

한국정보과학회
Korean Society of Information Science and Technology

제 24 권 제 1 호
Vol. 24 No. 1



2022

제 24 회 한국 소프트웨어공학 학술대회 논문집

Proceedings of the 24th Korea Conference on
Software Engineering (KCSE 2022)

- 일시: 2022년 1월 19일(수) ~ 1월 21일(금)
- 장소: 강원도 평창 한화리조트(휘닉스파크점)

주최: 한국정보과학회, 한국정보처리학회

주관: 한국정보과학회 소프트웨어공학 소사이어티
한국정보처리학회 소프트웨어공학연구회

후원: **ETRI** 한국전자통신연구원 Electronics and Telecommunications Research Institute, **SOLUTIONLINK** (주)비트컴퓨터,
한국소프트웨어기술진흥협회(KOSTA), T3Q(주),
(주)유정인베스트먼트, (주)다한테크, (주)브이웨이,
브이플러스랩(주), 슈어소프트테크(주), (주)이에스지,
(주)코어다트랩, STA 테스트컨설팅(주),
TTA 소프트웨어시험인증연구소,
서강대 지능형블록체인연구센터,
KAIST 빅데이터 엣지 클라우드서비스 연구센터

1 월 21 일 (금)				
시 간	행 사 내 용			
	논문 발표			
	E1: 요구사항 및 아키텍처 I 좌장: 고인영 (KAIST) 장소: 그랜드홀 2	E2: 요구사항 및 아키텍처 II 좌장: 유준범 (건국대) 장소: 세미나실 1	E3: SW 모델링 및 테스트 좌장: 이주용 (UNIST) 장소: 세미나실 2	워크숍 W2:
9:25-10:50 (85 분)	[우수 일반 논문] 소프트웨어 요구사항 분류 모델의 성능 향상을 위한 불균형 데이터 처리에 관한 연구 최종우, 이영준, 임채균, 최호진 (KAIST) [우수 일반 논문] 효율적인 아키텍처 패턴 적용을 위한 소프트웨어 요구사항 분류체계에 관한 연구 최종우, 민상윤 (KAIST) [단편 논문] 코드가시화의 고도화 연구 김장환, 문소영, 김영철 (홍익대) [단편 논문] 자카드 지수를 이용한 한국어 요구사항 유사도 분석 자동화 장우성, 문소영, 김영철 (홍익대) [단편 논문] Use case 명세서 기반 자동 가이드 코드 생성 정세준, 문소영, 김영철 (홍익대)	[일반 논문] Status Quo of Software Architecture and Technologies for Cloud-Edge Collaborative Intelligence Xiangchi Song, 고인영 (KAIST) [우수 일반 논문] AIAG-VDA FMEA 방법론 기반 SFMEA : 효과적인 자동차 SW 안전분석 김소연, 권기현 (경기대) [일반 논문] 엣지 클라우드 환경에서 다목적 최적화 알고리즘을 활용하여 비용 및 서비스 지연을 최소화하는 마이크로서비스 배포 방법 김민협, 고인영 (KAIST) [산업체 논문] 소프트웨어 수출을 위한 시장주도 요구공학 적용 사례 이종길, 길문주, 오정섭 (엔에스이)	[일반 논문] 다양한 디스플레이 환경을 위한 적응형 UI 레이아웃 프레임워크 윤진, 이은영, 이상현 (삼성전자), 백종문 (KAIST) [단편 논문] 멀티태스크 PLC 프로그램의 테스트를 위한 주기 별 테스트 시퀀스 실행 제어 방법 은형석, Lingjun Liu, 지은경, 배두환 (KAIST) [산업체 논문] 반도체 공정 유지 보수 비용 절감을 위한 공정 레시피 데이터 학습 및 장비 결함 검출 최민호(피에스케이 / 아주대), 유동연, 이정원 (아주대) [학부생 논문] 아동·청소년을 위한 인공지능 스피커: 심리 상담 통합 시스템 개발 김수아, 이예빈, 정수민, 최미소, 최혜빈 (이화여대) [학부생 논문] 스마트 팩토리-가스라이터 제조 공정에서의 스티커 부착과 관련된 불량 검출 프로그램 개선 신관수, 송호준, 조영진, 신연순 (동국대)	서강대 지능형블록체인 연구센터 장소: 세미나실 4 (10:00~11:50)
10:50-11:00	휴식			
	신진 연구자 초청 세미나 N4 좌장: 김미정 (UNIST) 장소: 그랜드홀 2	신진 연구자 초청 세미나 N5 좌장: 유신 (KAIST) 장소: 세미나실 1	신진 연구자 초청 세미나 N6 좌장: 지은경 (KAIST) 장소: 세미나실 2	
11:00 -11:50 (50 분)	오픈소스 소프트웨어에서의 버그 정정 관리를 위한 지능형 프레임워크 양근석 교수 (경남대)	비형식 요구사항 기반 유스케이스 추출을 통한 노력 추정 방법 박보경 교수 (진주교대)	소프트웨어 신뢰도 보증을 위한 효율적인 도달 가능성 분석 이낙원 박사 (KAIST)	
11:50-12:10 (20 분)	폐회식 장소: 그랜드홀 2 사회: 이정원 조직위원장 (아주대)			

