문의사 | | | | | |
| - 기존 웹사이트 | | | | | |
| 1 회 | 0.50 | 0.60 | 0.50 | 0.58 | 0.71 |
| 2 회 | 0.50 | 1.00 | 0.50 | 0.83 | |
| - 온톨로지 시스템 | | | | | |
| 1 회 | 1.00 | 1.00 | 0.50 | 1.00 | 1.00 |
| 2 회 | 1.00 | 1.00 | 0.50 | 1.00 | |
| 연구자 | | | | | |
| - 기존 웹사이트 | | | | | |
| 1 회 | 1.00 | 0.33 | 0.50 | 0.38 | 0.47 |
| 2 회 | 1.00 | 0.50 | 0.50 | 0.56 | |
| - 온톨로지 시스템 | | | | | |
| 1 회 | 0.50 | 0.33 | 0.50 | 0.36 | 0.51 |
| 2 회 | 0.67 | 0.67 | 0.50 | 0.67 | |

### 나. 시스템 사용성 평가

**2005**년 **6**월 7일부터 **16**일까지 총 **10**일간 온톨로지 기반 대화시스템의 사용성 평가를 위해 일반인과 의료인 총 **36**명을 대상으로 실제 시스템을 사용하도록 한 후 평가 결과를 평가서에 직접 기록하도록 하였다(부록 1).

대상자들은 각기 자유롭게 정한 위암 관련 질문 **5**가지를 기존 웹사이트와 개발 시스템에서 각각 검색함으로써 원하는 답변을 검색하는데 걸리는 시간과 검색 단계를 각각 측정하였으며, 제공된 답변에 대한 만족도를 **5**점 척도로 평가하는 방법을 사용하였다. 평가의 범주와 문항 및 측정방법은 표 **16**과 같다.

표 **16**. 평가 문항과 측정방법

| 평가 범주와 문항 | | 측정방법/측정척도 |
|---|---|---|
| 일반 사항 | 연 령 | 숫자로 직접 기록 |
| | 성 별 | 남, 여 中 선택 |
| | 직 종 | 일반인과 의료인 中 선택 |
| 위암 관련 질문 | 검색 질문 | 번호 **1~5**까지 검색질문을 직접 기록 |
| 시스템 사용 결과† | 검색 소요 시간 | 검색하는데 소요된 시간을 분단위로 기록 |
| | 검색 단계(Stage) | 검색 단계(Stage)를 횟수로 기록 |
| | 검색결과 만족도 | **5**점 척도<br>(매우 만족, 만족, 보통, 불만, 매우 불만) |

(†시스템 사용 결과는 **5**가지 질문을 기존 웹사이트와 개발 시스템에 각각 검색)

1) 대상자의 일반적 특성

평가에 참여한 대상자는 총 36명이었으며 이중 일반인은 20명(55.6%), 의료인은 16명(44.4%)이었다. 대상자들의 연령은 주로 30대였으며, 남자가 9명(25%)에 비해 여자는 27명(75%)으로 3배 많았다. 일반인과 의료인 집단 간 일반적 특성의 동질성을 분석하기 위해 $X^2$검정을 통해 비교해본 결과 통계적으로 유의한 차이를 보이지 않았다(표 17).

표 17. 일반적 특성 비교

| 특성 | 구분 | 일반인 | 의료인 | 총계 | $X^2$ | p값 |
|---|---|---|---|---|---|---|
| | | 명 (%) | 명 (%) | 명 (%) | | |
| 연령 | 30세 이하 | 7 (19.4) | 3 ( 8.3) | 10 (27.8) | 3.69 | 0.297 |
| | 31 ~ 35세 | 7 (19.4) | 4 (11.1) | 11 (30.6) | | |
| | 36 ~ 40세 | 4 (11.1) | 8 (22.2) | 12 (33.3) | | |
| | 41세 이상 | 2 ( 5.6) | 1 ( 2.8) | 3 ( 8.3) | | |
| 성별 | 남 | 8 (22.2) | 1 ( 2.8) | 9 (25.0) | 3.75 | 0.053 |
| | 여 | 12 (33.4) | 15 (41.6) | 27 (75.0) | | |
| 합 | 계 | 20 (55.6) | 16 (44.4) | 36 (100.0) | | |

2) 평가 질문 유형

위암에 관한 질문은 대상자 1인이 각 5가지의 질문을 기재하도록 하였으며, 질문은 총 180문항이었다. 치료 및 예후에 관한 질문이 53문항(29.4%)으로 가장 많았으며, 원인에 관한 질문이 31문항(17.2%)으로 많았으며, 나머지 질문 유형은 증상, 진단 및 검사, 질병 정의, 식이, 예방에 관한 질문이었다. 일반인과 의료인 집단간 질문 유형의 동질성을 분석하기 위해 $X^2$검정을 통해 비교해본 결과 통계적으로 유의한 차이를 보이지 않았다(표 18).

