



지능형 지식저작 도구: Intelligent-Knowledge Authoring Tool(I-KAT)

저자 (Authors)	방재훈, 허태호, Taqdir Ali, Musarrat Hussain, 이승룡
출처 (Source)	한국통신학회지(정보와통신) 37(9) , 2020.8, 18-27 (10 pages) The Journal of The Korean Institute of Communication Sciences 37(9) , 2020.8, 18-27 (10 pages)
발행처 (Publisher)	한국통신학회 Korea Institute Of Communication Sciences
URL	http://www.dbpia.co.kr/journal/articleDetail?nodeId=NODE09873406
APA Style	방재훈, 허태호, Taqdir Ali, Musarrat Hussain, 이승룡 (2020). 지능형 지식저작 도구: Intelligent-Knowledge Authoring Tool(I-KAT). 한국통신학회지(정보와통신), 37(9), 18-27.
이용정보 (Accessed)	경희대학교 국제캠퍼스 163.***.116.140 2020/09/14 17:31 (KST)

저작권 안내

DBpia에서 제공되는 모든 저작물의 저작권은 원저작자에게 있으며, 누리미디어는 각 저작물의 내용을 보증하거나 책임을 지지 않습니다. 그리고 DBpia에서 제공되는 저작물은 DBpia와 구독 계약을 체결한 기관소속 이용자 혹은 해당 저작물의 개별 구매자가 비영리적으로만 이용할 수 있습니다. 그러므로 이에 위반하여 DBpia에서 제공되는 저작물을 복제, 전송 등의 방법으로 무단 이용하는 경우 관련 법령에 따라 민, 형사상의 책임을 질 수 있습니다.

Copyright Information

Copyright of all literary works provided by DBpia belongs to the copyright holder(s) and Nurimedia does not guarantee contents of the literary work or assume responsibility for the same. In addition, the literary works provided by DBpia may only be used by the users affiliated to the institutions which executed a subscription agreement with DBpia or the individual purchasers of the literary work(s) for non-commercial purposes. Therefore, any person who illegally uses the literary works provided by DBpia by means of reproduction or transmission shall assume civil and criminal responsibility according to applicable laws and regulations.

지능형 지식저작 도구: Intelligent-Knowledge Authoring Tool(I-KAT)

방재훈, 허태호, Taqdir Ali, Musarrat Hussain, 이승룡
경희대학교

요약

건강과 의료 관련 추천 시스템은 올바른 추천 사항을 적시에 올바른 사용자에게 제공해야 되고, 이를 위해 시스템이 보유하고 있는 지식은 진화해야 하며, 오류 없이 항상 정확하고 최신성을 유지해야 한다. 지식은 의사와 의료 데이터로부터 획득할 수 있으며 이를 바탕으로 한 지식베이스의 생성과 업데이트는 다양한 기계학습 알고리즘과 진화형 지식 획득 기술들을 이용하거나 전문가들이 직접 경험지식을 입력하고 검증할 수 있어야 한다. 그러나 현재 많은 의료 추천 시스템들은 의사의 경험이나 의료 데이터로부터 지식을 획득하거나 관리를 적절하게 수행할 수 있는 저작도구가 충분하지 않다. 이를 해결하기 위하여 본고에서는 지능형 지식 저작 도구 I-KAT (Intelligent-Knowledge Authoring Tool)을 제안한다. I-KAT은 지식 공학자로부터 도움을 최소로 받으며 의료 전문가들이 쉽게 지식을 획득, 관리 및 검증할 수 있도록 보조한다. I-KAT은 다양한 분야에서 활용되도록 개발되었으나 본고에서는 의료분야로 한정한다. 전문가의 지식 생성 효율을 높이기 위하여 I-KAT은 병원경영정보시스템의 임상 개념을 바탕으로 생성된 도메인 임상 모델(Domain Clinical Model, DCM)을 사용한다. I-KAT에 대한 성능 실험은 여러 방법이 있지만 본고에서는 룰을 생성하는 시간의 효율성과 사용자 설문을 통한 만족도를 측정하는 방법으로 범위를 제한 하였다.

I. 서론

임상 의사결정 지원시스템(Clinical Decision Support System, CDSS)과 의료 추천 시스템은 의사가 환자를 진단, 치료, 사후관리를 하는 동안에 의사결정을 지원하며[1], 지능형 의료 서비스 모델을 제공함으로써 정확하고 신속하게 환자에 대한 서비스 품질을 향상시키고 비용 효과성도 증진시킨다[2]. 정확하고 지능적인 의사결정 제공을 위해 CDSS는 검증된 최신의 진화

형 지식베이스를 필수적으로 가져야 하며[3][4][5], 손쉬운 시스템의 사용성, 높은성능, 접근성, 용이한 지식의 생성, 유지 관리 및 검증이 CDSS의 범용적 확산에 중요한 요소들이다[6]. 그리고, 지식 공학 병목 현상을 줄이기 위해 수행되는 작업의 수를 최소화하도록 전문가에게 통제된 상호작용 지식 저작 환경 제공이 필요하다[7][8].

또한, 기존의 의료 시스템과의 정보 공유를 위해서는 CDSS에서 생성된 지식은 공유 가능하고 상호운용성이 있어야 된다[9]. 지식 획득의 3요소인 용이성, 공유성, 상호운용성은 표준 데이터 모델과 표준 용어 사용을 통해 달성할 수 있다[9][10]. 임상 분야에서는 SNOMED CT, LOINC, ICD10[11][12] 등의 표준 용어와 Virtual Medical Record(vMR)[13] 등의 표준 모델이 사용 가능하다. 지식 기반 추천 시스템에서 기존 시스템들이 가진 이기종 지식베이스를 통합하는 것은 해결해야 할 하나의 과제이다. 이와 마찬가지로 데이터 기반 시스템에서는 추론 엔진과 데이터베이스가 서로 다른 형태로 구성되어 있다[14]. 지식 공유 연구 학계에서는 지식을 공유하기 위한 다양한 방법을 제시하고 있지만, 지식베이스 개발 도구의 재사용에 관한 연구는 미흡한 실정이다[15].

전문가는 지식 저작 도구를 통해 정확한 지식을 생성하고 검증 가능하며 유지관리가 용이한 지식 저작 환경을 제공받을 수 있다. 제안하는 I-KAT는 의사, 간호사, 간병인, 영양사 등이 지식 공학자의 최소한의 도움을 받으며 쉽게 지식을 생성, 진화, 검증할 수 있는 룰 저작 편집기를 제공한다. 룰 편집기는 통제 용어 환경인 인텔리센스 기능을 탑재하고 있다. 이는 일부 입력된 개념에 대한 리스트와 선택한 개념에 대해 사용될 수 있는 값의 리스트를 사용자에게 보여주는 기능으로 더 쉽고 빠르게 룰 생성과 수정을 가능하게 한다.