자카드 지수를 이용한 한국어 요구사항 유사도 분석 자동화

장우성[○], 문소영, 김영철

홍익대학교 소프트웨어공학연구실

{uriel2001, whit2, bob}@hongik.ac.kr

Automatic Analysis Method using Jaccard Coefficient for Korean Requirements Similarity

Woo Sung Jang[○], So Young Moon, R. Young Chul Kim

SELab., Hongik University

요 약

요구공학의 영역에서 중요한 초점 중 하나는 고객의 비정형 요구사항을 올바르게 파악하는 것이다. 하지만 올바른 요구사항의 식별은 매우 어렵다. 그리고 한국어 요구사항 자동 분석 연구는 매우 드문 실정이다. 우리는 한국어 요구사항 기반의 원인/결과 모델링 및 테스트케이스 자동 생성에 초점을 두고 있다. 이 방법은 요구사항 문장을 단순화가 가장 중요하다. 하지만 요구사항의 원인/결과들의 유사성 식별을 해결하지 못한다. 문제의 해결을 위해 자카드 유사도를 이용한 요구사항 유사성 식별 방법을 제안한다. 이 방법은 유사도를 수학적으로 자동 계산함으로써 유사한 요구사항 내의 원인/결과를 일치화한다.

1. 서 론

많은 소프트웨어 개발 조직은 고품질의 소프트웨어를 개발하기 위해 다양한 테스트 방법을 사용하고, 테스트 비용을 최대한 절감하기 위한 방법을 고려한다[1]. 하지만 한국어 기반의 요구사항으로부터 테스트 케이스를 자동 생성하는 연구는 매우 드물다. 이는 한국어 문장 의미의 자동 분석이 힘들기 때문이다.

기존 어려움의 해결을 위해 한국어 요구사항으로부터 테스트케이스 자동 생성 방법을 연구한다[2]. 방법은 다음과 같다. 1) 한국어 요구사항을 단순화, 2) 단순화 과정을 C3Tree Model[3]로 표현, 3) C3Tree Model로부터 Cause-Effect Graph[4]를 생성, 4) Cause-Effect Graph로부터 Decision Table을 생성, 5) Decision Table로부터 테스트케이스를 생성하는 과정들을 자동화한다. 이 방법은 최소한의 테스트케이스로 100% 커버리지를 만족한다[2]. 하지만 유사한 요구사항 문장을 식별하지 못한다. 이는 중복되는 Cause와 Effect가 증가되고, 이는 테스트케이스의 불필요한 증가를 초래할 수 있다.