표 18. 위암 관련 질문 유형

| 질문 유형 | 일반인 | 의료인 | 전체 | $X^2$ | p값 |
|---|---|---|---|---|---|
| | 질문수 (%) | 질문수 (%) | 질문수 (%) | | |
| 치료 및 예후 | 26 (14.4) | 27 (15.0) | 53 (29.4) | 7.4 | 0.29 |
| 원 인 | 21 (11.7) | 10 ( 5.6) | 31 (17.2) | | |
| 증 상 | 14 ( 7.8) | 11 ( 6.1) | 25 (13.9) | | |
| 진단 및 검사 | 7 ( 3.9) | 13 ( 7.2) | 20 (11.1) | | |
| 질병 정의 | 13 ( 7.2) | 7 ( 3.9) | 20 (11.1) | | |
| 식 이 | 12 ( 6.7) | 7 ( 3.9) | 19 (10.6) | | |
| 예 방 | 7 ( 3.9) | 5 ( 2.8) | 12 ( 6.7) | | |
| 합 계 | 100 (55.6) | 80 (44.4) | 180 (100.0) | | |

3) 검색소요시간 및 검색단계 비교

기존 웹사이트와 온톨로지 대화시스템에서의 검색소요시간과 검색단계를 비교한 결과는 표 19와 같다. 검색 소요 시간을 살펴보면 기존 웹사이트 에 비해 온톨로지 시스템이 빠르게 검색되었는데, 일반인은 61.9초, 의료인은 45.6초의 차이가 났으며, 두 집단 모두 통계적으로 유의하였다(p=0.000, p=0.003). 검색 단계는 일반인의 경우 기존 웹사이트는 2.87회, 온톨로지 시스템은 1.98회로 통계적으로 유의한 차이를 보였으나(p=0.001), 의료인의 경우 기존 웹사이트와 온톨로지 시스템에서 통계적으로 유의한 차이를 보이지 않았다(p=0.253). 검색이 되지 않은 질문 13개 항목은 제외하였다.

표 19. 검색소요시간 및 검색단계 비교 (단위: 초, 회)

| 구 분 | 기존 웹사이트 검색 | 온톨로지 대화시스템 검색 | t값 | p값 |
|---|---|---|---|---|
| | 평균 ± 표준편차 | 평균 ± 표준편차 | | |
| 검색소요시간 | | | | |
| - 일반인 | 116.4 ± 142.4 | 54.5 ± 50.9 | 3.87 | 0.000* |
| - 의료인 | 125.6 ± 100.5 | 80.0 ± 82.4 | 3.05 | 0.003* |
| 검색 단계 | | | | |
| - 일반인 | 2.87 ± 2.22 | 1.98 ± 1.49 | 3.26 | 0.001* |
| - 의료인 | 2.74 ± 1.37 | 2.38 ± 2.30 | 1.15 | 0.253 |

* : p<0.05

질병에 대한 지식을 기본적으로 가지고 있는 의료인과는 달리 일반인의 경우 온톨로지 시스템에 비해 기존 웹사이트에서 원하는 정보를 찾는데 시간이 더 소요되고, 더 많은 횟수를 검색하게 되는 것으로 생각된다.

4) 검색 결과 만족도 비교

① 직종별 결과 만족도 비교

직종별 검색 결과에 대한 만족도를 살펴보면 표 20과 같다. 기존 웹사이트에서는 일반인은 44문항(24.4%)에 대해 만족하였고, 39문항(21.7%)에 대해 불만을 표시하였으며, 의료인은 45문항(25.0%)에 대해 만족하였고, 14문항(7.8%)에 대해 불만을 표시하였다. 기존 웹사이트에서의 두 집단간 만족도가 통계적으로 유의한 차이를 보였다(p=0.006). 반면에 온톨로지 대화시스템에서의 두 집단간 만족도는 통계적으로 유의한 차이를 보이지 않았다(p=0.518).

표 20. 직종별 결과 만족도 비교

| 구 분 | 일반인 | 의료인 | 전체 | $X^2$ | p값 |
|---|---|---|---|---|---|
| | 질문수 (%) | 질문수 (%) | 질문수 (%) | | |
| 기존 웹사이트 검색 | | | | | |
| - 만족 | 44 (24.4) | 45 (25.0) | 89 (49.4) | 10.1 | 0.006* |
| - 보통 | 17 ( 9.4) | 21 (11.7) | 38 (21.1) | | |
| - 불만 | 39 (21.7) | 14 ( 7.8) | 53 (29.4) | | |
| 온톨로지 대화시스템 | | | | | |
| - 만족 | 61 (33.9) | 45 (25.0) | 106 (58.9) | 1.31 | 0.518 |
| - 보통 | 18 (10.0) | 20 (11.1) | 38 (21.1) | | |
| - 불만 | 21 (11.7) | 15 ( 8.3) | 36 (20.0) | | |
| 합 계 | 100 (55.6) | 80 (44.4) | 180 (100.0) | | |

* : p<0.05

② 기존 웹사이트와 온톨로지 시스템의 검색결과 만족도 비교

사용자 검색 결과에 대한 만족도를 살펴보면 표 21과 같다. 검색 결과에 대한 만족도는 리커드(Likert) 5점 척도를 점수화하여 분석하였다(매우 만족=5점, 만족=4점, 보통=3점, 불만=2점, 매우 불만=1점).

일반인의 검색 결과에 대한 만족도는 기존 웹사이트보다 온톨로지 시스템에서 평균 0.67점 높았으며 이는 통계적으로 유의하였다(p=0.000). 반면에 의료인의 검색 결과에 대한 만족도는 기존 웹사이트보다 온톨로지 시스템에서 3.45점과 3.49점으로 통계적으로 유의한 차이를 보이지 않았다(p=0.808).

