

온톨로지를 이용한 선박 통관 프로세스의 유사성 측정

베르나르도 · 박재훈 · 배혜림[†] · 모정관

부산대학교 산업공학과

Similarity Measurement Using Ontology in Vessel Clearance Process

Bernardo N. Yahya · Jae Hun Park · Hye Rim Bae · Jung Kwan Mo

Department of Industrial Engineering, Pusan National University, Busan, 609-735

The demands of complicated data communications have issued a new challenge to port logistics systems. Customers expect ports to handle their generated administrative data while a vessel is docked in a port. One port logistics system, known as the Vessel Clearance Process (VCP), manages large numbers of documents related to port of entry. In the VCP, information flows through many organizations such as the port authority, shipping agents, marine offices, immigration offices, and others. Therefore, for effective management of the Business Process (BP) of the VCP, a standardized method of BP modeling is essential, especially in heterogeneous system environments. In a port, according to port policy, terms and data are sued that are similar to but different from those of other logistics partners, which hinders standardized modeling of the BP. In order to avoid tedious and time-consuming document customization work, more convenient modeling of BP for VCP is essential. This paper proposes an ontology-based process similarity measurement to assist designer for process modeling in port domain, especially VCP. We expect that this methodology will use convenient and quick modeling of port business processes.

Keywords: Vessel Clearance Process, Similarity Measurement, Business Process, Logistic Ontology

1. 서론

RFID(Radio Frequency Identification) 및 프로세스 관리 기술의 발달로 항만물류 환경은 급격한 변화를 이루고 있으며, 이러한 변화로 인해 항만 물류와 관련된 비즈니스 프로세스(BP : Business Process)는 각 항만마다 조금씩 다르게 정의되고 있는 실정이다. 각 항만마다 각기 다른 비즈니스 프로세스는 항만 관리 주체들 간의 유기적인 정보 교환 및 표준 업무 처리를 불가능 하게 하는 요인으로 작용한다. 또한, 각 항만마다 항만 관리 시스템들이 독립적으로 운영되고 있기 때문에 항만 전체적인 관점에서 데이터 통합 및 업무 관리가 효율적으로 이루어지지 못하고 있다. 이러한 업무 프로세스의 미 표준화는 각

기 다른 고객 요구사항과 독립적인 항만 프로세스를 만족시키기 위한 IT 개발 업체의 개발 비용 증가 및 개발 생산 저하에 지대한 영향을 미치게 된다. 이러한 문제점들을 해결하기 위한 목적으로 항만 통합 시스템 개발 및 항만 프로세스의 효율적 관리에 대한 연구들이 활발히 진행되고 있다(Rhee *et al.*, 2007). 항만 통합 시스템 개발을 위한 목적으로 항만 주체들 간의 유기적인 정보통합과 체계적인 관리를 위한 항만 커뮤니티 시스템(PCS : Port Community System)에 대한 관심이 증대되고 있다. PCS는 항만 주체들 간의 정보 교환 및 유기적인 결합을 통해 항만 주체들 간의 정보통합과 체계적인 관리를 위해 중요하게 여기고 있으며 이를 위해서는 BP를 통합하고 이를 효율적으로 관리하는 것이 필수적인 요소로 작용하고 있다. 일반

이 논문은 2009년 국토해양부의 재원으로 한국해양과학기술진흥원의 지원을 받아 수행된 연구임(위크플로우 엔진 기반의 PLUS 개발).

[†] 연락처 : 배혜림 교수, 609-736 부산광역시 금정구 장전동 부산대학교 산업공학과, Tel : 051-510-2733, Fax : 051-512-7603,

E-mail : hrbae@pusan.ac.kr

2011년 4월 29일 접수; 2011년 5월 22일 게재 확정.

