

온톨로지 매핑 기반 엔지니어링 정보 검색

정민*(한국과학기술원 산업공학과), 서효원(한국과학기술원 산업공학과)

Engineering Information Search based on Ontology Mapping

Min Jung (Industrial Eng. Dept. KAIST), Hyo-won Suh (Industrial Eng. Dept. KAIST)

ABSTRACT

The participants in collaborative environment want to get the right documents which are intended to find. In general search system, it searches documents which contain only the keywords. For searching different word-expressions for the same meaning, we perform mapping before searching. Our mapping logic consists of three steps. First, the character matching is the mapping of two terminologies that have identical character strings. Second, the definition comparing is the method that compares two terminologies' definitions. Third, the similarity checking pairs terminologies which were not mapped by two prior steps. In this paper, we propose Engineering Information Search System based on ontology mapping.

Key Words : Ontology (온톨로지), Semantic (의미), Mapping (매핑), Search (검색)

1. 서론

CPC 환경에서는 방대한 양의 정보 교환과 검색이 발생한다. 효과적인 협업을 위해서는 참여자들이 다른 참여자들이 작업(상품 개발 또는 연구) 해 놓은 정보를 바탕으로 다음 작업을 진행해야 한다. 그러나 참여자들은 보통 서로 다른 용어를 사용하기 때문에 정보(문서)를 검색하는데 어려움이 존재한다.

정보시스템 내부에서 사용하는 용어들이 온톨로지 정의를 갖는다는 가정하에서, 본 연구는 정확성 있는 검색에 대한 해결책으로 온톨로지 매핑을 기반으로 한 검색을 제안한다. 본 연구의 구성은 다음과 같다. 2 장에서는 온톨로지 매핑 기반 엔지니어링 정보 검색 방법과 OntSE (Ontology Search Engine) 시스템에 대해서 설명한다. 3 장에서는 OntSE 시스템에서 온톨로지 매핑 알고리즘이 어떻게 적용되는지를 설명한다.

본 연구에서는 온톨로지의 기본 요소를 일반적으로 많이 사용되는 5-tuple [1] 구조를 채택하였다. 또한 온톨로지 표현 언어로서 First Order Logic(FOL)을 선택하였다[2].

2. 온톨로지 매핑 기반 엔지니어링 정보 검색

온톨로지 매핑 기반 엔지니어링 정보 검색은 세가지 단계로 진행된다. 사용자에게서 검색을 하고자 하는 키워드와 온톨로지 정의를 받아오는 과정, 온톨로지 매핑을 통해 유사어를 찾아내는 과정, 그리고 찾아낸 유사어를 엔지니어링 정보가 존재하

는 문서나 정보를 검색해내는 과정으로 진행된다. 이때 온톨로지 매핑 알고리즘은 문자열 비교, 정의 비교, 유사도 검사의 3 단계로 진행된다[3].

Fig 1 에서는 OntSE 시스템의 구조를 보여준다. OntSE 시스템 크게 UI, 내부 로직, Database 로 구성되어 있다. OntSE 시스템은 사용자로부터 키워드와 온톨로지 정의를 받아오는 웹페이지를 제공하고, 입력 받은 온톨로지를 가공한 뒤 온톨로지 매핑을 수행한다. 매핑을 통해 검색어에 매칭된 용어들을 구한 뒤, 그 용어들이 존재하는 문서에 접근한다. 이러한 과정들을 통해 사용자에게 검색어에 대한 결과(문서)를 제공한다.