표 21. 검색 결과 만족도 비교

| 구 분 | 기존 웹사이트 검색 | 온톨로지 대화시스템 | t값 | p값 |
|---|---|---|---|---|
| | 평균 ± 표준편차 | 평균 ± 표준편차 | | |
| 일반인 | 2.95 ± 1.45 | 3.62 ± 1.26 | -3.48 | 0.000* |
| 의료인 | 3.45 ± 0.94 | 3.49 ± 1.00 | -0.24 | 0.808 |

* : p<0.05

질병에 대한 지식이 상대적으로 부족한 일반인의 경우 정보를 찾기 어려운 기존 웹사이트에 비해 온톨로지 대화시스템의 검색 결과에 더 높은 만족도를 표시하였다고 생각된다. 따라서 온톨로지 시스템은 전문가보다는 일반인들에게 더 유용하다고 할 수 있다.

### 다. 시스템 활용방안 제시

시스템 성능 평가 및 사용성 평가 결과를 바탕으로 온톨로지 시스템의 경우 전문가 보다는 일반인을 대상으로 전문 지식을 제공할 때 더 효과적이라고 할 수 있다.

국가암정보센터 웹사이트에서 키워드 검색, 내용 검색, 자료별 검색, 용어 사전 등 다양한 방법들을 통해 암정보를 제공하고 있다. 따라서 이 연구의 제한점이 보완되고 보다 발전된 온톨로지 기반의 대화형 에이전트를 국가암정보센터 웹사이트에 추가하여 활용한다면 일반인들에게 보다 친숙하고 빠르게 암정보를 제공할 수 있을 것이다(그림 **26**).

그림 **26**. 국가암정보센터 웹사이트† 에 온톨로지 대화시스템 추가 화면

(† Available from URL: http://www.cancer.go.kr/ Accessed June **20**, **2005**)

# V. 고찰

이 연구는 온톨로지를 이용하여 원하는 정보를 보다 쉽고, 간편하게 제공받을 수 있도록 온톨로지 기반의 대화형 에이전트를 개발하고자 하였다. Protege를 통해 질문 온톨로지를 구축하여 사용자 질의에 대한 적절한 답변을 찾을 수 있도록 하고, 대화형 에이전트를 통해 보다 실질적으로 활용이 가능한 시스템을 구축하였다. 실제 위암정보 제공을 위한 도우미 에이전트를 구현해 봄으로써 그 가능성을 평가해 보았다.

이 연구와 관련된 선행 연구들을 살펴보면 다음과 같다. 암정보 제공에 대한 분석 연구(김추봉, **2002**; 강건욱, **2003**)들이 최근 국립암센터를 중심으로 국가적인 연구가 진행 중이다. 선진 외국의 경우 근거에 기반한 표준화된 내용과 형식으로 암정보 제공 시스템을 통해 사용자 친화적으로 제공하고 있다. 미국은 국립암연구소(NCI, National Cancer Center Institute)에서 국가적인 차원에서 암정보 관리를 하고 있으며, CancerNet이라는 포탈사이트를 운영하여 전문가나 일반인들을 위한 광범위한 과학적 지식에 근거한 표준화된 암관련 정보를 제공하고 있다. 캐나다와 영국의 경우도 미국 NCI의 자료를 활용하여 암정보를 국가적으로 제공·관리하고 있으며, 일본의 경우 의학영상참고자료센터를 구축하여 표준화된 암관련 영상 정보를 제공하고 있다(강건욱, **2003**).

국내에서는 전문 분야별 온톨로지 구축 연구가 최근 진행되고 있다. Protege를 이용한 온톨로지 구축에 관한 연구들로는 건축용어 및 건설정보

온톨로지 구축(김학래 등, 2004; 윤진혁, 2003), 온톨로지 기반 지능형 역사유물 검색시스템 개발(이진중, 2003)이 있으며, 의학 분야에서는 한의학 증상 온톨로지 개발(박경모, 2003; 임희숙 등, 2002), 병리 온톨로지 구축(하병현, 2003; 이재일, 2004) 등이 있다. 이 외에도 간호학에서의 온톨로지 모델링(박현애, 2004), UMLS 의학용어 연구(한승빈, 2001), 호흡기 내과 한글주소 표현에 관한 연구(김미정, 2004) 등이 있다. 해외의 경우는 온톨로지 개발 방법론(Noy, 2001; Sure, 2002; Webster, 2001)에 관한 연구와 의학적 의사결정 시스템으로서의 적용 연구 - 비뇨기계 감염(Karlsson, 2001), 수술실 분야(Hajdukiewicz, 2001), 염증 온톨로지(Pisanellia, 2004) - 등에 대한 연구가 이미 진행되고 있다.