적으로 BP를 통합하고 효율적으로 관리하기 위한 목적으로 BPM(Business Process Management) 기법이 널리 적용되고 있으며(Smith and Fingar, 2003), 항만 물류 분야에서도 BPM을 도입에 대한 필요성이 증대되고 있다. 일반적으로 BPM은 기업과 기업 간 또는 고객과 기업 사이에서 일어날 수 있는 BP를 정의하여 가시화하고 업무의 수행과 관련된 사람, 시스템을 프로세스에 맞게 실행/지휘(Orchestration)하며, 전체 업무 프로세스를 효율적으로 관리하고 최적화하여 기업의 가치를 높이는 지속적인 활동으로 정의할 수 있다. 이러한 BPM 기법을 기존의 PCS에 접목시키기 위해서는 항만 프로세스에 대한 정확한 이해가 필요하며 이러한 이해를 바탕으로 BPM에서 관리 가능한 프로세스로의 변환이 필수적이다. 하지만, 항만 프로세스는 그 특성상 정부를 비롯하여 많은 관련 업체들이 유기적으로 업무를 공유하고, 각 항만마다 항만 정책과 법에 따라서 서로 다른 프로세스를 가지고 있기 때문에 정확한 프로세스 정의는 쉬운 작업이 아니다. 또한 각 항만마다 사용하는 용어의 차이로 디자이너에게 혼란을 야기할 수 있다. 예를 들면, 같은 의미의 업무이지만 다른 용어를 사용하는 경우를 볼 수 있다. 실례로 말레이시아 Johor 항만에서는 'Sign On Crew List'와 'Sign Off Crew List'로 구분되어 두 종류의 업무로 사용되고 있지만, Hong Kong 항만에서는 'Crew List'란 이름으로 쓰이고 있다. 이와 같이 동일한 업무임에도 불구하고 각 항만마다 서로 다른 이름으로 업무를 정의하기 때문에 서로 다른 업무로 구별되어 프로세스 표준화를 저해하는 요소가 발생한다.

이처럼 서로 상이한 업무 표현 방식 및 프로세스의 복잡성으로 인해 BPM 기법을 항만 시스템에 적용하기 위한 프로세스 정의 단계에서 많은 시간과 노력을 요구하게 된다. 이러한 요인은 BPM 기법을 항만 시스템에 적용하는데 단점 요인으로 작용하고 있으며, 이를 보완하고 BPM 기법을 항만 시스템에 효과적으로 통합 운영하기 위해서는 먼저 효율적인 프로세스 정의의 기술이 선행되어야 한다. 항만 프로세스의 효율적인 정의를 위해서는 항만 프로세스들 간의 구조적, 의미적 공통 요소를 도출하고, 이를 표준화 하여 이후 프로세스 정의 시 참고 모델로 제시하는 방법을 고려할 수 있다. 하지만 이를 위해서는 항만 프로세스에 대한 정확한 분석과 각 프로세스들 간의 구조적 의미적 유사관계에 대한 분석이 필요하다. 프로세스 정의 시간을 단축하기 위한 목적으로 BPM에 대한 기존 연구들 중 기존에 정의한 프로세스 중 생성하려는 프로세스와 유사한 프로세스를 참고하여 새로이 정의하는 프로세스 유사성 측정 방법들이 제시되었다(Jung et al., 2007; Bae et al., 2007; Jung et al., 2009). 일반적으로 프로세스 유사성 측정 방법은 구조적인 부분을 고려한 것과 의미적인 부분을 고려한 측정 방법으로 구분될 수 있다. 구조적 유사성은 프로세스를 실행하는데 필요한 업무의 순서와 의존 관계를 바탕으로 유사성을 측정하는 방법이다. 반면, 의미적 유사성은 비즈니스 프로세스의 목적과 업무 수행에 관련된 구성요소의 근사한 정도를 측정할 수 있다. 구성요소로는 프로세스와 업무들을 비롯하여, 이를 수행하는 데 사용되는 입출력 데이터와 문서, 업무 수행