제안 시스템에서는 지식의 획득, 통합, 재사용을 위한 DCM을 제안하며 이를 기반으로 전문가의 경험지식을 지식베이스로 입력하는 역할을 수행한다. DCM은 현지화된 개념 온톨로지로서, 의사들이 많은 시간을 소요하여 어려운 표준 용어를 기억할 필요 없이 익숙한 용어를 사용하여 쉽고 빠르게 지식을 다룰 수 있게 한다. DCM은 전문가가 복잡하고 어려운 용어나 문법을 기억할

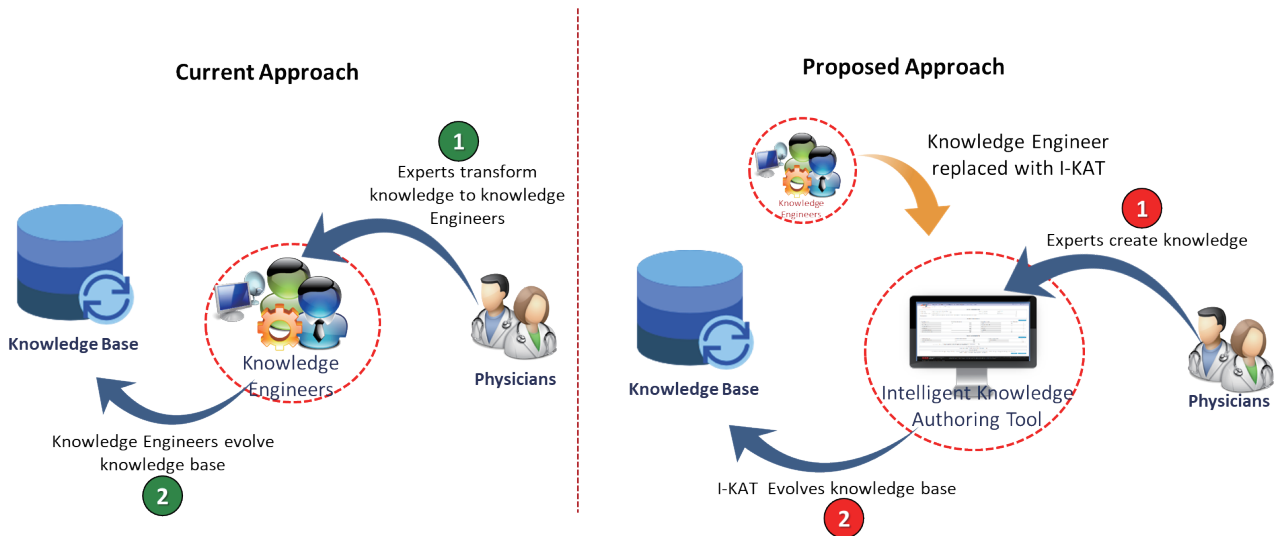


그림 1. 기존 방법과 제안 방법과의 비교

필요 없이 선택 가능한 개념의 리스트를 전문가에게 제공함으로써 쉽고 빠르게 룰을 생성하고 오류율을 감소시킨다.

〈그림 1〉은 제안하는 I-KAT과 기존 방법과의 차별성을 보여 준다. 기존 방법에서는 의사가 그들의 지식을 지식 공학자에게 제공하고, 지식 공학자는 이를 이용하여 시스템의 지식베이스를 진화시키는 방법으로 동작하여 의사는 지식 공학자의 도움 없이 직접 지식을 생성하고 표현할 수 없다. 반면 제안 방법에서는 지식 공학자의 역할을 I-KAT이 수행하여 의사가 직접 지식을 구축가능하다.

본고의 구성은 다음과 같다. 2장에서는 관련 연구를 기술하고, 3장에서는 제안하는 저작도구의 작동원리와 시스템 구성 컴포넌트들의 기능을 설명을 한 뒤, 4장에서 평가, 5장에서 결론을 내린다.

II. 관련 연구

연구 [16]에서는 모든 전문가 시스템의 목표는 지식을 항상 최신으로 유지하며 지식의 수정 시간을 최소화해야 된다고 말하며, 이를 위한 전문가 기반 지식 획득 도구를 제안하였다. 연구 [17]에서는 XML 방식으로 룰을 생성하는 XML 기반 지식 저작 도구를 제안하였다. 룰의 구조는 Clinical Document Architecture(CDA) 기반으로 형성되지만, 이 도구를 사용하기 위해서는 시스템이 CDA를 준수해야만 한다. 연구 [18]에서는 EMR 기반 지식 획득 도구를 제안하였지만 서술적 지식만을 지원하고, EMR 정보에만 의존하여 룰을 생성하며, 표준을 사용하지 않아 지식의 공유가 불가능하다는 단점을 가진다.

연구 [19]에서는 지식 공학자의 도움 없이 의사가 직접 지식을 생성할 수 있는 시스템을 제안하였다. 이는 마취 분야에서 “패턴과 결과” 방법으로 동작한다. 그러나 범위가 한정적이며 다른 분야에서는 동작할 수 없다. 이 시스템은 룰을 XML 형태로 생성하며, 이에 따라 의사들은 XML 작성법을 먼저 교육 받아야 한다. 연구 [20][21]에서는 지식베이스 유지관리 분석법을 제안하였으며, 이는 지식의 각 부분이 선언 규칙의 집합으로 구성되어 관리되는 정규 모델을 제안하였다. 이 모델을 수정함으로써 지식베이스를 쉽게 관리할 수 있다. 연구 [22]에서는 WebProtege라는 협업 온톨로지 편집기 및 지식 획득 도구를 제안하였다. 이 시스템은 백엔드 단에서의 협업을 위하여 오픈 소스 온톨로지 편집기인 protégé를 이용한다. WebProtege는 지식 획득 관점에서 우수하지만, 전문가가 시맨틱 웹 지식을 보유하고 있어야 사용이 용이하다.

한편, 최근에는 증가하는 데이터량과 발전된 하드웨어의 연산 능력 및 효율성 향상으로 인하여 기존의 수동으로 지식을 생성하는 방법에서 반자동 혹은 자동으로 지식을 생성하는 방법으로 데이터 기반 지식 획득 연구가 활발히 진행되고 있다[23]. 즉, 전문가가 데이터로부터 지식을 생성할 수 있는 직관적인 도구들이 많이 활용되고 있다. 이러한 관점에서, Rapid Miner는 매우 다재다능한 도구 중 하나로 다양한 작업을 수행할 수 있는 모듈식 환경을 제공한다[24]. Weka는 쉽게 사용 가능한 그래픽 사용자 인터페이스(GUI)를 제공하지만 Rapid Miner와 달리 의사 결정 모델로부터 지식을 추출하는 기능은 갖추지 않고 있다[25]. Konstanz Information Miner (KNIME)은 모듈식 환경과 GUI를 모두 제공하여 데이터 파이프라인을 쉽게 연결하고 쌍방향적

인 실행이 가능하며, 새로운 알고리즘과 도구들을 쉽게 통합할 수 있는 장점을 가진다[26].