온톨로지를 이용한 응용 연구들을 살펴보면, 국내의 경우 기업간 전자상거래 표현과 지식관리(조영희, 2002), 의류 사이트 안내 도우미로써의 대화형 에이전트 구현(김경민, 2002) 등이 있으며, 의료 분야에서는 의료 에이전트 시스템 활성화 고찰(김민철, 2003), 한의학 처방 지식관리시스템 개발(이현실, 2003), 한의학 원격문진시스템 개발(김선호, 2004) 등이 있다. 해외 연구로는 미국의 보건의료분야의 최대 기업인 Kaiser사의 의학용어 개발(Dolin, 2004), Ontology-Based Dialogue Systems(Milward et al, 2003) 등이 있다.

온톨로지 시스템의 평가에 관한 연구(Aitken, 2000; Gruber, 1993; Gruber, 1995; Campbell, 2001; Guarino, 2002; Doan, 2003; 양정진, 2003; Pisanellia, 2004 )들도 매우 활발하게 진행되고 있으나, 국내의 경우는 시스템 개발이 실험적인 수준에 그치고 있어서 아직까지 정확한 평가 기준을 마련하기에는 어려운 실정이다(최호섭, 2004).

위에서 제시한 연구들 중에 Ontology-Based Dialogue Systems(Milward et

al, 2003) 연구는 서로 다른 2가지 영역에 대해 일반적 대화시스템과 영역 지식 결합을 위해 일반의를 대상으로 유방암이 의심되는 환자를 전문의에게 의뢰하기 위한 시스템인 Breast cancer referrals과 가정용 기구에 대한 제어 시스템인 Networked home appliances에서의 온톨로지 사용을 비교하였다. 온톨로지 기반 환자-의사 대화시스템에 관한 연구(김선호, 2004)에서는 환자가 직접 입력한 증상을 통해 의사가 한의학적 진단을 내리고 필요한 정보나 권고사항을 제시하는 원격진료시스템을 시험적으로 구현하였다. 미국 국립암연구소(NCI)는 MEGD 온톨로지를 제공하고 있는데, 이것은 생물정보(Bio-informatics)를 위한 유전자 발현 데이터의 교환과 공유를 위한 표준화를 위해 개발되었으며, 대략 17,000개의 암관련 용어 및 그 정의와 Manchester 대학에서 개발된 잘 알려져 있는 GALEN 의학 온톨로지의 OWL 버전을 포함하고 있다(Golbeck, 2003).

국내에서는 실질적인 온톨로지가 소규모 연구실에서의 실험적인 수준으로 구축되거나 이론적 수준에 그치고 있는 실정이다(최호섭, 2004). 온톨로지의 필요성과 그 가치를 높이기 위해서는 실제적인 활용이 중요한 만큼 이러한 연구 시도를 통해서 향후 관련 연구가 활성화되기를 기대한다.

# VI. 결론

이 연구는 위암정보에 대한 전문가의 지식을 수집하여 분석하고, 정보 제공을 효율적으로 하기 위해 질문 온톨로지를 구축하여 대화형 에이전트를 개발하였다.

국내의 병원, 협회, 업체에서 운영하고 있는 암정보 관련 웹사이트에서 위암정보 FAQ 자료를 수집하여 분석하였다. 1차 분석된 자료를 질문에 해당하는 자료에 대해서 Protege를 이용하여 온톨로지를 구축하였다. 질문 자료에 해당되는 답변 자료에 대해서는 일반적인 관계형 Database로 구축하였다. 이러한 자료를 이용하여 대화형 에이전트와 연결하여 시스템을 구축하였다. 개발된 시스템에 대하여 암관련 전문의사인 전문가를 통한 시스템 성능평가와 36명을 대상으로 사용자 중심의 사용성 평가를 시행하였다.

이 연구의 제한점은 다음과 같다. 첫째, 위암 정보의 수집 및 분석시 정보의 신뢰성을 위해서 기존 연구에서 확인된 내용 위주로 제한하다보니 보다 많은 질문과 답변 사례가 충분하지 않았다는 것이다. 둘째, Protege를 이용하여 질문 온톨로지만 단순하게 구축하였고, 답변에 대한 온톨로지는 구축하지 못한 점에서 부분적인 온톨로지 기반이라는 제한점이 있다. 셋째, 완전한 자연어 처리를 통한 유연한 대화 시스템이 아닌 선택을 통한 정형화된 대화를 진행하는 기술난이도 측면에서 용이한 시스템을 구축하였다. 넷째, 시스템의 평가 부분에서 검색 성능과 사용성 평가만 시행하였으며, 온톨로지 자체에 대한 객관적인 평가를 실시하지 못하였다.

이 시스템의 자체 검색기능 및 성능은 DB의 디자인과 저장된 자료에 달려있다고 할 수 있다. 즉, 얼마나 많은 용어들이 저장되었는지, 얼마나 많은 관계들이 저장되어 있는지에 따라 그 검색효과가 달라지기 때문에 향후 현재의 온톨로지를 보완하여 신뢰성 있는 다양한 암정보를 추가하는 것이 필수적이다. 또한 암정보 관련 전문가들을 통해서 구축 온톨로지의 분류체계에 대한 검증을 통해 내용을 보완하는 것이 필요하며, 답변 부분의 온톨로지를 추가로 구축하고, 질문 유형 클래스인 어휘집을 충분히 활용하고, 보다 구체적인 관계 설정을 통해 다양한 질문을 이해하고 구분해 낼 수 있도록 해야 한다. 위에서 언급했던 연구의 제한점이 보완된다면 질문분석 온톨로지는 누구든지 손쉽게 재사용이 가능하며, Protege를 통해 표준 용어와 다른 용어간의 관계를 설정하여 확장 또는 수정이 가능하며, 다양한 Library Program을 이용해서 처리하기가 용이할 것이다.