자 등 다양한 요소들이 존재하며, 이러한 다양한 요소들을 비교함으로써 프로세스간의 유사성을 측정할 수 있다(Jung et al., 2007). 유사성 측정 방법에 관한 기존 연구들은 구조, 의미 그리고 구조와 의미를 함께 고려한 유사성 측정으로 구분할 수 있다. Bae et al.(2007)은 프로세스 의존 그래프를 분석하여 다른 프로세스 사이의 구조적 유사성 정도를 측정하고, 프로세스 간의 거리 측정을 위한 정량적인 측도를 제시하고 있다. Jung et al.(2009)은 비즈니스 프로세스의 구조적 유사성을 이용한 프로세스 클러스터링 방법을 제안하였다. Ehrig et al.(2007)은 구조와 의미를 고려한 프로세스 유사성 측정을 하고 있으며, 특히 프로세스 요소 이름에 대한 동의어(synonym), 동음이의어(homonym)의 유사성 측정 방법을 통해 의미 유사성을 측정하고 있다.

본 연구에서는 항만 프로세스의 일부인 선박 통관 프로세스 대상으로 구조와 의미를 함께 고려한 프로세스 유사성 측정 방법을 제시하는 것을 목적으로 한다. 이를 위해 항만 프로세스의 액티비티 근접 스코어(APS)를 이용한 구조(structure) 유사성 측정 방법을 제시한다. 또한 의미(semantic) 유사성 측정을 위해 string 값을 고려하기 위한 String similarity와 서로 상이한 용어를 적용하는 항만 프로세스의 효율적 유사성 측정을 위한 온톨로지를 이용한 Taxonomy similarity 방법을 제시하고 이를 고려한 확장된 의미 유사성 측정 방법을 제시한다. 항만 프로세스 온톨로지 표현을 위해 항만 프로세스의 일부인 말레이시아의 Johor 항만의 선박 통관(vessel clearance) 프로세스를 대상으로 하였다. 또한, 이들 구조와 의미를 모두 고려한 프로세스 유사성 측정 방법을 제시하고 선박 통관 업무를 대상으로 유사성 분석을 시행한다.

본 연구는 다음과 같이 구성되어 있다. 제 2장에서는 관련 이론에 대한 내용으로 프로세스 유사성과 온톨로지에 대한 관련 이론을 설명과, 제 3장에서는 프로세스 사이의 유사성을 측정하기 위한 방법으로 본 논문에서 제시하는 구조 유사성과 의미 유사성 방법에 대하여 정의하였으며, 앞서 정의한 구조 유사성과 의미 유사성을 기반으로 프로세스 유사성 측정 방법을 제시한다. 제 4장에서는 항만 프로세스를 대상으로 본 논문에서 제시하는 방법을 적용하여 프로세스 유사성을 측정하고 이전 연구와의 정성적, 정량적 비교를 통해 본 논문과 차별성을 보인다. 그리고 마지막으로 제 5장에서는 본 연구의 결론을 제시한다.

2. 배경 연구

2.1 온톨로지(Ontology)

온톨로지란 원래 철학에서 나온 개념이지만 정보시스템 분야에서는 특정 영역의 개념과 그들 간의 관계를 정의하는 명세로 정의하고 있다(Gruber, 1993). 즉, 온톨로지는 특정분야를 모델링하여 표현하기 위해 사용되는 개념들의 논리적인 집합

으로, 인간 뿐만 아니라 소프트웨어 에이전트 간의 커뮤니케이션 시 특정 단어가 나타내는 개념의 의미를 이해하는데 사용되고 있다. 따라서 온톨로지는 지식의 공유와 재사용을 가능하게 한다. 온톨로지를 구축하는 과정 측면에서 보면, 온톨로지의 구성요소는 개념을 나타내는 클래스와 실체를 의미하는 인스턴스, 그리고 개념을 구성하고 있는 속성(properties), 그리고 클래스와 클래스간의 관계(relationship), 속성들 간의 관계 등을 들 수 있다.