III. I-KAT 작동 및 구조

전문가시스템 성공의 관건은 해당분야의 전문가로부터 지식 획득과정을 통해 구축된 지식베이스의 양과 품질에 의해 좌우된다[1]. 특히, 의료분야는 환자의 건강과 직결되는 특성상 임상 지식베이스는 높은 정확도와 검증 가능하며 유지관리가 용이하고 공유 가능한 지식 저작 환경 갖추어야 한다. 임상 의료지식 획득 과정은 의사들이 자신의 경험지식, 가이드라인, 기타 온라인 문헌 등을 참고하여 마인드맵을 작성 후, 이를 가시성이 좋은 결정 트리(Decision Tree, DT), PR(Production Rule), RDR(Ripple Down Rules), MLM(Medical Logic Module)등의 형태로 전산화하여 시스템에 저장하는 과정을 거친다. 그러나 이 과정에서 IT 비전공자인 의료 전문가는 지식의 전산화 과정을 구현하기 어렵고 이를 지원해 주는 저작 도구가 미흡하여 지식 공학자에 의존적으로 지식베이스를 구축해 왔다. 이는 지식 공학자의 임상 의료 지식의 이해 정도에 따라 지식의 품질이 영향을 받아 정확도 높은 지식을 획득하기에는 어려움이 있다.

따라서, 본고에서는 의료 전문가의 임상 의료 지식 획득과정에서 지식 공학자의 의존도를 최소화하고 의료진이 쉽게 자신의 임상 의료지식을 시스템에 입력 및 수정하여 양질의 지식을 획득할 수 있는 사용자 친화적인 지식 저작 도구인 I-KAT을 제안한다.

〈그림 2〉는 본고에서 다루는 I-KAT의 구현 환경을 개념적으로 보여준다. I-KAT는 저자가 개발한 지능형의료플랫폼(Intelligent Medical Platform, IMP)[27]과 연동되어 의료전문가들에게 지식 획득, 구축, 유지보수 및 검증할 수 있는 수단을 제공하며, 사용자에게 다양한 지능형 의료 서비스(Silo)를 제공할 수 있도록 지원한다.

I-KAT은 텍스트(가이드라인), 대화, 의료 이미지, EMR/EHR, 의료진 경험지식의 다양한 정형/비정형 데이터를 입력받아 전문가 기반, 데이터 기반 두가지 방식으로 지식을 획득한다. 전문가 기반 지식획득 방법은 Rule 편집기, DT 편집기, RDR 편집기, MLM 편집기의 4가지 편집 기능을 제공하여, 전문가의 경험 지식을 시스템 실행 가능한 지식베이스로 변환한다. 데이터 기반 지식 획득 방법은 축적된 의료 이미지, EMR/EHR, 대화, 텍스트(가이드라인) 데이터를 기반으로 각 데이터 소스마다 룰을 생성하고 이를 통합하여 지식베이스를 구축한다. 이렇게 생성된 전문가 기반, 데이터 기반 지식베이스는 여러 병원에 공유 가능하도록 유지보수가 뛰어난 RDR 모델로 통합되어 실행 가능한 지식으로 저장된다. I-KAT을 통해 획득된 지식은 CDSS 및 지능형 의료 플랫폼에 적용되어 병원의 의료진에게 서비스 된다.

〈그림 3〉은 제안하는 I-KAT의 전체 동작 과정을 보여준다. 동작 순서는 그림에서 보여주는데 색깔별로 붉은색의 전문가 기반(Expert Driven, ED), 초록색의 데이터 기반(Data Driven, DD), MLM 기반 혼합(MLM based Hybridization, MBH)으로 구성된다.

① ED: 의사가 선호도에 따라 Rule 편집기, DT 편집기, RDR 편집기, MLM 편집기를 사용하여 룰을 생성한다.