의료분야의 지식은 대용량, 복잡한 특성 때문에 지식을 체계화하기 매우 어렵기 때문에 다른 어느 분야보다도 온톨로지의 필요성이 강조된다. 현재 표현하기 힘든 지식을 체계적으로 구축할 수 있는 열쇠가 되기 때문이다. 현재 개발되어 있는 전문적인 온톨로지 구축도구를 국내 특성에 맞고, 또한 사용자가 쉽게 개념을 표현할 수 있도록 하는 상용화된 도구를 개발하다면 고부가가치를 달성할 수 있고, 의료 지식 컨텐츠를 다양한 e-business 분야의 상품으로 개발하여 활용할 수 있을 것이다. 또한 도메인 지식의 명확하고 조직화된 추론 계층 구조를 통해 예기치 않은 상황에서 필요로 되는 의사결정지원을 돕는 정보까지도 제공받을 수 있을 것이다.

최근 온톨로지의 중요성을 인식하고 보건복지부에서 의과학 지식 및 온톨로지 관리기술 개발센터 MORC(Medical Ontology Research Center)를

지원하여 연구를 진행하고 있다. **2005**년 **5**월 개최된 MORC과제발표회에서 이재일은 온톨로지 기반의 통합 모델링의 목적은 결국 진단과 치료 등의 의사결정지원시스템의 개발이라고 설명하였으며, 김홍기는 각각 필요한 분야에서 다양한 형태의 지식베이스인 온톨로지를 구축하기 위한 온톨로지 획득 도구를 우선적으로 개발하여 시험 중이라고 발표하였다.

다양한 방식으로 표현된 의료정보를 온톨로지를 통해 공유하기 위해서는 정부의 도움도 필요하겠으나, 의료 정보의 공동 주인인 국민들에게 가치있는 지식과 지혜로 되돌려 주고자 하는 책임감을 바탕으로 한 의료인들의 강한 의지와 실천이 뒷받침되어야 할 것이다.

# 참 고 문 헌


강건욱. 대국민 암 관리 정보 Call-Center 구축을 위한 암정보 표준화 개발 연구. 보건복지부 보건의료기술 연구개발사업 최종보고서, 2003

구미숙, 황정희, 류근호. 온톨로지 기반의 쇼핑 사이트 고객을 위한 검색 시스템. 한국정보처리학회 추계학술발표대회 논문집 2004; 11(2): 51-54

김경민, 임성수, 조성배. 인공지능 기법을 이용한 웹 가상 도우미로서의 대화형 에이전트. 15회 한글 및 한국어 정보처리 학술대회 2002

김미정. 한글 주소(Chief complaint) 표현에 관한 연구. 연세대학교 보건대학원 석사학위논문, 2004

김민경, 조미정, 정혜진 등. 내용기반 검색기법을 이용한 인터넷 기반 유방종양 조직병리영상 검색 시스템 개발. 대한의료정보학회지 2005; 11(1): 87-96

김민철, 이동철. 온톨로지 기술을 통한 의료 에이전트 시스템 활성화 고찰. 기초과학연구 제주대학교 2003; 16(1): 125-135

김병찬. 묵시적 평가정보를 이용한 사례기반추론 추천시스템. 창원대학교 대학원 컴퓨터공학과 석사학위논문, 2001

김선호, 안충희, 박경모. 온톨로지를 이용한 한의학 원격 문진시스템의 설계와 시험적 구현. 대한의료정보학회지 2004; 10(2): 61-64

김은경. 시맨틱웹을 위한 온톨로지 구축방법에 관한 비교연구. 중앙대학교 대학원 석사학위논문, 2004

김지숙. Seed keyword를 이용한 비감독학습 대표 색인어 추출 기법. 창원대학교 대학원 컴퓨터공학과 석사학위논문, 2001

김추봉, 임민경, 김현영 등. 국내 웹싸이트 제공 암관련 상담 내용에 대한 분석 연구. 대한의료정보학회지 2002

김학래, 박의준, 김홍기 등. 통합건설정보분류체계 기반 건설정보 온톨로지 구축. 한국전자거래 학회지 2004; 9(3): 95-112

김홍기, 김명기. 의료정보학에서의 온톨로지 기술. 대한의료정보학회지 2003; 9(3): 213-219

류광택. 웹기술 발전방향 및 표준화 개발전략 연구. 정보통신부 정보통신 연구개발사업 최종보고서, 2004

박경모, 박정훈. Protage를 이용한 한의학의 구조화된 증상 입력을 위한 온톨로지 개발. 동의생리병리학회지 2003; 17(5): 1151-1156