이러한 문제의 해결을 위해 자카드 유사도[5]를 사용하여 C3Tree Model에서 유사 노드들의 일치화 후, Cause-Effect Graph로 자동 변환하는 방법을 제안한다. 결과적으로 중복되는 테스트케이스를 최소화할 수 있다.

본 논문의 구성은 다음과 같다. 2장은 관련 연구를 언급한다. 3장은 자카드 유사도 분석을 이용한 한국어 요구사항 자동 분석 방법을 언급한다. 4장은 적용 사례를 언급한다. 마지막으로 결론 및 향후 연구를

언급한다.

2. 관련 연구

2.1 Cause-Effect Graph

Cause-Effect Graph[4]는 시스템에서 가장 작은 기능적 단위를 식별하여 원인(Cause)과 결과(Effect)의 관계를 가지는 기능들을 서로 연결한다. 이 관계에서 왼쪽의 노드가 원인이고, 오른쪽의 노드가 결과이다. Cause-Effect Graph의 연결 관계는 그림 1과 같다. Identity는 단순한 원인-결과 관계이다. NOT 관계는 원인이 참이라면 결과가 거짓이고, 원인이 거짓이라면 결과가 참인 관계이다. OR 관계는 여러 개의 원인 중 하나라도 참이라면 결과가 참인 관계이다. AND 관계는 여러 개의 원인이 모두 참이어야 결과가 참인 관계이다.

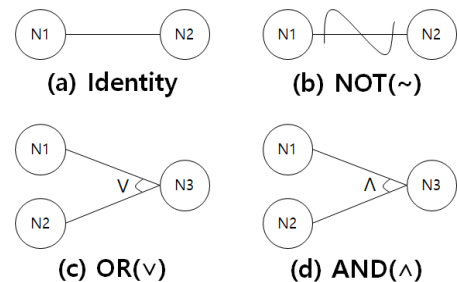


그림 1 Cause-Effect Graph의 연결 관계

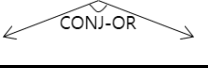
2.2 C3Tree Model

C3Tree Model[3]은 복잡한 요구사항 문장의 간소화

과정을 트리 구조로 표현한다. 표 1은 C3Tree Model의 구성요소에 대해 설명한다.

표 1 C3Tree Model의 구성요소

노드	설명
<<Sentence>> Korean Sentence	요구사항 문장
<<Clause>> Korean Sentence	간소화된 문장
<<Complex-Clause>> Korean Sentence	여러 개의 Cause 노드를 포함하는 문장

연결관계	설명
	긍정적 조건 관계 (좌측 자식 노드가 참이라면, 우측 자식 노드가 참)
	부정적 조건 관계 (좌측 자식 노드가 참이라면, 우측 자식 노드가 거짓)
	AND 관계 (모든 자식 노드가 참이라면, 부모 노드가 참)
	OR 관계 (모든 자식 노드 중 하나라도 참이라면 부모 노드가 참)

2.3 자카드 유사도

자카드 유사도[5]는 두 집합 사이의 유사도를 측정하는 방법이다. 자카드 지수(J)는 0~1 사이의 값을 가진다. 두 집합의 유사도가 동일하면 1의 값을 가지고, 두 집합이 전혀 유사하지 않으면 0의 값을 가진다. 집합 A, B에 대해 자카드 유사도를 측정하는 수식은 아래의 그림 2와 같다.

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|} = \frac{|A \cap B|}{|A| + |B| - |A \cap B|}$$

그림 2 자카드 유사도 계산식

3. 자카드 유사도 분석을 이용한 한국어 요구사항 유사도 자동 분석 방법

한국어 요구사항을 간소화를 위해, 간소화 과정을 C3Tree Model로 표현하고, C3Tree Model에서 유사한 노드를 합친다. 마지막으로 C3Tree Model을 Cause-Effect Graph로 변환하는 과정을 설명한다.