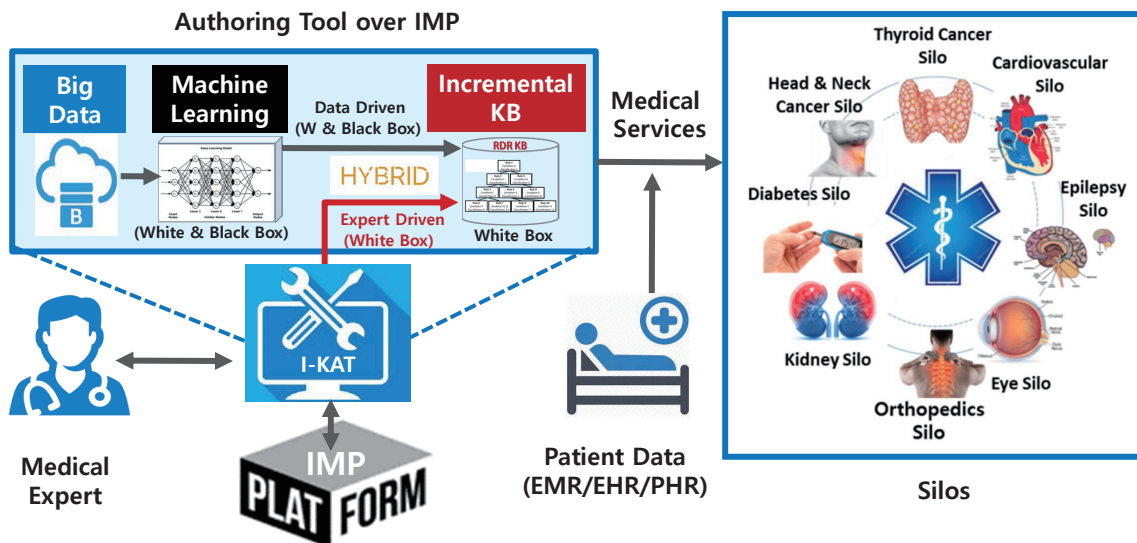


그림 2. IMP, I-KAT, Silo 연관도

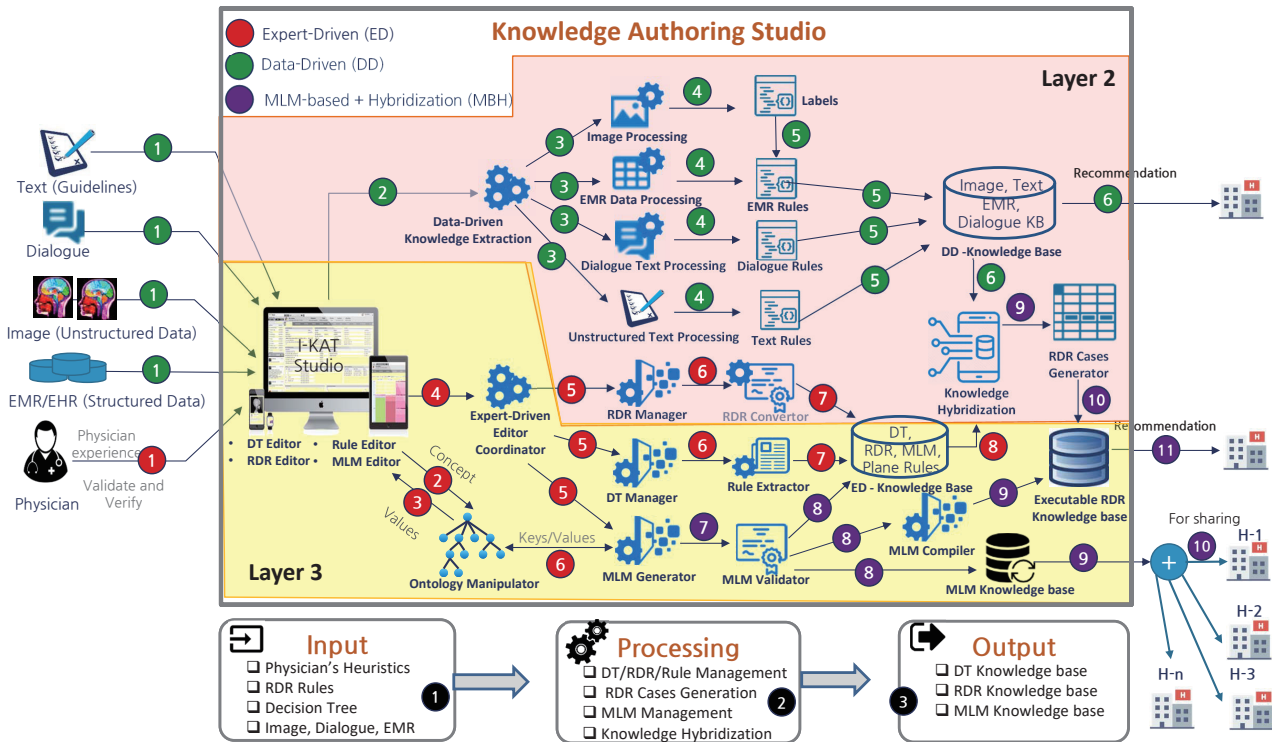


그림 3. I-KAT 구조 및 지식획득 순서도

DD: 의료 이미지, EMR/EHR, 대화, 텍스트의 정형/비정형 데이터가 입력된다.

- ② ED: 전문가의 편의를 위해 온톨로지 모델을 활용한 컨셉 리스트를 제공하는 인텔리센스 기능을 통해 원하는 컨셉을 선택한다.

DD: 입력된 정형/비정형 데이터를 데이터 기반 지식 추출 모듈로 전달된다.

- ③ ED: 선택된 개념에 대한 값을 리턴한다.

DD: 데이터의 종류에 따른 지식 추출을 위해 각기 다른 처리기로 데이터가 전달된다.

- ④ ED: 전문가 기반 편집기 조정자가 사용 중인 편집기와 통신한다.

DD: 각 데이터 소스의 특성에 따라 룰이 추출된다(의료 이미지 룰, EMR 룰, 대화 룰, 텍스트 룰).

- ⑤ ED: 전문가 기반 편집기 조정자가 각각의 룰 종류에 해당하는 관리자 룰을 전달한다.

DD: 각 데이터 소스에서 추출된 룰이 데이터 기반 지식베이스에 저장된다.

- ⑥ ED: RDR 관리자는 RDR 변환기로, DT 관리자는 룰 추출기로 룰을 전달한다. MLM 생성기는 온톨로지 조작기의 룰이 가진 키의 표준용어코드를 불러온다.

DD: 데이터 기반 지식베이스의 룰을 바탕으로 추천을 제공하며 지식 혼합 모듈로 전송된다.

- ⑦ ED: RDR 변환기는 입력된 룰을 RDR로 변환하고, 룰 추출기는 입력된 결정 트리를 룰로 변환하여 전문가 기반 지식베이스에 저장한다.

MBH: MLM 생성기에서 MLM을 생성하고 검증 위해 MLM 검증기로 전달된다.

- ⑧ ED: 저장된 룰은 지식 혼합 모듈로 전송된다.

MBH: 검증된 MLM은 지식 혼합 모듈, MLM 지식베이스, 그리고 실행 가능한 형태로 변환하기 위해 MLM 컴파일로 전달된다.

- ⑨ MBH: 컴파일된 MLM은 실행 가능한 MLM 지식베이스에 저장된다. 지식 혼합 모듈에 모인 룰들이 혼합되어 단일 형태의 지식으로 변환되며, 이는 RDR 변환에 필요 요소인 코너스톤 케이스 생성을 위해 RDR 케이스 생성기로 전달된다.

- ⑩ MBH: 지식베이스에 저장된 MLM들은 다른 병원과 공유되며, RDR 룰은 실행 가능한 지식과 연계되어 실시간으로 추천을 생성한다. 또한 RDR 룰 및 코너스톤 케이스가 실행 가능한 RDR 지식베이스에 저장된다.

- ⑪ MBH: 저장된 지식을 기반으로 입력에 따른 추천이 제공된다.

제안 I-KAT은 1) I-KAT 편집기, 2) 정형 지식 획득, 3) 이미지 지식 획득, 4) 서술적 지식 획득, 5) 대화 지식 획득, 6) 지식베이스 저장소, 7) 지식 혼합의 7개 컴포넌트로 구성되며, <그림 4>는 I-KAT의 구조도를 보여준다.