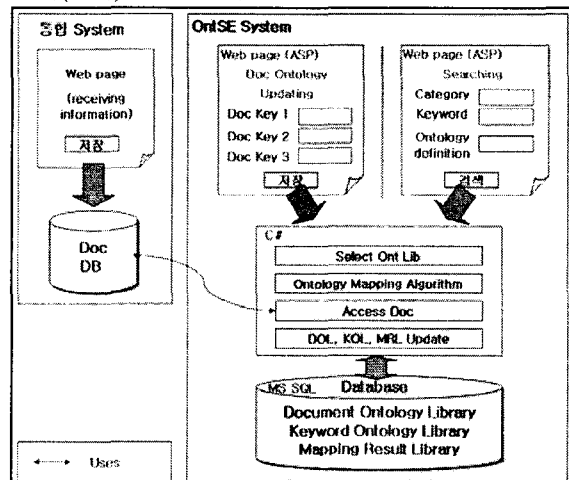


Fig. 1 Architecture of OntSE System

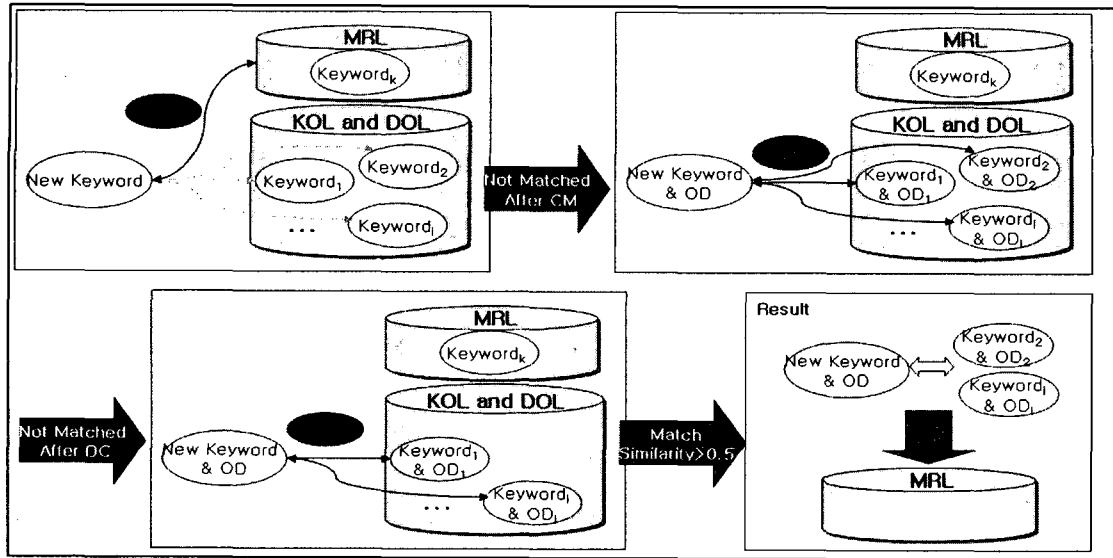


Fig 2. Ontology Mapping Algorithm on OntSE System

3. OntSE 시스템에서의 온톨로지 매핑

OntSE 시스템의 Database에는 Document Ontology Library(DOL), Keyword Ontology Library(KOL)와 Mapping Result Library(MRL)가 구분되어 있다.

통합시스템에서 문서가 새로 등록될 때, 사용자는 문서에 해당하는 키워드와 온톨로지 정의를 입력한다. 이 키워드와 온톨로지 정의는 DOL에 저장되고, 사용자가 검색을 할 때 입력하는 키워드와 온톨로지 정의는 KOL에 저장된다. MRL에는 온톨로지 매핑을 통해 매칭이 된 용어들에 대한 정보가 존재한다.

온톨로지 매핑 알고리즘은 문자열 비교(Character Matching, CM), 정의 비교(Definition Comparing, DC)과 유사도 검사(Similarity Checking, SC)의 세 단계로 진행된다. Fig 2에서와 같이, CM은 MRL에 저장되어 있는 용어에 대해서 진행한다. 이는 매핑 로직의 수행 시간을 줄일 수 있게 한다. 한편, DC와 SC 단계는 검색어의 온톨로지 정의와 KOL, DOL에 저장되어 있으면서 MRL에 존재하지 않은 용어들에 대해 매칭을 수행한다.

DC에서는 용어의 온톨로지 정의를 컨셉과 릴레이션으로 분류하여 각각을 CM으로 매칭 여부를 판단한다. CM과 DC 단계에서 매칭이 되지 않은 용어에 대해서 SC를 수행한다. SC에서는 용어 자체의 유사도와 온톨로지 정의의 유사도를 구한 뒤 가중합을 한다. 이 때 온톨로지 정의의 유사도에 가중치를 더 크게 준다. 유사도가 0.5 이상이라면 두 용어가 매칭이 된다고 결정한다.

이때, 매칭된 용어들을 MRL(Mapping Result Library)에 입력하여, 다음 검색에서 활용할 수 있도록 한다.

4. 결론

본 연구에서는 일반 검색이 아닌 의미 기반의 검색을 연구하였다. 온톨로지 매핑 로직은 문자열 비교, 정의 비교, 유사도 검사의 3 단계로 이루어진다. OntSE 시스템에서는 매핑 로직을 통해 유사어를 찾은 뒤 엔지니어링 정보가 저장되어 있는 문서와 Database를 검색한다. 이 시스템의 결과는 엔지니어링 정보를 열람할 수 있는 링크가 담긴 웹페이지이다. 따라서 사용자는 원하는 정보가 담긴 문서를 참고하여 업무에 임할 수 있다.

본 연구에서 제안된 온톨로지 매핑 기반 검색 방법은 의미가 다르고 표현이 다른 용어에 대해서는 매핑이 되지 않고 검색이 되지 않는다는 제한점을 가지고 있다. 따라서 일반적인 경우에 대한 매핑 로직의 개발이 유연한 정보 공유를 위해서 필요하다.

온톨로지 매핑 기반 검색의 적용을 위해서는 먼저 온톨로지를 구축해야 한다. 그러나 온톨로지 디자인은 방대한 양의 지식공학 작업, 많은 시간과 노력을 필요로 한다. 따라서 이에 대한 간편하고 효율적인 접근법이 개발되어야 할 것이다.

참고문헌

1. E. Bozsak, et al., "KAON: Towards a large scale Semantic Web," E-commerce and Web technology, pp304-313, 2002.
2. Stuart Russell and Peter Norvig, "Artificial Intelligence," Prentice Hall, 1995.
3. M. J. Jung, "Ontology-based Semantic Mapping Approach using a Bayesian Network for Communication in CPC Environment," Master thesis, KAIST, 2006.