2.2 BPM에서의 프로세스 유사성

프로세스는 업무 흐름과 통제를 컴퓨터를 이용하여 자동으로 실행하기 위해 설계된 모델을 말하며, 프로세스들은 특수한 형태의 방향성 그래프(labeled directed graph)를 이용하여 모델링 된다(WfMC, 1999). 또한 프로세스의 유사성은 크게 두 가지 측면에서 논의될 수 있다. 하나는 의미적 유사성이며, 다른 하나는 구조적 유사성이다. 의미적 유사성은 비즈니스 프로세스의 목적과 업무 수행에 관련된 구성요소의 근사한 정도를 측정할 수 있다. 반면에 구조적 유사성은 프로세스를 자동으로 실행하는데 필요한 업무의 순서와 의존관계를 바탕으로 수행될 수 있다(Jung *et al.*, 2007). 프로세스를 정의하는 방법은 크게 방향성 그래프(directed graph)방법과 블록 구조(block structured)방법으로 분류되는데, 방향성 그래프에 기반한 방법은 사용자의 모델링 및 이해 용이도가 뛰어나고 표현능력이 풍부하

여 비즈니스 프로세스 디자인 시스템에서 많이 사용된다(Shapiro, 2002). 이에 비하여 블록 구조 방법은 흐름 통제가 명확하고 자동화하기가 용이하여 최근 전자거래를 위한 BPEL(Business Process Execution Language), BPML(Business Process Modeling Language)과 같은 비즈니스 프로세스 표준 언어에서 많이 사용된다.

3. 유사성 측정 방법

본 장에서는 항만 커뮤니티 시스템의 일부로 말레이시아 Johor 항만의 선박 통관 프로세스를 대상으로 프로세스 유사성 측정 방법에 대해 설명한다.

Johor 항만의 선박 통관 프로세스는 단일 창구 전자 항만 통관 시스템으로 선사 정부 대리인의 선박 통관 및 승인 작업을 제어하는 것을 목적으로 한다. 선박 통관 프로세스는 다음 <Figure 1>과 같이 선박 정보(Vessel Information), 문서(Document) 그리고 업무흐름(Workflow)과 같은 3가지 단위 업무로 구성되어 있다. 먼저 선박 정보 업무는 선박의 제원 정보 및 입항 후 스케줄(schedule)에 대한 정보를 관리한다. 다음으로 문서 업무는 승원, 승객 그리고 밀항자에 대한 문서를 포함하여 선박이 입항 및 출항하는데 필요한 관련 문서에 대한 정보를 관리한다. 마지막으로 업무흐름에서는 정부 부서 및 관련 회사별 업무 절차 흐름을 관리한다.