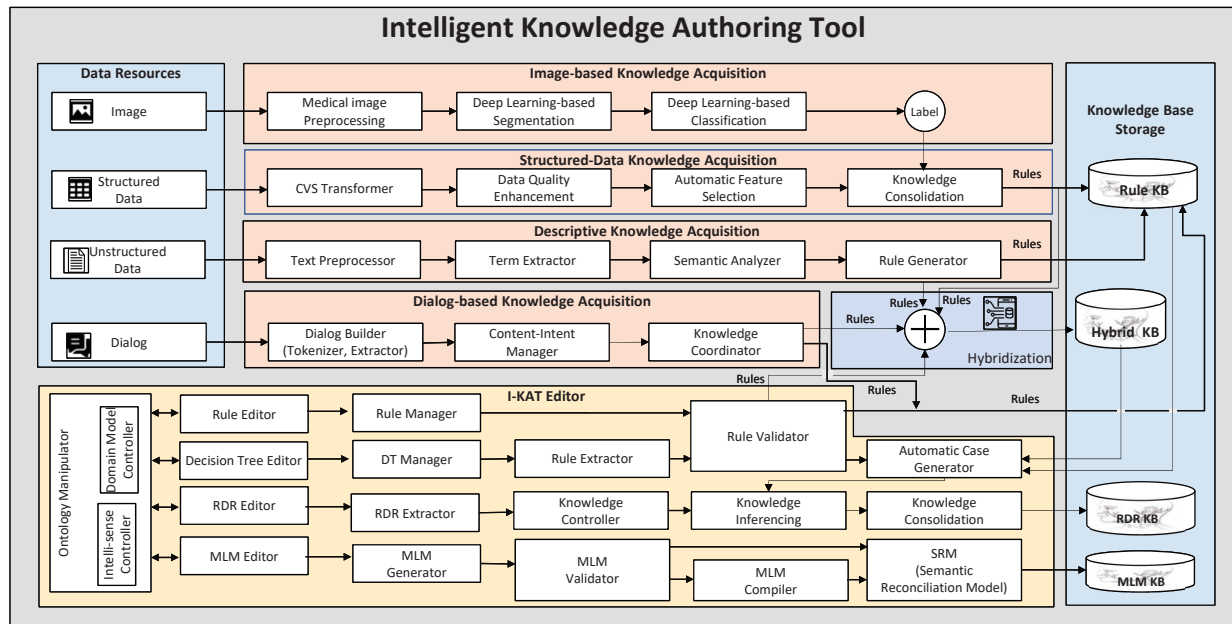


그림 4. I-KAT 구조 및 지식획득 순서도

1. I-KAT Editor(I-KAT 편집기)

I-KAT 편집기는 전문가의 경험과 지식을 입력하고 지식을 유지 관리할 수 있는 사용자 인터페이스를 제공한다. I-KAT 편집기는 Rule 편집기, DT 편집기, RDR 편집기, MLM 편집기를 제공한다. 편집 기능은 DCM 컨트롤러, 인텔리센스 컨트롤러, 편집기의 3가지를 기반으로 동작한다. 의사는 원하는 개념을 DCM 트리나 인텔리센스의 조건 및 결론 칸에서 선택할 수 있다.

편집기를 통해 생성된 룰들은 각각 룰 관리자, DT 관리자, RDR 추출자, MLM 생성자를 통하여 생산 룰, RDR 룰, MLM 형태의 지식으로 변환된다.

DT 편집기: 드래그 앤 드롭을 지원하는 결정 트리 모양의 지식 표현을 제공한다. 생성된 결정 트리는 자동으로 룰로 변환된다.

RDR 편집기: 시스템에 의해 제대로 분류되지 않은 새로운 사례를 데이터베이스에 추가한 후 지식베이스를 점진적으로 진화시키는 지식 획득 프로세스를 지원한다.

Rule 편집기: 지식을 키값쌍 형태로 변환하는 지식 저작 환경을 제공한다. 룰은 지식을 표현하기 위해 상황과 결론으로 구분되며, 추후 발생할 수 있는 룰의 변경을 지원한다.

MLM 편집기: 지식 공유를 위한 MLM 생성을 위한 룰을 입력한다.

온톨로지 조작기: 인텔리센스를 제공하기 위한 맞춤형 온톨로지를 보유하고 있다. 전체 SNOMED CT 온톨로지는 30만개의 개념을 가지고 있어 인텔리센스 제공에 지연이 생기게 된다. 따라서 제안 시스템에서는 전체 SNOMED CT에서 vMR 스키마

개념과 그 속성인 15%만을 따로 추출한 맞춤형 온톨로지를 사용한다.

MLM 생성기: 입력한 룰을 바탕으로 MLM을 생성한다.

컴파일 모듈: DT, RDR, 생산 룰은 추천을 생성하기 위해 바로 실행이 가능하나, MLM은 텍스트 형태로 되어 있어 바로 실행이 불가능하다. 따라서 본 모듈은 MLM을 실행 가능한 형태로 변환하기 위해 MLM을 컴파일하여 실행 가능한 exe 파일을 생성한다.

MLM 검증기/룰 검증기: 전문가에 의해 작성된 룰과 생산 룰이 표준 지식 표현법인 HL7 Arden Syntax의 문법을 따르는지와, RDR 룰이 RDR 문법을 따르는지를 검증한다. 또한, 룰 간의 중복과 충돌이 있는지도 검사한다.

자동 케이스 생성기: 기존의 다양한 지식베이스를 지식 진화가 가능하고 유지보수가 용이한 RDR 모델로 변환하기 위하여 코너스톤 케이스를 자동으로 생성한다.

2. 정형 지식 획득(Structured Knowledge Acquisition)

대부분 병원들은 의료 검사 결과, 징후, 증상, 약물, 치료 요법, 진단 결과 등의 환자 데이터를 제공하는 전자 의료 기록을 보유하고 있으며, 이들은 지식 생성에 중요하게 사용된다[28]. 전자 의료 기록은 표 형태로 되어 있으며 열(column)은 환자에 대한 속성을, 행(row)은 각 환자에 대한 정보를 담고 있다. 본 모듈에서는 이러한 의료 데이터로부터 지식을 추출하는 작업을 수행한다.

CSV 변환기: 테이블 형식의 전자 의료 기록의 데이터를 뽑아

내어 CSV 형식으로 변환한다.

데이터 품질 향상자: 데이터 품질을 계량화하여 데이터 기반 지식 획득에서 중요한 고품질의 데이터만을 이용한 지식을 획득한다.

특징 선택: 분류 모델이 높은 예측 정확도를 가질 수 있도록 높은 관련성과 낮은 중복성을 가진 중요한 특징 집합만을 선택한다.

지식 통합기: 선택된 특징들을 바탕으로 해석 가능한 여러 분류 알고리즘을 생성한다. 생성된 분류 모델들은 트리 형태 또는 의사결정의 목록 형태로 구성되며, 이들은 일련의 생산 규칙으로 변환된다.

3. 이미지 지식 획득(Image Knowledge Acquisition)

최근 비전 시스템에서 의료 이미지는 환자의 상태를 알 수 있는 중요한 데이터로 활용되고 있다. 예를 들어 폐, 안저, 뇌 등의 의료 이미지를 통해 질병, 결점, 이상 상황 등을 파악하는데 활용되고 있어 의료 이미지는 환자의 상태 파악에 매우 중요한 요소이다. 본 이미지 지식 획득 모듈에서 지식을 추출하는 것은 크게 전체 이미지에서 관심 영역만을 분리하는 과정과 해당 영역을 바탕으로 이상 여부를 판단하는 두 가지 과정으로 수행된다.

의료 이미지 전처리: 노이즈 감소를 위한 웨이블릿 변환과 대비 향상을 위한 동적 히스토그램 평활화의 고급 이미지 처리 기술을 활용하여 의료 이미지의 품질을 향상시킨다.

딥러닝 기반 분할: 브라켓 형태의 디코더를 이용하는 합성곱 신경망(CNN)을 이용하여 이미지에서 의미 있는 영역만을 분할한다.

딥러닝 기반 분류: 분할된 영역에서 결점 탐지를 위해 CNN 분류기를 사용한다.

레이블: 질병의 정도가 분류의 결과 레이블로 나타나며, 정형 지식 획득 컴포넌트로 전달된다.

4. 서술적 지식 획득(Descriptive Knowledge Acquisition)

본 모듈에서는 인터넷에서 획득 가능한 가이드라인, 임상 노트, 연구 문헌 등의 비정형 텍스트 데이터로부터 자동으로 관련 정보를 추출하고, 지식으로 변환하며, 시스템이 해석 가능하고 쉽게 진화할 수 있는 형태로 표현하는 기능을 수행한다.