박현애. 간호학에서의 온톨로지 모델링. 대한의료정보학회 춘계학술대회 초록집 2004; 98-104

양정진. 시맨틱 웹에서의 온톨로지 공학. 정보과학회지 2003; 21(3): 28-35

윤진혁, 유상봉, 김인한. 건축 용어의 온톨로지 구축과 활용. 한국전자거래 학회지 2003; 8(2): 49-62

이두영. 정보검색이론. 중앙대 문헌정보학과 Lecture, 2003

이재일. 의과학 지식 및 온톨로지 관리기술 개발센터. 보건복지부 특정센터 연구 지원계획서, 2004

이진중. 온톨로지 기반 지능형 검색시스템-역사유물 검색 모델링 중심으로. 원광대학교 교육대학원 석사학위논문, 2003

이현실, 이두영. 온톨로지 기반 한의학 처방 지식관리 시스템 설계에 관한

연구. 정보관리학회지 2003; 20(1): 341-371

임희숙, 박경모. 전자의무기록시스템에서의 구조화된 증상입력을 위한 인터페이스 설계 및 초보적 구현. 대한의료정보학회지 2002; 8(S1): 217-220

조영희. 온톨로지 기반의 기업간 전자상거래 표현과 지식관리. 자연과학연구 2002; 제2호: 65-87

정은경, 김영민, 변영철 등. 시맨틱 웹 환경에서의 온톨로지 기반의 정보검색. 제주대학교 기초과학연구 2003; 16(2): 111-116

최정화, 박영택. 유비쿼터스 시스템을 위한 시맨틱 다중 에이전트. 정보과학회 논문지: 소프트웨어 및 응용 2005; 32(3): 192-201

최호섭, 옥철영. 정보검색시스템과 온톨로지. 정보과학회지 2004; 22(4): 62-71

통계청. 2003년 사망원인통계결과, 2004

하병헌, 김홍기, 이재일 등. 재사용 가능한 병리 온톨로지 구축을 위한 지식 모델링 시스템 개발. 한국정보과학회 가을학술발표논문집(2) 2003; 64-66

하상범, 박영택. 온톨로지기반 추론을 이용한 시맨틱 검색시스템. 정보과학회 논문지: 소프트웨어 및 응용 2005; 32(3): 202-213

하영국. Information retrieval system for the semantic web. CS774 Term Project Report 2003

한승빈, 최진욱. UMLS(Unified Medical Language System)의 증상용어와 국내의무기록에서 사용되는 증상용어와의 비교연구. 2001

Achour SL. A UMLS-based knowledge acquisition tool for rule-based

clinical decision support system development. JAMIA 2001; 8(4)

Aitken S, Reid S. Evaluation of an ontology-based information retrieval tool. 12th ECAI Workshop on applications of ontologies and problem-solving method, 2000

Buckland M. The Trade-off between recall and precision. Journal of the American Society for Information Science 1994; 12-19

Campbell JR. Sementic features of an enterprise interface terminology for SNOMED RT. MEDINFO 2001

Corch O, Mariano FL, Asuncion GP. OntoWeb technical roadmap v1.0. IST Programme of the Commission of the European Communities as Project No. IST-2000-29243, 2000.

Doan A, Madhavan J. Learning to match ontologies on the semantic Web. The VLDB Journal 2003; 12: 303-319

Dolin RH, Mattison JE, Cohn S, et al. Kaiser permanente CMT team. Kaiser permanente's convergent medical terminology. MEDINFO 2004; 346-350

Fischer G, Ostwald J. Knowledge management: problems, promises, realities, and challenges. IEEE Intelligent System. Jan/Feb 2001; 60-72

Golbeck J, Fragoso G. The National cancer institute's sésaurus and ontology. Journal of Web Semantics 2003; 1(1)

Gruber T. A translation approach to portable ontology specifications. Knowlege Acuquisition 1993; 5(2): 199-220

Gruber T. Toward principles for the design of ontologies used for

knowledge sharing. IJHCS 1995; 43(5,6): 907-928

Guarino N, Welty C. Evaluating ontological decisions with OntoClean. Communications of the ACM 2002; 45(2): 61-65

Hajdukiewicz JR, Vicente KJ. Modeling a medical environment: an ontology for integrated medical informatics design. International Journal of Medical Informatics 2001; 62: 79-99

Karlsson D. A Design and prototype for a decision-support system in the field of urinary tract infections – application of OpenGALEN techniques for indexing medical information. MEDINFO 2001

Lewis D, Gale W. A sequential algorithm for training text classifiers. In proc of 17th annual international ACM-SIGIR Conference on Research and Development in Information Retrieval, London, Springer-Verlag. 1994; 3-12

McGuinness DL. Ontological issues for knowledge-enhanced search. In formal ontology in information systems. Proceedings of the First International Conference(FOIS'98). Trento, Italy, 1998; 302-316

Milward D, Beveridge M. Ontology-based dialogue systems. 3rd IJCAI Workshop on knowledge and reasoning in practical dialogue systems 2003

Mizoquchi R, Ikeda M. Ontology engineering - towards the basic theory and technology for content-oriented research. 人工知能學會誌 1997; 12(4): 559-569

Noy NF, McGuinness DL. 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, 2001

Pisanellia DM, Gangemia A, Battagliab M, et al. Coping with medical polysemy in the semantic web : the role of ontology. MEDINFO 2004; 416-419

Rector A, Rogers J. Ontological issues in using a description logic to represent medical concepts : Experience form GALEN. IMIA WG6 Workship. 1999

Salton G, Fox EA, Wu H. Extended boolean information retrieval. commun. ACM 1983; 26(11): 1022-1036