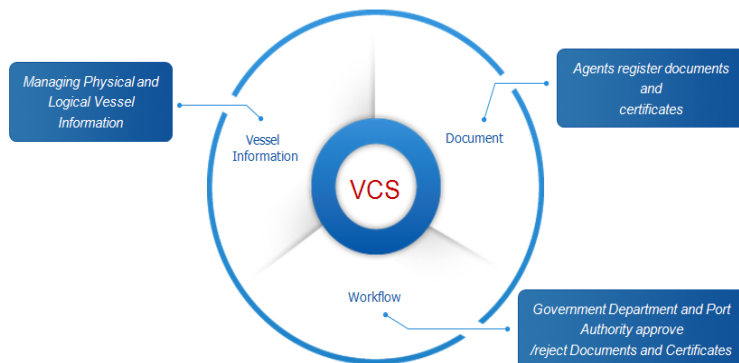


Figure 1. Structure of vessel clearance

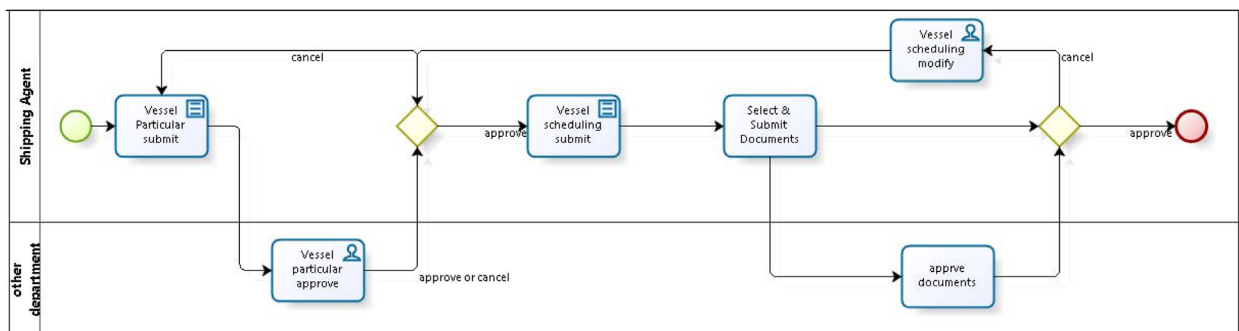


Figure 2. Process model of vessel clearance

선박 통관 프로세스를 BizAgi라는 프로세스 모델링 툴을 이용하여 표현하면 다음 <Figure 2>와 같이 나타낼 수 있다.

선박 회사(Shipping Agent)에서 선박 및 선원과 관련한 제원 정보를 생성하여 제출(Vessel particular submit)하면 관련 부서에서 이를 승인(Vessel particular approve)하게 된다. 승인이 되고 나면 선박회사는 다시 선박이 입항 후 하게 되는 업무 등과 관련한 스케줄 정보(Vessel scheduling submit)와 각 부서별로 승인이 필요한 문서를 제출(Select and submit documents)한다. 제출된 문서는 해당 부서에서 승인(approve documents)을 하게 되고 모든 승인이 확인되면 해당 프로세스는 완료된다. 그렇지 않고 하나의 부서라도 완료 승인을 하지 않을 경우 문서는 수정(Vessel scheduling modify)되고 스케줄 정보 제출(Vessel scheduling submit)부터 다시 실행되게 된다. 승인과 관련한 부서는 일반적으로 세관, 출입국 사무소, 보건국, 항만공사가 있을 수 있고 각 항만의 특성에 따라 화학물품 취급 관리소, 정부 해양부서 등이 추가로 존재할 수 있다. 해당 부서에서 승인하는 업무는 하위 단위 업무를 포함하고 있으며, 7개의 승인 부서(Immigration, Healthy, Chemistry, Marine, Customer, Johor port authority, JPB marine)에서 배의 제원 정보, 승원 정보, 세관 승인 및 입항과 관련된 문서들을 확인하여 승인하도록 구성되어 있다. 하지만 이는 Johor 항만에 적용되는 프로세스이고 각 국가마다

유사한 업무임에도 불구하고 상이한 프로세스 구조와 용어를 사용하고 있다. 아래 <Figure 3>은 선박통관 프로세스에서 두 항만(항만 A와 항만 B)의 프로세스 예제를 보여주고 있다.