텍스트 전처리: 비정형 텍스트 데이터를 입력받아 자연 언어 처리 기술을 사용하여 n-gram을 추출한다.

용어 추출기: n-gram에서 개념을 추출한다.

시멘틱 분석기: 개념들을 특정 유형의 관계를 기반으로 연결하여 개념 계층 구조를 생성한다.

룰 생성기: 계층적 룰 구조에서 개념을 추출하여 기존 온톨로지 모델에 추가하는 방식으로 진화 가능한 온톨로지 룰 모델을 구축한다.

5. 대화 지식 획득(Dialogue Knowledge Acquisition)

의사 또는 환자와 의료 SW 시스템 간의 상호작용을 통하여 획득한 대화 데이터에서는 환자의 식습관, 생활패턴 등이 내포되어 있어 진료에 필요한 히든 정보를 도출할 수 있는 중요한 데이터이다. 본 지식획득 방법론에서는 환자와의 대화 내용을 분석하여 룰을 생성하는 기능을 제공한다.

대화 생성자: 대화 생성자는 대화 모델에 따라 입력된 대화 텍스트를 이해 가능한 형식으로 변환한다. 대화 텍스트를 토큰화하고 명명된 엔티티 인식을 사용하여 가장 관련성이 높은 키워드를 필터링한다. 또한, 적절한 모델을 이용하여 토큰화된 단어에서 의미론적인 개념을 식별한다.

문맥-의도 관리자: 대화 인스턴스 검증, 문맥 식별, 의도 추론 등 분류되지 않은 대화 인스턴스를 위한 다양한 기능을 제공한다. 또한, 지식베이스를 분석하여 지식의 중복을 방지한다.

지식 조정자: 대화 생성자, 문맥-의도 관리자, 저장소 간의 통신 관리를 담당한다. 또한, 룰 추출과 보안, 성능, 투명성 등의 다양한 이벤트 로그를 추적한다.

6. 지식베이스 저장소(Knowledge Base Storage)

본 모듈에서는 지식베이스의 유형에 따라 지식을 영구히 저장하거나 임시로 저장하는 기능을 제공한다. RDR, MLM, 하이브리드 저장소의 3가지 종류가 있으며, RDR과 MLM은 영구히 저장되는 지식베이스로 각각 추천의 생성과 지식의 공유를 위해 사용되고 하이브리드 지식 저장소는 최종 지식 생성을 위해 혼합된 지식이 임시로 저장된다.

7. 지식 혼합(Knowledge Hybridization)

본 지식 혼합 기법은 전문가로부터 획득된 지식 정보와 데이터로부터 추출된 예측 모델(Prediction Model, PM)을 통합하여 보다 신뢰성 높은 지식베이스를 구축한다. 형태가 다른 두 종류의 지식을 혼합하는 방법은 아래의 7단계를 거쳐 수행된다.

- ① 의료 전문가가 검증 및 일치성에 대한 기준 C_v (하이브리딩 기준)를 설정한다.
- ② PM으로부터 생성된 모든 경로를 탐색할 때까지 각각의 경로 P_i 를 선택한다.
- ③ 기준 C_v 로부터 모든 기준을 확인할 때까지 각각의 C_i (정확도, 충돌, 확인, 증거)를 선택한다.
- ④ 선택한 각각의 P_i 에 대해 C_i 와의 부합성을 확인한다.
- ⑤ P_i 가 C_i 를 충족하면 다음 기준을 확인하며, 경로 P_i 를 점검하고 필요한 경우 정제하여 경로 P_j (Iterative DT 경로)를 생성한다. 경로 P_i 가 기준 C_i 를 충족시키지 못한다면 C_i 의 우선순위를

확인하여, 1순위일 경우 경로를 건너뛰고, 1순위가 아닐 경우 다음 기준을 확인한다.

- ⑥ 5단계 작업을 모든 경로와 기준에 대해 반복한다.
- ⑦ P를 추가하여 최종 혼합 지식을 진화시키고 필요한 경우 기존 경로를 갱신한다.

IV. 평가

제안한 I-KAT의 효율성을 평가하기 위해 실험은 룰 생성 소요 시간과 사용자 만족도 평가의 두 가지 방법으로 진행되었다. 실험에는 각각 2명의 운동 지도사, 영양사, 의사 총 6명이 실험에 참여하였다. 룰 생성 소요 시간 성능 평가를 위하여 참여자들은 5개의 룰을 생성하도록 지시되었으며, 각각의 룰 생성 시간이 측정되었다. 룰 생성은 DCM을 사용한 I-KAT, DCM을 미사용한 I-KAT, 그리고 통제 어휘 바탕의 DCM을 사용한 I-KAT의 3개 환경에서 시행되었다. 룰 생성 소요 시간을 정확하게 측정할 수 있도록 실험 시작 전 참여자들에게 I-KAT 사용법에 대한 사용설명서가 사전 제공되었다. 사용자 만족도는 인터페이스 디자인, 레이아웃 복잡도, 룰 생성에 걸리는 시간 요소를 기반으로 측정하였으며 이는 전문가가 I-KAT을 사용하고 난 후의 설문 결과를 토대로 평가되었다.

측정 결과는 <표 1>과 같다. 실험 결과 DCM을 사용한 I-KAT이 DCM을 쓰지 않았을 때보다 30%, 통제 어휘를 사용했을 때는 50% 더 빠른 성능을 나타냈다.

표 1. 룰 생성 시간 측정 결과

전문가	룰	I-KAT (DCM 포함)	I-KAT (DCM 미포함)	I-KAT (통제 어휘)
운동 지도사 1	R-1	1분 42초	1분 9초	59초
	R-2	1분 53초	1분 34초	1분 2초
	R-3	1분 52초	1분 29초	1분
	R-4	2분 58초	1분 42초	1분 11초
	R-5	1분 35초	1분 14초	56초
운동 지도사 2	R-1	2분 36초	1분 15초	54초
	R-2	4분 4초	1분 47초	1분 3초
	R-3	2분 59초	2분 29초	1분 49초
	R-4	4분 29초	2분 50초	2분 1초
	R-5	3분 18초	2분 30초	1분 48초
영양사 1	R-1	2분 19초	1분 50초	1분 13초
	R-2	2분 28초	1분 53초	1분 3초
	R-3	2분 58초	2분 3초	1분
	R-4	3분 30초	2분 18초	1분 41초
	R-5	2분 23초	1분 46초	1분 9초

