

문제기술서의 문장 분석을 통한 개념클래스 도출 방법 제안

박가영, 이용훈, 이상범
단국대학교 학교 전자계산학과
e-mail:tuesday16@hanmail.net

Method of making a conception class from problem description using sentence analysis

Ga-Young Park, Yong-Hun Lee, San-Bum Lee
Dept of Computer Science, Dankook University

요 약

본 논문에서는 고객의 초기 요구 문서인 문제 기술서로부터 문장 분석을 통하여 개념클래스를 도출하는 방법에 대하여 제안한다. 문제기술서는 개발할 시스템에 대한 고객의 요구사항을 개괄적으로 작성된 문서로서 요구사항을 분석하고 설계할 때 이용되는 초기 문서 역할을 한다. 개발자는 이 문서의 내용을 바탕으로 시스템에 중요한 개념클래스를 도출하는 작업을 시작한다. 시스템 개발에 사용될 개념클래스들을 도출하는 것은 객체지향적인 설계에서 중요한 과정이다. 문제 기술서로부터 개념클래스 도출 과정은 주로 개발자의 경험과 직관에 의존하는 경향이 있으며 경험이 많은 전문 개발자들은 문제 기술서로부터 개념클래스들을 올바르게 도출할 수 있지만 초보 개발자의 경우 도출에 어려움이 따른다. 이러한 문제점의 개선 방법으로 기술서의 문서에서 문장 단위로 형태소 분석을 통하여 명사와 동사를 추출한다. 추출된 명사를 통하여 2가지 분류 기준에 따라 개념클래스 도출하고, 이후 동사를 이용하여 개념클래스 간의 관계 설정하는 방법을 제안한다.

1. 서론

소프트웨어 개발이란 사용자로부터 요구된 기능들을 정확히 수행할 수 있는 프로그램을 개발하는 것을 의미한다. 프로그램의 거대화로 인해 정확한 기능을 수행하는 프로그램의 요구와 개발 후 프로그램의 유지·보수를 최소화하기 위해서 사용자로부터 정확한 요구사항을 받는 것이 중요하다. 또한 개발 할 프로그램을 정확히 설계하는 작업도 매우 중요시 되고 있다. 소프트웨어의 개발과정에서 요구사항 분석단계는 사용자로부터 전달 받은 요구사항을 분석하고 어떠한 프로그램을 개발해야 하는 지에 대한 밑그림을 그리는 작업이며 이후 설계, 구현 및 유지보수 단계를 지원하기 위한 가장 중요한 단계이다. 그러므로 개발자는 사용자의 요구사항을 정확히 분석하기 위해 사용자와의 많은 대화를 통해 문서화 작업을 하여 사용자로부터 정확한 요구사항을 이끌어 내야 한다. 사용자가 생각하는 요구사항과 개발자가 이해한 요구사항이 정확히 일치해야 하는 것이다. 이러한 작업을 도와주기 위해 문제기술서를 만든다.

문제기술서는 개발할 시스템의 대한 사용자의 요구사항을 개괄적으로 작성한 문서로서 시스템에 대한 요구사항을 분석할 때 이용되는 초기 문서 역할을 한다. 사용자는 자신이 원하는 시스템의 기능과 특성을 기술하므로 개발자는 이 문서에서 소프트웨어 개발에 있어 중요한 개념

을 담고 있는 객체들을 찾을 수 있다. 이 객체를 클래스 다이어그램을 이용하여 가시적으로 표현한 것을 개념클래스라고 한다. 개념클래스를 이용하여 객체지향 소프트웨어 시스템의 기본 단위인 클래스를 식별하고 그 관계를 쉽게 정의 할 수 있으며, 요구사항이 정확히 분석되었는지 사용자와의 의사소통이 원활 할 수 있는 도구로 이용된다.

이러한 개념클래스를 도출하는 과정을 돕기 위해서 본 논문에서는 문장 분석을 통하여 개념클래스를 좀 더 쉽게 도출하기 위한 방법을 제안한다. 개념클래스 도출 방법으로 문제기술서에서 형태소 분석을 통하여 명사와 동사를 추출한다. 추출한 명사는 후보개념클래스가 되고 동사는 후보개념클래스들의 관계를 설정하는 역할을 하게 된다. 후보개념클래스는 개념클래스가 될 수 있는 클래스들로 개념클래스 분류 카테고리를 이용하여 1차 개념클래스 도출 작업을 실행한다. 1차 개념클래스 도출 작업이 끝난 후, 개념클래스가 될 수 없는 카테고리를 이용하여 후보개념클래스를 제거하는 2차 개념클래스 도출 작업을 진행한다. 이 과정을 거치면 개념클래스가 선택되어 지며 형태소 분석에서 나온 동사와의 관계를 파악하여 개념클래스들 간의 관계를 설정할 수 있다.