두 항만은 선사의 편의를 위해 또는 항만의 규정에 따라 두 항만이 서로 다른 프로세스를 가지고 있음을 보여주고 있으며, 같은 의미이지만 다른 용어를 사용하는 것을 발견할 수 있다. <Figure 3>(a)는 주요항만들의 메인 선박통관 프로세스를 나타내며, <Figure 3>(b)는 항만 A에서는 다루지 않는 서류(Dutiable Clearance)를 항만 B에서만 다루고 있는 프로세스의 예제를 나타낸다. <Figure 3>(c)와 <Figure 3>(e)는 같은 서류(Health Clearance)를 각각의 항만 A와 항만 B에서 다루는 프로세스 예제이다. 즉 두 프로세스는 각 항만에서 다른 이름으로 사용되고 있지만 같은 특성을 가진 같은 서류임을 알 수 있다. 이와 달리 <Figure 3>(d)와 <Figure 3>(f)는 이름이 다른 단위 업무에 의해서 같은 서류(Ship Security)를 다루는 두 프로세스를 보여주고 있다. 이는 같은 서류이지만 다른 용어를 사용하거나, 유사한 용어를 사용함에 따라 단위 업무 속성에 따른 프로세스 모델링 시 혼돈을 야기할 수 있는 부분이 된다. 이러한 혼돈을 줄이기 위한 방법으로 온톨로지를 이용한 의미 유사성을 고려할 수 있다. 다음 <Figure 4>는 선박 통관 시스템에 대한 온톨로지를 나타낸다.

