

국방 컴포넌트그리드 기반의 시맨틱 검색 기술의 연구

허윤, 김수경, 최호진
한국과학기술원 전산학과
e-mail : { hurosy, ksk0314, hojinc }@kaist.ac.kr

A Study on Techniques for Semantic Search based on Defense Software Component Grid

Yun Her, Su-Kyoung Kim, Ho-Jin Choi
Dept. of Computer Science, Korea Advanced Institution for Science and Technology

요 약

본 연구는 국방 소프트웨어 컴포넌트그리드 환경의 자산저장소를 지원하는 시맨틱 검색 시스템을 설계하고 개발하는데 바탕을 두고 있다. 컴포넌트그리드 환경의 자산저장소의 중요한 특성은 재사용성과 상호운용성 그리고 유용성을 보장하는 것이다. 이러한 특성을 만족하는 시맨틱 검색 시스템을 개발하기 위해서는 기반 기술에 대한 심도 있는 기초 연구가 필요하다. 본 논문에서는 이러한 기술들 중 현재 연구 중인 몇 가지를 소개하고 적용 방법을 제안하고자 한다. 이러한 기술로는 사례기반추론을 이용한 소프트웨어 개발 경험재사용 연구, 유사한 컴포넌트들의 추출을 위한 의미기반의 유사도 연구, 그리고 사용자 질의의 추론과 매칭을 위한 추론규칙 연구 등이 있다. 본 연구에서는 다양한 형태의 산출물들의 저장 및 검색을 위한 기술들을 조사하고 이를 연구하여 향후 컴포넌트그리드 환경의 자산저장소의 시맨틱 검색을 제공하기 위한 기초로 활용할 예정이다.

1. 서론

국방 도메인의 소프트웨어 개발 분야에서의 중요한 것은 재사용성과 상호운용성 지원이다. 이를 위해 연구되고 있는 컴포넌트그리드는 서로 다른 국방 응용 소프트웨어 개발자들이 개발한 소프트웨어 컴포넌트들을 효과적으로 공유하고 체계적으로 재사용할 수 있도록 하기 위한 많은 기반 기술 들로 구성되어 있다[1]. 컴포넌트그리드 환경의 검색 시스템의 요건은 재사용성과 상호운용성의 관점에서 다양한 종류의 자산을 검색할 수 있어야 한다는 것이다. 따라서 자산 종류(혹은 대상)에 따라 적합한 검색 방법이 무엇인지 찾아내고 그에 맞는 기술들을 연구하는 것이 중요하다. 이러한 기술들을 찾아내고, 또한 이를 기초로 시맨틱 검색 시스템 개발을 위해 해당 연구들의 개념을 더욱 확장시키고 발전시키는 것이 본 연구의 주 목표이다. 나아가 이렇게 개발된 기술들이 검색 시스템 개발을 위해 활용되는 방법을 연구할 것이고, 이러한 기술들이 유기적으로 접목되어 시맨틱 검색 시스템으로 구현 될 수 있도록 연구하고자 한다.

2. 시맨틱 검색 기술

본 연구에서 다루는 해당 기술들을 간략히 소개하면 첫 번째로 명세화 및 표현된 소프트웨어 컴포넌트들 간의 의미적 유사도를 측정하여 사용자의 의도에 가장 유사한 컴포넌트들을 추천해주기 위한 기초 기

술로서 의미유사도(semantic similarity) 측정 기술이 있고, 하나의 새로운 문제를 해결하기 위해 과거에 경험했던 문제의 해결 방법을 변형시켜 이용하는 방법으로서[2], 사람이 문제를 해결하기 위하여 생각하고 추론하는 방법을 정형화한 사례기반추론(case-based reasoning 이하 CBR)기법, 그리고 컴포넌트 명세를 작성하고 검색함으로써 개발자가 검색 대상 컴포넌트와 의미적으로 밀접한 컴포넌트들을 검색할 수 있도록 하기 위한 기반 기술인 시맨틱 검색 기술 등이다.

2.1 개념 근사 기반의 컴포넌트 검색

소프트웨어 자산저장소에서 사용자들이 컴포넌트 검색을 할 때, 필요한 컴포넌트를 효과적으로 찾아내기 어려운 이유는 사용자들이 원하는 컴포넌트 자체에 대해 무지하기 때문이다. 즉, 다양한 소프트웨어 컴포넌트들 중에서 사용자가 원하는 컴포넌트를 효과적으로 찾아내기 위한 관련 키워드에 익숙하지 않거나, 혹은 사용자들이 필요로 하는 컴포넌트 자체의 존재를 모르기 때문이다. 이러한 문제를 해결하기 위한 새로운 대안으로 컴포넌트를 추천해주는 RASCAL과 같은 방법이 제안되었다[3]. 하지만 이 방법도 사용자들이 명확하고 정확한 질의를 알고 있다는 가정에 기초한다. 본 연구는 두 가지의 차별성을 바탕으로 이러한 제한을 해결하고자 한다. RASCAL 이 정확한 매칭에 기반하는 반면에 본 연구는 의미 기반 매칭을 활용한다는 것과 RASCAL 이 협력적 필터링 방

법을 이용하는 반면 본 연구에서는 FCA 기법을 활용한다는 것이다.

2.2 CBR 기반의 컴포넌트 검색 모델

컴포넌트 검색 기술을 위한 몇 가지 방법들이 연구되고 있는데, 이러한 기술들의 필수 요건은 사용자의도에 부합하는 기능이 제대로 검색되는지 여부이다. 현재 연구되고 있는 기존의 방법들은 다음과 문제들 안고 있기에 이를 해결한 더욱 진화된 기술이 필요하게 되었다.