2. 관련연구

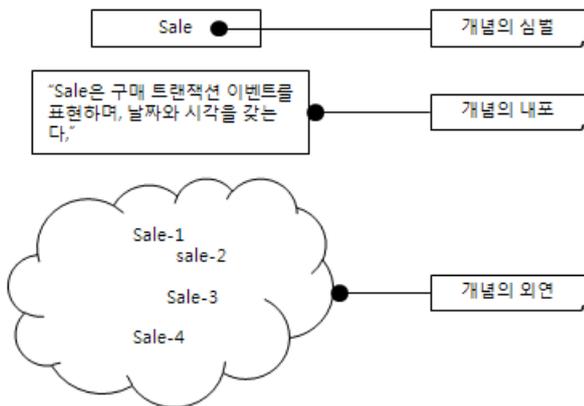
2.1 형태소 분석

형태소 분석은 문서 또는 문장을 구성하는 단어들의 품사를 파악하여 문장의 구조를 분석해내는 분야이다[1]. 한국어 형태소 분석은 많은 문법적 변화를 가지고 있으며 단어가 축약, 탈락, 불규칙 활용 등으로 인하여 여러 가지 형태로 변경되어 문장에서 표현된다[2]. 이러한 변화가 많은 한국어 문장에서 품사를 결정하는데 어려움을 겪게 된다. 문서에서 중요한 단어 추출에 사용하는 품사로서 명사를 들 수 있다. 명사는 문장의 중심적인 역할을 하며 문장 및 문서를 표현하는데 중요한 의미를 지니고 있다. 명사의 역할을 하는 주어와 목적어가 중요하긴 하지만 명사가 문장의 전체의 의미를 대변한다고는 할 수 없다. 주어와 목적어의 관계가 동사에 의해 결정되어 지므로 동사 파악은 문장의 의미를 파악하는데 중요한 역할을 한다.

이러한 명사와 동사를 파악하는 것은 문장을 분석하는데 있어서 중요한 과정이다. 문제기술서로부터 추출된 명사와 동사 파악으로 시스템에서 요구되는 개념들을 도출하고 클래스 간의 관계를 설정하는데 사용할 수 있게 된다.

2.2 개념클래스 도출

개념클래스는 아이디어, 사물, 혹은 객체이다. 좀 더 형식적으로 표현하면, 개념적 클래스는 심벌, 내포, 외연의 측면에서 볼 수 있다. 심벌(symbol)는 개념클래스를 표현하는 단어나 상징이며, 내포(intension)는 개념클래스의 정의를 나타낸다. 외연(extension)는 개념클래스가 적용되는 예들이다. <그림 1>은 심벌, 내포, 외연에 대해 나타내고 있다.



<그림 1> 개념클래스의 심벌, 내포, 외연

개념클래스를 도출하는 방법 중 명사 분석 방법은 개념클래스를 도출하는 가장 간단하면서도 효과적인 방법이다. 문제기술서를 바탕으로 명사를 찾는 방법인데 물론 모든 명사가 클래스에 해당되는 것은 아니며, 문제기술서의 명사가 클래스보다는 다른 클래스의 속성이나 연산에 해당될 수도 있다. 그래서 개념클래스 카테고리 목록과 부적절한 클래스를 찾아 제거해야 적합한 개념클래스를 도출

할 수 있다. 개념클래스 카테고리 목록은 일반적으로 객체가 될 수 있는 카테고리를 목록화한 것으로 무작정 개념클래스를 도출하는 것 보다 개념클래스 카테고리 목록을 통해 고려할 가치가 있는 개념클래스를 도출하는 것이 낫다.

3. 개념클래스 도출 방법

3.1 문제기술서의 문장 분석과정

본 연구에서 먼저 실행 할 작업은 문제기술서에서 명사와 동사를 정확히 분석하는 일이다. 문제기술서는 개발할 시스템에 대해서 사용자가 기대하는 모든 기능을 포함하고 있는 문서로서, 사용자와 개발자간의 이해를 도와준다. 그러므로 문제기술서에는 개발할 시스템의 영역이 명확히 정의되어야 하며, 가능한 자세하게 작성해야 한다. 완벽하게 작성된 문제기술서에서 좋은 개념클래스를 도출할 수 있다.

문제기술서를 이용한 문장 분석 방법은 형태소 분석을 통하여 명사와 동사를 추출한다. 추출된 명사와 동사는 개념클래스 후보목록에 기록된다.