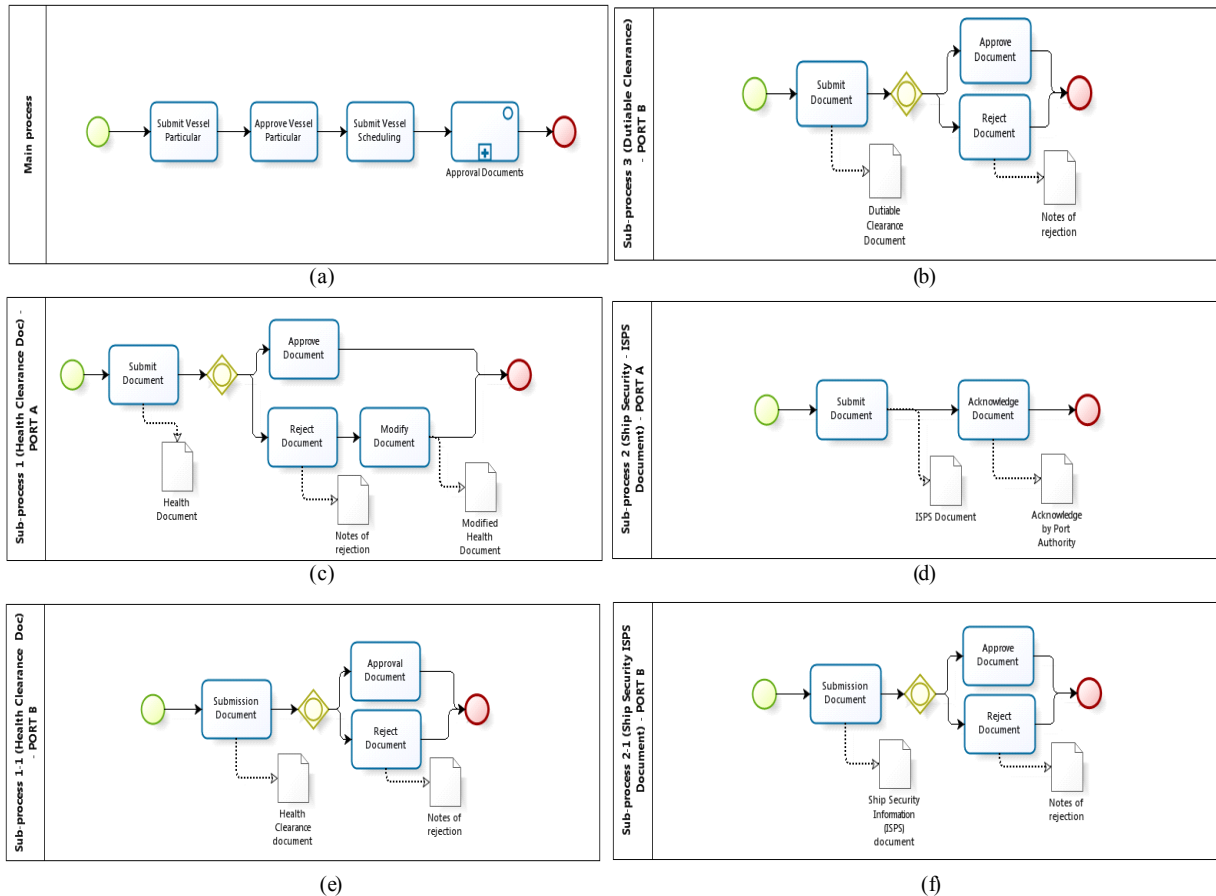


Figure 3. Example processes of vessel clearance in Port A and B

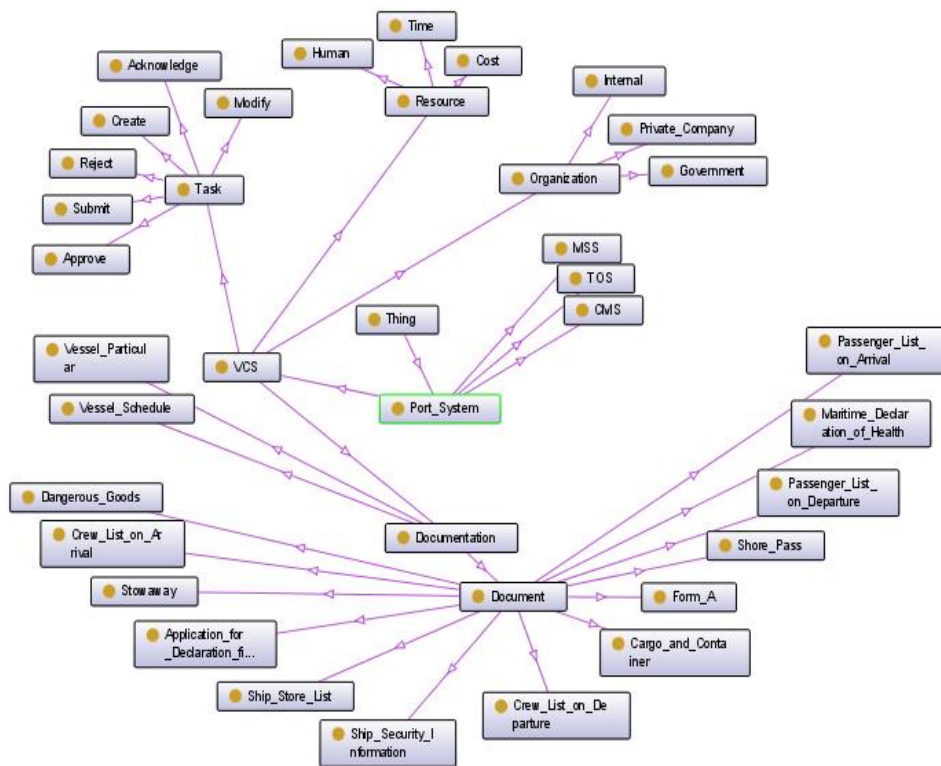


Figure 4. Ontology of vessel clearance process

다음 장에서는 구조적 유사성 측정과 온톨로지를 이용한 의미적 유사성 측정에 대한 새로운 방법과 두 방법을 함께 고려한 유사성 측정 방법에 대해 설명한다.

3.1 구조 유사성(Structure similarity)

본 연구에서 사용하는 프로세스 모델은 Yahya *et al.*(2009)에서 제시한 다음의 프로세스 모델 표기법을 따른다.

정의 1 : 프로세스 모델(Process Model)

프로세스는 다음과 같이, 유방향 그래프 $p = (A, L)$ 와 분기 및 병합을 위한 라벨링 함수 $f(0)$ 로 정의된다.