- 결정 지원 부재 (inability to provide decision support)

검색된 컴포넌트들이 사용자가 필요로 하는 기능을 갖고 있지만, 반드시 최적은 아니다. 이는 재사용 관점에서 컴포넌트 선택에 대한 어떠한 정보도 제공하고 있지 않고 있다는 점에서 비롯된다.

- 검색 결과 설명 미흡 (inability to explain retrieval)

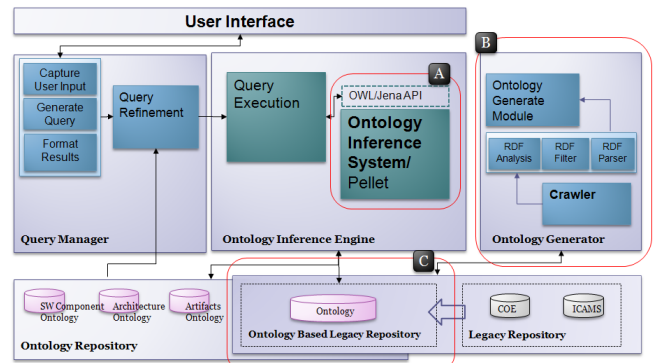
검색된 컴포넌트의 검색 결과에 대해서 설명을 제공하지는 않는다. 예를 들어 사용자의 질의와 검색된 컴포넌트들의 연관성에 대해서 아무런 설명을 제공하지 않는 것이다.

성공적인 컴포넌트 검색과 선택을 위해서 이러한 문제들을 해결하는 방향으로 검색 기술의 진화가 이루어져야 한다. 이러한 문제를 해결하기 위해 컴포넌트 검색 과정에서 과거의 컴포넌트를 선택했던 경험에 근거하여 재사용하는 방법을 제안한다. 본 연구에서는 과거의 컴포넌트를 선택했던 결정을 재사용하기 위해서 CBR의 프레임워크를 적용하였다.

2.3 시맨틱 검색

본 연구에서는 컴포넌트그리드 환경 기반의 레거시(legacy) 데이터베이스를 온톨로지 지식베이스로 변환하는 온톨로지 생성기를 적용하고자 한다. 컴포넌트그리드에서는 기존에 활용되던 컴포넌트 개발 방법론 기반의 레거시 지식베이스들이 존재한다[1][4]. 레거시 지식베이스는 관계형 데이터베이스 형태로 존재하므로 적용 계층 성격의 데이터베이스 중재자를 지원한다[1]. 본 연구의 목적은 새로운 온톨로지 기반의 지식베이스가 기존의 지식베이스와 상존함으로써 생기는 복잡성을 제거하고자 하는 것이다. 이원화된 지식저장소 통합을 통해 상호운용성 증대를 기대한다. 이를 위해 기존의 지식베이스를 온톨로지로 변환하는 온톨로지 생성기(B)에 대한 연구가 필요하다. [그림 1]에서 블록 B는 온톨로지 생성기를 나타낸다. 내부에는 온톨로지 크롤러(crawler)와 RDF(resource description framework) 해석기, 필터, 파서 모듈이 존재할 것이다. 온톨로지 크롤러는 지속적으로 갱신되는 기존의 자산저장소의 관계형 데이터베이스 기반의 데이터들을 추출하여 주어, 술어, 객체의 RDF 트리플로 수집하여주는 역할을 한다[5]. 이후 RDF 분석과 파싱이 이루어지고 온톨로지 생성 모듈을 통해 온톨로지 지식베이스(C)로 저장된다. [그림 1]은 컴포넌트그리드 환경을 기반으로 온톨로지 추론 엔진과 온톨로지 생성기를 적용하여 시맨틱 검색 시스템의 구조를 구상해

본 것이다.



(그림 1) 온톨로지 추론 엔진과 온톨로지 생성기를 적용한 시맨틱 검색 시스템 설계 구상도

3. 결론

지금까지 살펴본 각각의 기술들은 컴포넌트그리드 환경에서 생산되는 다양한 산출물 형태의 자산들이 효과적으로 저장 및 검색될 수 있는 시맨틱 검색 시스템을 개발하는데 중요한 기술들로 활용될 것이다. 앞으로의 연구는 CBR 기법과 의미유사도 측정 기법, 그리고 시맨틱 검색의 기본 개념을 활용하는 관점에서 좀 더 연구의 폭과 깊이를 심화시킬 것이다. 향후 가장 핵심이 되면서도 난관으로 예상되는 연구 이슈는 개별적으로 연구되고 있는 기술들의 유기적인 접목과 통합이다.

감사의 글

본 연구는 방위사업청과 국방과학연구소의 지원으로 수행되었습니다. (2009-SW-12-DM-01)

참고문헌

- [1] 고인영, 조혜경, 최호진, “국방 소프트웨어 컴포넌트 그리드 환경 설계”, 정보과학회지 제 25 권 제 9 호, pp.5-17, 2007. 9.
- [2] Aamodt, A, Plaza. E, “Case-based reasoning: foundational issues, methodological variations, and system approaches.”, AI Communications 7(1), 39-59, 1994.
- [3] Frank McCarey, Mel O Cinnéide, and Nicholas Kuskmerick. “Rascal: A recommender agent for agile reuse.” Artificial Intelligence Review, 24(3-4):253-276, November. 2005.
- [4] 국방과학연구소, “국방 아키텍처 프레임워크 (MND-AFV1.0 사용자 지침서)”, 2005. 2.
- [5] 공현장, 황명권, 김관구, “온톨로지 기반 의미적 정보검색 시스템 설계”, 한국지능정보시스템학회 2005 추계학술대회, pp 304~308, 2005.