Sowa JF. Knowledge representation : logical philosophical and computational foundations, books/Cole Press 2000

Studer RV, Benjamins R, Fensel D. Knowledge engineering : principles and methods. Data & knowledge engineering. 1998; 25(1-2): 161-197

Sure Y, Staab S, Studer R. Methodology for development and employment of ontology based knowledge management applications. SIGMOD Record 2002; 31(4): 18-23

Webster C. A Methodology for web-enabling a computer-based patient record with contributions from cognitive science. MEDINFO 2001

국가암정보센터 자주묻는질문, Available from URL: http://www.cancer.go.kr. Accessed June 20, 2005

국립암센터 암정보서비스, Available from URL: http://www.ncc.re.kr. Accessed May 6, 2005

서울대학교 혈액종양내과 암종합정보망 FAQ, Available from URL: http://hmo.snu.ac.kr/cgi-bin/CrazyWWWBoardcgi?_db=faq. Accessed May 6, 2005

연세대학교 세브란스병원 암센터 암정보 건강정보 FAQ, Available from URL: http://www.severance.or.kr/cancer/health_info/health_faq/disease/list.asp?cat_no=13324&og_cd=15. Accessed May 6, 2005

고려대학교 의료원 종양혈액내과 FAQ, Available from URL: http://www.korea.ac.kr/~kuho/main8/main1.html. Accessed May 6, 2005

대한암협회 암예방·치료 궁금증 질문답변, Available from URL: http://kocancer.org/technote/main.cgi?command=main_htm&board=sense. Accessed May 6, 2005

암닥터 위암 FAQ, Available from URL: http://www.amdoctor.com. Accessed May 6, 2005

하이닥 사이버종합병원클리닉 위암클리닉 자주하는 질문과 답, Available from URL: http://www.hidoc.co.kr/Mic/common/faqList.jsp?clinic =3210&kP=1. Accessed May 6, 2005

서울대암연구소 일반인 게시판, Available from URL: http://cri.snu.ac.kr/public/index.asp?Action=bulletin. Accessed May 6, 2005

암이란닷컴 공개상담실 위암게시판, Available from URL: http://www.am2ran.com/g/main1.htm. Accessed May 6, 2005

코어메드 질병정보센터 암정보, Available from URL: http://www.coremed.co.kr/amdoctor. Accessed May 6, 2005

에버케어 건강정보센터 암정보, Available from URL: http://www.evercare.co.kr/mychart/inform1.jsp. Accessed May 6, 2005

원자력병원 암유전상담 클리닉, Available from URL: http://www.geneclinic.re.kr. Accessed May 6, 2005

NCI Terminology Browser, Available from URL: http://nciterms.nci.nih.gov/NCIBrowser/Dictionary.do. Accessed May 6, 2005

OWL Web Ontology Language Test Cases, Available from URL: http://www.w3.org/TR/2004/REC-owl-test-20040210/#testEntailment. Accessed May 6, 2005

< 부록 1 > 평가서

# 위암정보 제공을 위한 온톨로지 기반 대화시스템 평가

안녕하십니까? 최근 지식의 체계화와 활용에 대한 관심이 높아지면서 각 분야에서 온톨로지의 연구와 발전에 많은 노력을 기울이고 있습니다. 조달청에서는 이미 온톨로지 상품검색 시스템을 개발하여 활용함으로써 다양한 상품정보간 상호호환과 운용성을 높여 국내 e-Commerce 활성화에 기여하고 있습니다. 또한 보건복지부에서도 의과학 지식 및 온톨로지 관리기술 개발센터(Medical Ontology Research Center; MORC)를 통해 관련 연구를 진행하고 있습니다.

온톨로지란 철학의 존재론(존재하는 것들에 대한 체계적인 설명)에서 기원하였습니다. 정보학에서는 개념과 개념과의 관계를 설정하여 컴퓨터가 처리할 수 있도록 설명하는 컴퓨터와 인간과의 커뮤니케이션 기술이라고 할 수 있습니다. 초기에 지식의 공유와 재사용을 극대화하기 위한 차세대 데이터베이스로 개발되었으며, 최근에는 인공지능, 정보검색, 전자상거래, 지식관리 시스템과 같은 분야로 확대되고 있습니다.

본 평가의 목적은 기존 웹사이트와 온톨로지 대화시스템에서 각각 위암관련 정보검색을 실시하여 두가지 시스템에서의 검색 소요 시간, 검색 단계, 검색 결과 만족도 항목을 비교하기 위함입니다. 즉, 온톨로지 활용 시스템에 관한 사용자의 의견을 수렴하고 이를 토대로 향후 활용 방안을 수립하기 위한 것입니다. 본 평가에서 수집된 자료는 연구용도 이외에는 사용하지 않을 것을 약속드리며, 바쁜 시간임에도 불구하고 평가에 협조해 주신 점에 대해 깊이 감사드립니다. 평가과정이나 전·후에 의문사항이 계시면 아래 연락처로 문의하여 주십시오.