전문가	룰	I-KAT (DCM 포함)	I-KAT (DCM 미포함)	I-KAT (통제 어휘)
영양사 2	R-1	3분 5초	2분 29초	2분 1초
	R-2	3분 28초	2분 36초	2분 11초
	R-3	3분 10초	2분 23초	1분 58초
	R-4	3분 3초	2분 29초	1분 48초
	R-5	4분 2초	3분 29초	2분 47초
의사 1	R-1	2분 48초	1분 52초	1분 6초
	R-2	3분 49초	2분 59초	2분 14초
	R-3	3분 28초	2분 45초	2분 1초
	R-4	2분 53초	2분 23초	1분 58초
	R-5	2분 30초	2분 22초	1분 44초
의사 2	R-1	2분 40초	1분 50초	1분 10초
	R-2	1분 55초	1분 17초	57초
	R-3	2분 39초	1분 20초	1분
	R-4	2분 59초	1분 42초	1분 5초
	R-5	2분 39초	1분 41초	1분 11초

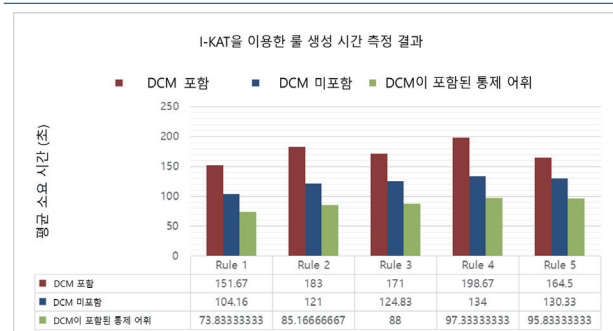


그림 5. I-KAT의 룰 생성 시간 측정 결과

두 번째 실험에서는 작성된 설문조사를 토대로 사용자 만족도를 측정하였다. 측정은 AttrakDiff[29] 측정법을 이용하여 실용 품질(Pragmatic Quality), 쾌락 품질(Hedonic Quality), 매력(Attractiveness) 차원을 측정하였다(그림 5), (그림 6).

- Pragmatic Quality (PQ): 제품의 유용성을 설명하고 사용자가 제품을 사용하여 얼마나 성공적으로 목표를 달성했는지를 나타낸다. 첫 실험에서는 90.4%의 만족도를 보였으며, 피드백 결과를 바탕으로 I-KAT을 수정 후 재실험 결과 96.5%의 만족도를 보였다.
- Hedonic Quality-Simulation (HQ-S): 제품의 참신성, 흥미성, 기능, 내용에 대한 사용자의 요구사항을 얼마나 충족시키는지를 나타낸다. 측정 결과 I-KAT은 93.2%의 만족도를 보였다.
- Hedonic Quality-Identity (HQ-I): 사용자가 제품을 어느 정도 수준으로 사용 가능한지를 나타낸다. 첫 실험에서는 78.4%의 만족도를 보였으나, 피드백 결과를 바탕으로

I-KAT을 수정 후 재실험 결과 91.4%의 만족도를 보였다.

- Attractiveness (ATT): 품질에 대한 전반적인 가치 점수를 나타낸다. 첫 실험에서는 75.3%의 만족도를 보였으나, 피드백 결과를 바탕으로 I-KAT을 수정 후 재실험 결과 93.2%의 만족도를 보였다.

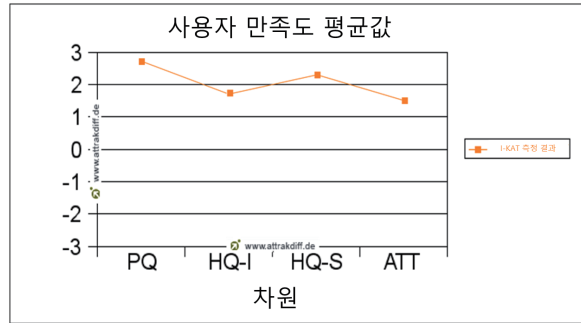


그림 6. I-KAT의 사용자 만족도 측정 결과

V. 결론

지식 저작 도구는 전문가가 지식 공학자의 도움을 최소화하면서 쉽고 편하게 지식을 생성, 관리, 검증할 수 있는 효과적인 저작 환경을 제공해야 한다. 이를 위해서는 멀티모달 데이터 처리, 지식 추출, 그리고 전문가에 의한 지식베이스의 진화 등 종합적인 핵심 기술을 뒷단(back-end)에서 지원해야만 한다. 제안하는 I-KAT은 의료 전문가가 쉽게 지식을 생성 및 관리할 수 있도록 인터페이스를 제공하고, 전문가 기반, 데이터 기반, 혼합 기반 지식획득의 핵심 기술을 뒷단에서 지원하여 진화형 지식베이스를 의료 전문가가 손쉽게 직접 생성할 수 있다. 실험을 통하여, 전문가들이 제안하는 I-KAT을 사용하여 지식을 빠르게 생성할 수 있으며, 사용 만족도가 높음을 입증하였다. 향후에는 보다 신뢰성 높은 지식을 획득하기 위해 지식 의미론적 검증 연구를 수행하여 I-KAT으로부터 수집되는 지식의 품질을 향상시킬 계획이다.

Acknowledgement

이 논문은 2020년도 정부(과학기술정보통신부)의 재원으로 정보통신기술진흥센터의 지원(IITP-2017-0-01629, IITP-2017-0-00655, IITP-2020-0-01489)과 한국연구재단의 지원(NRF-2016K1A3A7A03951968, NRF-2019R1A2C2090504)을 받아 수행된 연구임.

참고 문헌

- [1] A.Soumeiya, D.Michel, R.Claire, B.Philippe and L.Eric. "A UMLS-based knowledge acquisition tool for rule-based clinical decision support system development." Journal of the American Medical Informatics Association, vol 8, no. 4, pp. 351-360, 2012.
- [2] E. Sanchez, C. Toro, E. Carrasco, G. Bueno, C. Parra, P. Bonachela, M. Graña, F. Guijarro, An architecture for the semantic enhancement of clinical decision support systems, in: Knowledge-Based and Intelligent Information and Engineering Systems, Springer, 2011, pp. 611 - 620.
- [3] K. Fehre, K.-P. Adlassnig, "Service-oriented arden-syntax-based clinical decision support", Proceedings of eHealth2011. Vienna: Austrian Computer Society (2011) 123-8.
- [4] D. F. Sittig, A. Wright, J. A. Oshero, B. Middleton, J. M. Teich, J. S. Ash, E. Campbell, D. W. Bates, "Grand challenges in clinical decision support", Journal of biomedical informatics 41 (2) (2008) 387-392
- [5] D. Dustin, D. Jemery, B. Christopher, F. Simon and A. J. Mark. "A Knowledge Authoring Tool for Clinical Decision Support" Journal of Clinical Monitoring and Computing, vol 22, no. 3, pp. 189-198, 2008
- [6] A. Wright, D. F. Sittig, A four-phase model of the evolution of clinical decision support architectures, International journal of medical informatics 77 (10) (2008) 641-649.
- [7] Shortliffe, E. H. (1986). Medical expert systems—knowledge tools for physicians. Western Journal of Medicine, 145(6), 830.
- [8] Shortliffe, Edward H., and Martin J. Sepúlveda. "Clinical decision support in the era of artificial intelligence." Jama 320.21 (2018): 2199-2200.
- [9] Ali, T., Hussain, M., Ali Khan, W., Afzal, M., & Lee, S. (2013, July). Authoring tool: acquiring sharable knowledge for Smart CDSS. In Engineering in Medicine and Biology Society (EMBC), 2013 35th Annual International Conference of the IEEE (pp. 1278-1281). IEEE.
- [10] J. A. Oshero, M. J. Teich, B. Middleton, E. B. Steen, A. Wright, and D. E. Detmer. "A roadmap for national action