“24시 편의점 관리 시스템이다. 시스템은 다양한 외부 시스템과 연계되어 24시간 내내 작동되어야한다. 고객이 편의점에서 물품구입을 원하면 점원은 바코드 스캐너를 이용해서 물품 판매를 기록하고 영수증을 발급해준다. 이때 고객은 현금 또는 카드를 사용해서 물품을 구입할 수 있다.

편의점에는 고객이 원하는 물건이 있는지 살펴볼 수 있도록 상품 카달로그(Catalog)가 비치되어있으며 카달로그에는 상품 설명과 가격이 적혀있다. 편의점 관리자는 금전등록기에 동전이나 지폐가 적정량 있는지 확인하고 남은 금액은 수거한다.”

위의 글은 24시 편의점 관리 시스템의 문제기술서이다. <표 1>는 위의 문제기술서를 형태소 분석을 통해 명사와 동사를 정리한 개념클래스 후보목록이다.

<표 1> 개념클래스 후보목록

연번	문장번호	추출명사	추출동사
1	2	시스템	
2	2	외부시스템	연계되다
3	2	24시간	작동되다
4	3	고객	
5	3	편의점	
6	3	물품구입	원하다
7	3	점원	
8	3	바코드스캐너	기록하다
9	3	물품판매	
10	3	영수증	발급하다
11	4	고객	
12	4	현금	
13	4	카드	이용하다
14	4	물품	구입하다

15	5	편의점	
16	5	고객	
17	5	물건	살펴보다
18	5	상품카탈로그	비치되다
19	5	카탈로그	
20	5	상품설명	
21	5	가격	적혀있다
22	6	편의점 관리자	
23	6	금전등록기	확인하다
24	6	동전	
25	6	지폐	수거하다

3.2 개념클래스 후보 추출 방법

형태소분석을 통해 명사와 동사를 추출하는데 여기서 추출된 명사는 후보개념클래스를 나타내고, 동사는 개념클래스를 연결하기 위한 관계를 나타낸다. 문제기술서의 문장 분석과정에서 작성한 개념클래스 후보목록을 이용하여 2단계 과정을 거쳐 개념클래스를 도출한다. 개념클래스를 도출하기 위한 2단계는 다음과 같다.

1단계 개념클래스 도출과정은 개념클래스 카테고리 목록을 이용하여 카테고리에 해당되는 명사를 선택하고, 선택된 명사들은 1단계 후보개념클래스가 된다. 개념클래스 카테고리 목록은 고려할 가치가 있는 일반적인 카테고리들을 포함하는데, 여기서 순서는 우선순위를 나타내지 않는다. <표 2>는 개념클래스 카테고리 목록을 나타낸다.

<표 2> 개념클래스 카테고리 목록(Conceptual Class Category List)

연번	List
1	물리적 유형 객체
2	사물에 대한 명세, 설계, 혹은 기술들
3	장소
4	트랜잭션(Transaction)
5	트랜잭션 라인 아이템
6	사람의 역할, 조직
7	컨테이너(Container)
8	외부시스템
9	추상명사 개념
10	프로세스
11	법칙과 정책
12	카탈로그
13	재정, 작업, 계약, 기록
14	서비스
15	매뉴얼, 문서, 참고문헌, 책

<표 3>는 1단계 개념클래스 도출 목록을 나타낸다. 각 카테고리에 해당이 된다면 선택한다. 선택되지 않는 명사는 개념클래스에서 제외되어 개념클래스 후보목록에서 제외된다.

<표 3> 1단계 개념클래스 도출 목록

카테고리 단어	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
시스템	√														
외부 시스템	√														
24시간									√						
고객						√									
편의점			√												
물품구입															
점원						√									
바코드 스캐너								√							
물품판매										√					
영수증													√		
고객						√									
현금	√														
카드	√														
물품				√											
편의점			√												
고객						√									
물건				√											
상품															
카탈로그															√
카탈로그															√
상품설명		√													
가격		√													
편의점 관리자						√									
금전 등록기							√								
동전	√														
지폐	√														

2단계 개념클래스 도출 과정은 부적절한 개념클래스를 제거하는 과정이다. 1단계에서 찾은 개념클래스들은 아직 완벽한 개념클래스라고 보기 어렵다. 중복된 단어와 클래스가 아닌 연산이나 속성인 개념도 있다. 또 구현과 관련된 개념과 UseCase에서 액터와 같이 시스템 외부에 존재하는 개념도 있을 수 있다. 2단계에서는 후보개념클래스에서 개념클래스가 될 수 없는 개념을 제거함으로써 시스템에서 필요로 하는 개념만 도출하게 된다. <표 4>는 부적절한 개념클래스 목록을 나타낸다.