3.1 복잡한 요구사항의 간소화

한국어 요구사항의 간소화 과정은 아래와 같다. 분석하는 문장의 구성에 따라 과정의 순서가 바뀔 수 있다.

- 긍정적 조건절, 부정적 조건절 식별
- AND 연결절, OR 연결절 식별

- 주어 복구
- 수동문을 능동문으로 변환

간소화 과정의 예는 다음과 같다. 표 2는 예제로 사용되는 요구사항 문장이다.

표 2 요구사항 문장의 예제

A가 입력되고 B가 입력되면 C를 출력하다.

표 3은 표 2 요구사항 문장의 형태소 분석 결과이며, 16개의 형태소가 분석된다.

표 3 요구사항 문장의 형태소 분석 결과

형태소	종류(태그)	형태소	종류(태그)
A	외국어(SL)	되	동사파생접미사(XSV)
가	주격조사(JKS)	면	연결어미(EC)
입력	일반명사(NNG)	C	외국어(SL)
되	동사파생접미사(XSV)	를	목적조사(JKO)
고	연결어미(EC)	출력	일반명사(NNG)
B	외국어(SL)	하	동사파생접미사(XSV)
가	주격조사(JKS)	다	종결어미(EF)
입력	일반명사(NNG)	.	마침표(SF)

표 4는 표 3을 사용하여 문장의 절을 식별한 결과이며, 문장의 절은 일반적으로 연결어미(EC)를 기준으로 나누고, 결과절의 경우 수동문이 능동문으로 변환된다.

표 4 간소화된 문장

절	세부 절	단순화된 문장
원인절	선행절	A가 입력되다
	후행절	B가 입력되다
결과절		C가 출력되다

3.2 C3Tree Model의 생성

요구사항 문장의 간소화 과정은 C3Tree Model로 나타낼 수 있다. 완성된 C3Tree Model은 XMI 코드로 표현된다. 표 4의 문장 간소화 과정을 C3Tree Model로 나타낸 결과는 그림 3과 같다. 총 3개의 Clause 노드가 생성된다.

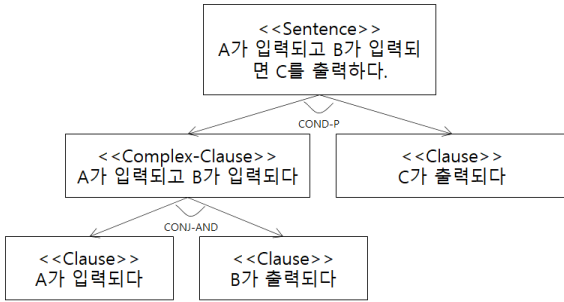


그림 3 C3Tree Model 예제

3.3 자카드 유사도를 이용한 C3Tree Model의 말단 노드 유사도 분석

C3Tree Model이 생성되고, 모델 내의 모든 Clause 노드들 간의 자카드 유사도를 분석한다. 두 문장의 주어, 동사, 목적어가 같고, 수식어 또는 일반 명사에 해당하는 형태소들의 자카드 유사도가 0.6 이상이라면 유사한 문장이라고 판단한다.

표 5는 두 개의 요구사항을 나타낸다. 그림 4는 표 5에 대한 C3Tree Model을 나타내고, 가장 유사도가 높은 두 개의 노드를 표현한다.

표 5 두 개의 요구사항 문장 예제

A가 입력되고 B가 입력되면 C를 출력하다.
D가 입력되고 C가 출력되면 E가 출력된다.