- on clinical decision support.” Journal of the American medical informatics association 14.2 (2007): 141-145.
- [11] Ali, T., Hussain, M., Khan, W. A., Afzal, M., Kang, B. H., & Lee, S. (2014, June). Arden syntax studio: Creating medical logic module as shareable knowledge. In Innovations in Intelligent Systems and Applications (INISTA) Proceedings, 2014 IEEE International Symposium on (pp. 266-272).
- [12] Summary of SNOMED CT Benefits, <http://www.ihtsdo.org/snomedct/whysnomedct/benefits/> [Visited on June, 2020].
- [13] K. Kawamoto, G. Del Fiore, H. R. Starsberg, N. Hulse, C. Curtis, J. J. Cimino, H. R. Beatriz et al. “Multi-national, multi-institutional analysis of clinical decision support data needs to inform development of the HL7 virtual medical record standard.” AMIA Annual Symposium Proceedings. Vol. 2010. American Medical Informatics Association, 2010.
- [14] B. Johansson, N. Shahsavari, H. Ahlfeldt, O. Wigertz, Database and knowledge base integration - a data mapping method for arden syntax knowledge modules., *Methods of information in medicine* 35 (4-5) (1996) 302 - 308.
- [15] V. K. Chaudhri, A. Farquhar, R. Fikes, P. D. Karp, J. P. Rice, Okbc: A programmatic foundation for knowledge base interoperability, in: AAAI/IAAI, 1998, pp. 600 - 607.
- [16] Regier, R., Gurjar, R., & Rocha, R. A. (2009). A clinical rule editor in an electronic medical record setting: development, design, and implementation. In AMIA Annual Symposium Proceedings (Vol. 2009, p. 537). American Medical Informatics Association.
- [17] Hulse, N. C., Rocha, R. A., Del Fiore, G., Bradshaw, R. L., Hanna, T. P., & Roemer, L. K. (2005). KAT: a flexible XML-based knowledge authoring environment. *Journal of the American Medical Informatics Association*, 12(4), 418-430.
- [18] Regier, R., Gurjar, R., & Rocha, R. A. (2009). A clinical rule editor in an electronic medical record setting: development, design, and implementation. In AMIA Annual Symposium Proceedings (Vol. 2009, p. 537). American Medical Informatics Association.
- [19] D. Dustin, et al. “A Knowledge Authoring Tool for Clinical Decision Support” *Journal of Clinical Monitoring and Computing*, vol 22, no. 3, pp. 189-198, 2008.
- [20] Debenham, John. “An analysis of knowledge base maintenance.” *Computer Science and Information Systems* 2.1 (2005): 1-29.
- [21] Debenham, John. “A rigorous approach to knowledge base maintenance. “*Developments in Applied Artificial Intelligence*. Springer Berlin Heidelberg, 2003. 219-228.
- [22] Tudorache, T., Nyulas, C., Noy, N. F., & Musen, M. A. (2013). WebProtégé: A collaborative ontology editor and knowledge acquisition tool for the web. *Semantic web*, 4(1), 89-99.
- [23] Ali, M., Ali, S. I., Kim, D., Hur, T., Bang, J., Lee, S., ... & Hussain, M. (2018). uEFS: An efficient and comprehensive ensemble-based feature selection methodology to select informative features. *PloS one*, 13(8), e0202705.
- [24] Hofmann, M., & Klinkenberg, R. (Eds.). (2016). *RapidMiner: Data mining use cases and business analytics applications*. CRC Press.
- [25] Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., & Witten, I. H. (2009). The WEKA data mining software: an update. *ACM SIGKDD explorations newsletter*, 11(1), 10-18.
- [26] Berthold, Michael R., Nicolas Cebrian, Fabian Dill, Thomas R. Gabriel, Tobias Kötter, Thorsten Meinl, Peter Ohl, Kilian Thiel, and Bernd Wiswedel. “KNIME-the Konstanz information miner: version 2.0 and beyond.” *ACM SIGKDD explorations Newsletter* 11, no. 1 (2009): 26-31. beyond. *ACM SIGKDD explorations Newsletter*, 11(1), 26-31.
- [27] Ali, Taqdir, et al. “The Intelligent Medical Platform: A Novel Dialogue-Based Platform for Health-Care Services.” *Computer* 53.2 (2020): 35-45.
- [28] Ali, S. I., & Lee, S. (2020, January). Ensemble based Cost-Sensitive Feature Selection for Consolidated Knowledge Base Creation. In 2020 14th International Conference on Ubiquitous Information Management and Communication (IMCOM) (pp. 1-7). IEEE.
- [29] Hassenzahl, M. (2006). Hedonic, emotional, and experiential perspectives on product quality. *Encyclopedia of human computer interaction*, 266-272

약 력



방재훈

2011년 평택대학교 디지털응용정보학과 이학사
 2013년 경희대학교 컴퓨터공학과 공학석사
 2019년 경희대학교 컴퓨터공학과 공학박사
 2019년~현재 경희대학교 지능형의료플랫폼 연구센터
 연구박사
 관심분야: 인공지능, 패턴인식, 유비쿼터스 컴퓨팅



허태호

2010년 경희대학교 컴퓨터공학 학사
 2020년 경희대학교 컴퓨터공학과 박사
 2020년~현재 경희대학교 지능형의료플랫폼 연구센터
 연구박사
 관심분야: 인공지능, 패턴인식, 상황인지



Taqdir Ali

2006년 Kohat University of Science and Technology 학사
 2019년 경희대학교 컴퓨터공학 박사
 2019년~현재 경희대학교 박사 후 연구원
 관심분야: 임상 의사결정지원시스템

Musarrat
Hussain

2010년 파키스탄 University of Peshawar 컴퓨터공학 학사
 2015년 파키스탄 National University of Science and
 Technology 소프트웨어공학 석사
 2016년~현재 경희대학교 컴퓨터공학 박사과정
 2008년~2016년 소프트웨어 엔지니어
 관심 분야: 임상 텍스트 마이닝, 자연 언어 처리, 인공지능



이승룡

1978년 고려대학교 학사
 1987년 Illinois Institute of Technology 전산학 석사
 1991년 Illinois Institute of Technology 전산학 박사
 1993년~현재 경희대학교 컴퓨터공학과 교수
 관심분야: 유비쿼터스컴퓨팅, 상황인지 미들웨어,
 지능형 컴퓨팅, 디지털 헬스케어