<표 4> 부적절한 개념클래스 목록

연번	부적절한 개념클래스
1	시스템 외부의 존재
2	중복된 클래스
3	시스템의 기능과 무관한 클래스
4	속성
5	연산
6	구현과 관련된 클래스

1단계 개념클래스 도출 방식과 같은 방식으로 2단계에서도 부적절한 개념클래스를 선택한다. 선택된 명사는 <표 1>의 개념클래스 후보목록에서 제외시킨다.

<표 5> 2단계 개념클래스 도출 목록

카테고리 단어	1	2	3	4	5	6
시스템	✓					
외부시스템	✓		✓			
24시간			✓			
고객						
편의점						
물품구입						
집원						
바코드스캐너	✓					
물품판매						
영수증					✓	
고객		✓				
현금				✓		
카드				✓		
물품						
편의점		✓				
고객		✓				
물건		✓				
상품카달로그						
카달로그		✓				
상품설명				✓		
가격				✓		
편의점 관리자						
금전등록기						
동전		✓		✓		
지폐		✓		✓		

개념클래스는 2단계를 거쳐 도출된다. 1단계에서는 개념클래스 카테고리 목록을 이용하여 도출된 단어들을 간추려 내는 작업을 하고, 2단계에서는 시스템에서 개념이 되는 클래스를 찾아내는 작업을 한다. 최종 개념클래스는 개념클래스 후보목록에 남아있게 되고 동사를 이용하여 다음 장에서 개념클래스 간의 관계설정 방법을 설명할 것이다.

3.3 개념클래스 간의 관계 설정 과정

개념클래스 간의 관계는 UML에서 사용하는 집합관계, 의존관계 등 자세한 관계 표현은 의미가 없다. 그래서 본 연구에서는 다른 관계를 사용하지 않고 가장 간단하고 일반적인 연관관계를 사용하여 개념클래스 간의 관계를 설정한다. 개념클래스 간의 관계 설정은 다음에 2가지 방법을 사용한다.

첫 번째, 형태소 분석을 통해 추출한 동사의 앞과 뒤의 명사를 파악하고 앞과 뒤의 두 명사는 관계가 있다고 판단하여 관계를 설정한다.

두 번째, 한 문장에 있는 명사는 관련이 있는 명사라고

판단하고 같은 문장번호를 가진 명사들의 관계를 설정한다.

<그림 2>은 24시 편의점 관리 시스템의 개념클래스를 추출하여 관계를 설정한 것이다.



<그림 2> 24시 편의점 관리 시스템의 개념클래스

4. 결론 및 향후 연구과제

본 연구에서는 문제기술서를 이용하여 형태소 분석 후 명사와 동사를 이용하여 개념클래스를 도출하고 개념클래스들 간의 관계를 설정하는 방법을 제안하였다. 이로써 개념클래스 도출에 있어 범하기 쉬운 오류들을 최소화하여 좋은 개념클래스를 도출할 수 있게 되었다. 그 결과 설계와 구현단계에 기초 바탕이 되었으며, 개념클래스 도출이 어려운 초보 개발자에게 가이드라인을 제공했다.

하지만 이 문제기술서를 이용하여 개념클래스를 도출하는 방법은 기계적인 매핑은 어려우며 자연어는 모호성을 가지고 있어서 개발자의 관여가 필요하다. 또한 문제기술서에서 나타나 있지 않은 개념클래스에 대해서 개발자는 주의를 기울여야 한다.

앞으로 본 연구에서 제시한 방법을 이용하여 시스템을 구현한 후, 개념클래스 도출의 정확성에 관한 연구를 계속할 것이다.

참고문헌

[1] 강승식, “음절 특성을 이용한 한국어 불규칙 용언의 형태소 분석”, 정보과학회논문지, 제22권, 제10호, pp. 1480-1487, 10월, 1995.
 [2] 심광섭, 양재형, “인접 조건 검사에 의한 초고속 한국어 형태소 분석”, 정보과학회논문지, 제31권, 제1호, pp. 89-99, 1월, 2004.
 [3] 이서정, “클래스 도출을 지원하는 세가지 완충 다이어그램”, 한국전자거래학회지, 제9권, 제3호, pp. 145-154, 8월, 2004.
 [4] 이해선, 박재년 “비즈니스 시스템 분석을 위한 확장된 정보구조 모델링”, 한국정보과학회, pp. 379-381, 7월, 2005.