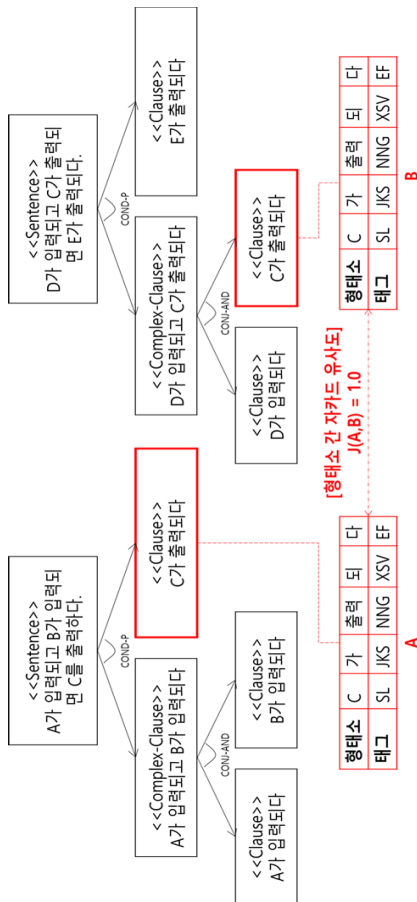


그림 4 자카드 유사도 계산 예제

3.4 C3Tree Model로부터 Cause-Effect Graph 자동 생성

마지막으로 자카드 유사도 지수가 높은 Clause 노드들을 통합 후 Cause-Effect Graph로 변환한다. 완성된 Cause-Effect Graph는 XMI 코드로 표현된다.

그림 5는 그림 4로부터 자동 생성된 Cause-Effect Graph를 나타낸다. C3Tree Model의 Clause 노드 6개가 5개의 Cause-Effect Graph 노드로 표현된다.

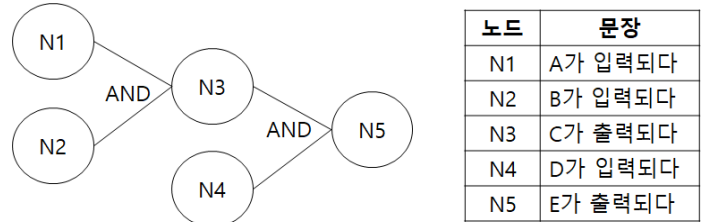


그림 5 Cause-Effect Graph 예제

4. 적용 사례

본 논문에서 제안하는 방법에 대한 적용 사례는 다음과 같다. 표 6은 두 개의 요구사항 예시이다.

표 6 요구사항 1과 요구사항 2

1. 프로그램이 켜지고 사용자가 확인 버튼을 클릭하면 1~10 사이의 난수를 발생시킨다.
2. 1부터 10 사이의 난수가 발생되면, 해당 레벨의 게임을 시작한다.

그림 6과 그림 7은 표 6의 요구사항으로부터 생성된 C3Tree Model이다.

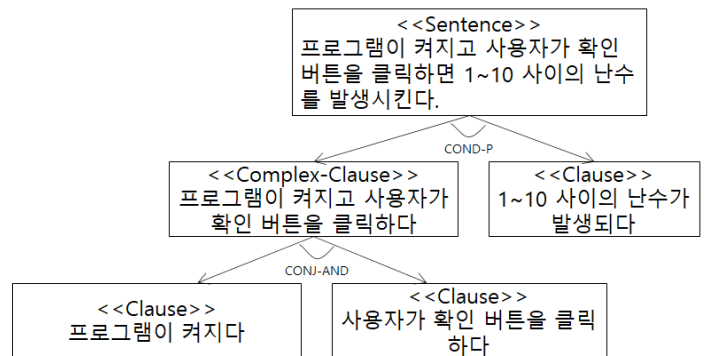


그림 6 요구사항 1의 C3Tree Model

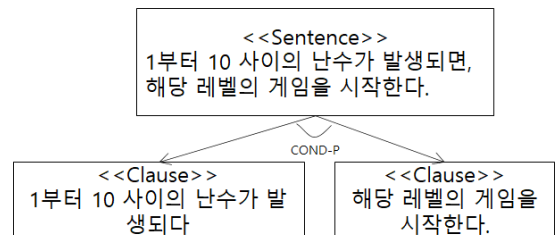


그림 7 요구사항 2의 C3Tree Model

그림 8은 그림 6, 그림 7에서 가장 유사도가 높은 두 개의 노드들의 형태소를 나타낸다. 보조사(JX)와 기호(SY)를 제외하고는 모두 같은 형태소를 가진다.