연 구 자 : 연세대학교 보건대학원 보건정보관리학과 이 수경
(Tel : 02-772-6186, Fax : 02-772-6004, e-mail : soo1005@naver.com)
지도교수 : 연세대학교 보건대학원 보건정보관리학과 채 영문교수

**[참조 1] 기존 웹사이트**

| 웹사이트명 | URL |
|---|---|
| 국립암센터 암정보서비스 FAQ | http://www.ncc.re.kr/ |
| 서울대학교 혈액종양내과 암종합정보망 | http://hmo.snu.ac.kr/ |
| 연세대학교 세브란스병원 암센터 | http://www.severance.or.kr/cancer |
| 고려대학교의료원 종양혈액내과 | http://www.korea.ac.kr/~kuho/ |
| 사단법인 대한암협회 | http://www.kocancer.org |
| 암닥터 | http://www.amdoctor.com/ |
| Hi-doc 사이버종합병원클리닉 | http://www.hidoc.co.kr/ |

**[참조 2] 개발 온톨로지 대화시스템** (URL : http://cancerq.ezzin.com/)

# 평가 기록지( No. ____ , 평가일자 2005.06.__ )

1. 괄호안에 해당 내용을 직접 기록하거나 (√) 표시하여 주시기 바랍니다.
(의료법상 의료인은 의사, 치과의사, 한의사, 간호사, 조산사에 해당됩니다.)

| | | | |
|---|---|---|---|
| 일반 사항 | (　　) 세 | 남 (　), 여 (　) | 의료인 (　　), 일반인 (　　) |

**2**. 위암과 관련된 검색 질문을 우측의 빈칸에 직접 기록하여 주십시오.

| 번 호 | 검색 질문 |
|---|---|
| 1 | |
| **2** | |
| 3 | |
| 4 | |
| **5** | |

3. 아래 각각의 항목 빈칸에 직접 기록하거나 (√)표시하여 주시기 바랍니다.

| 질문 번호 | 시스템 구분 | 검색소요 시간(분) | 검색단계 (횟수) | 검색 결과 만족도 | | | | |
|---|---|---|---|---|---|---|---|---|
| | | | | 매우 불만 | 불만 | 보통 | 만족 | 매우 만족 |
| 1 | 기존 웹사이트 검색 | | | | | | | |
| | 온톨로지 대화시스템 | | | | | | | |
| **2** | 기존 웹사이트 검색 | | | | | | | |
| | 온톨로지 대화시스템 | | | | | | | |
| 3 | 기존 웹사이트 검색 | | | | | | | |
| | 온톨로지 대화시스템 | | | | | | | |
| 4 | 기존 웹사이트 검색 | | | | | | | |
| | 온톨로지 대화시스템 | | | | | | | |
| **5** | 기존 웹사이트 검색 | | | | | | | |
| | 온톨로지 대화시스템 | | | | | | | |

**평가에 응답해주셔서 대단히 감사드립니다!**

## ABSTRACT

# Development of Ontology-Based Dialogue System for Efficient Cancer Information Supply

### Focused on the provision of Gastric Cancer information to non-experts

Soo-Kyung Lee
Dept. of Health Informations
Graduate School of Public Health
Yonsei University

(Directed by Professor Young-Moon Chae, Ph.D)

Cancer is number one killer in Korea. Koreans want to learn about cancer and the importance of managing the disease is ever-increasing. Under the circumstances, however, people have difficulties in finding information customized to their needs among abundant information available to them. This paper aims to develop an ontology-based dialogue system for the effective provision of customized information to the public and present measures to utilize the system through

assessment.

Research was conducted on FAQs on web sites providing information about cancer. A number of cancer related web sites were run by hospitals, associations, and companies in Korea. A total of 140 FAQs on gastric cancer information were collected and finally arranged into 126 sets of question and answer. A total of 286 key words were selected from the questions. Definitions of 95 terms out of the key words were excerpted from a terminology dictionary.

In this study, an ontology dialog system was built with the Protege 2000 Ver 3.0. This ontology system was linked with the Dialogue Agent. Users entered a question through the Interface to ask the Dialogue Agent for an answer. The Interface was programmed through the Visual Studio.NET 2003. Similar to the existing dialogue agents, ASP.NET was used so as for the Interface to be run on the Internet and C# was used as behind Code. MS-SQL Sever was chosen as DBMS which can be approached through ADO.NET for the protection of unauthorized connection. The resulting CancerQ system was run on the web(URL : http://cancerq.ezzin.com).

To assess performance of the ontology system, searching capability and usability was measured. For the assessment of searching capability, a cancer specialist and a researcher did searching on the system two times to measure the recall, precision, and F-measure. They both reported that the ontology system performed better than existing web

sites. For the usability assessment, a total of 36 non-experts and medical experts did searching on gastric cancer information both on the existing web sites and the ontology system. Searching time, searching stage, and satisfaction with the results were compared. Unlike the medical experts who already had knowledge on the disease, the non-experts spent more time and stage in finding needed information, searched more times and expressed less satisfaction when using the existing system compared to the ontology system. Therefore, the ontology system was found to be more effective for the provision of specialized knowledge to non-experts than to experts.

The web site of the National Cancer Information Center (NCIC) provides information on cancer through a variety of ways including keyword searching, content searching, category searching, and terminology dictionary. If the ontology-based dialogue agent is improved into a more advanced system and then added to the web of the NCIC, it could provide faster and more user-friendly access to non-experts.

---

Key words : Ontology, Dialogue system, Cancer information