문장A: 1부터 10 사이의 난수가 발생된다

형태소	1	부터	10	사이	의	난수	가	발생	되	다
태그	SN	JX	SN	NNG	JKG	NNG	JKS	NNG	XSV	EF

문장B: 1~10 사이의 난수가 발생된다

형태소	1	~	10	사이	의	난수	가	발생	되	다
태그	SN	SY	SN	NNG	JKG	NNG	JKS	NNG	XSV	EF

그림 8 두 Clause 노드 간의 형태소 차이

표 7은 그림 8의 두 노드 간 자카드 유사도를 계산한 결과이다. 주어, 동사, 목적어가 같고, 형태소의 자카드 유사도 지수가 0.82이기 때문에 두 문장은 유사하다고 판단한다.

표 7 자카드 유사도 계산 결과

$$J(A,B) = 9 / (10 + 10 - 9) = 0.82$$

C3Tree Model에서 유사도가 높은 노드를 통합하고 Cause-Effect Graph로 변환한 결과는 그림 9와 같다.

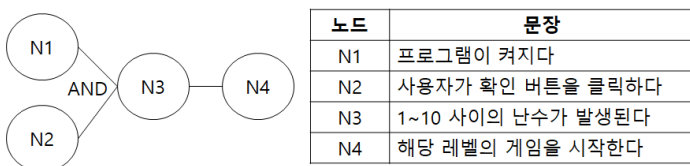


그림 9 생성된 Cause-Effect Graph

5. 결론 및 향후 연구

본 논문은 한글 요구사항으로부터 테스트케이스 생성 과정의 중간 단계이다. 1) 요구사항을 간소화하고, 2) 간소화 과정을 C3Tree Model로 표현하고, 3) C3Tree Model에서 유사도를 측정하여 통합하고, 4) 통합된 결과를 Cause-Effect Graph로 변환하는 방법을 제안한다. 모든 과정은 알고리즘 기반 변환 방법을 사용하고, 모든 모델은 XMI 코드로 저장된다. 그렇기 때문에 요구사항으로부터 Cause-Effect Graph 생성의 자동화가 가능하고, 모델의 XMI 코드 분석을 통해 모델 생성 과정에 대한 자동 추적이 가능하다.

향후 연구로써, 제안한 방법을 프로그램으로 구현하여 자동화하고, Cause-Effect Graph 생성의 정확도를 검증할 것이다.

ACKNOWLEDGMENT

이 논문은 교육부 및 한국연구재단의 4단계 두뇌한국21 사업의 지원(F21YY8102068)과 2021년도 정부(교육부)의 재원으로 한국연구재단의 지원(No.

2021R1I1A305040711, No. 2021R1I1A1A01044060)을 받아 수행된 연구임.

참고문헌

- [1] 권오승, 홍사능, "로그 기반 효과적 반복 테스트", 한국경영정보학회 2009년 추계학술대회, pp. 685-690, 2009.11.
- [2] 장우성, 김영철, "말뭉치 정규화 의미 규칙 기반 요구사항 정제를 통한 원인-결과 그래프 자동 생성", 한국정보처리학회, Vol. 26, No.2, pp.691-693, 2019.11.
- [3] 장우성, 박보경, 김영철, "비정형 요구사항으로부터 생성된 원인-결과 그래프의 검증을 위한 C3Tree 모델 설계", 국방소프트웨어학회, 2019.12.
- [4] Myers, G. L., "The Art of Software Testing", Wiley-Interscience, 1979.
- [5] J Santisteban, J Tejada-Carcamo, "Unilateral Jaccard Similarity Coefficient.", GSB'15, 2015.