- $A = \{a_i | i = 1, \dots, l\}$ 는 단위 업무 집합이며, a_i 는 i 번째 단위 업무이고, l 는 단위 업무의 전체 수이다.
- $L = \{l_{ij} = (a_i, a_j) | a_i, a_j \in A\}$ 는 링크의 집합이며, l_{ij} 는 두 단위 업무 a_i 와 a_j 사이의 연결을 나타낸다. (a_i, a_j) 는 a_i 는

a_j 를 바로 앞선다는 것을 의미한다.

- a_{i+} 는 a_i 다음으로 오는 단위 업무이고, a_{i-} 는 a_i 앞에 오는 단위 업무이다.
- a_i 로부터 분기된 단위 업무 집합을 의미하는 SA_i 는 $SA_i = \{a_{i+} | (a_i, a_{i+}) \in L\}$ 와 같이 정의되며, $|SA_i| > 1$ 일 때, 만약 모든 a_{i+} 들이 실행되어지면 $f(a_{i+}) = 'AND'$ 이고, 그 외에는 $f(a_{i+}) = 'XOR'$ 이다.
- a_i 까지 병합된 단위업무 집합을 의미하는 MA_i 는 $MA_i = \{a_{i-} | (a_{i-}, a_i) \in L\}$ 와 같이 정의되며, $|MA_i| > 1$ 일 때, 만약 모든 a_{i-} 들이 실행되어지면 $f(a_{i-}) = 'AND'$ 이고, 그 외에는 $f(a_{i-}) = 'XOR'$ 이다.

정의 1에 의해 다음 <Figure 5>와 같은 프로세스 모델을 도출 가능하다. <Figure 5>는 6개의 단위 업무로 구성되었으며, 각 단위 업무는 \rightarrow 에 의해 선후 연결 관계가 정의되어 있다.

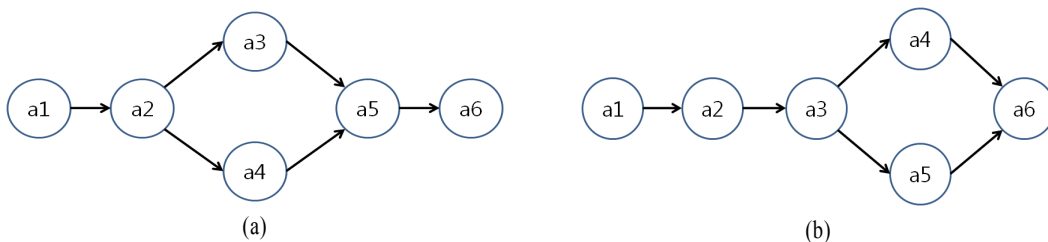


Figure 5. Example process

앞서 언급한 것과 같이 구조 유사성은 각 단위 업무 간의 구조적 선후 관계에 따라 각 단위 업무 간의 근접 값을 기반으로 유사성을 분석하는 것으로, 본 연구에서는 각 단위 업무간의 거리를 계산하기 위해 다음과 같이 액티비티 근접 스코어(APS : Activity Proximity Score)를 이용한 구조 유사성을 정의한다. APS는 각 단위 업무들 간의 근접성에 기반한 거리 값으로 다음 식 (1)로부터 구할 수 있다.

$$d_{ij}^k = \frac{h^k(i, j)}{d_{ij}^k} \quad (1)$$

식 (1)에서 d_{ij}^k 는 k번째 프로세스의 i번째 단위 업무와 j번째 단위 업무와의 APS를 의미한다. 만약 k번째 프로세스에서 i번째 단위 업무에서 j번째 단위 업무로 도달가능하면 $h^k(i, j)$ 는 1이라 정의하고, 아니면 0으로 정의한다. d_{ij}^k 는 i번째 단위 업무와 j번째 단위 업무사이의 거리 값으로 가장 인접한 두 단위 업무간의 거리 값은 1로 가정한다. 예를 들면, <Figure 6>에서 단위 업무 a1에서 a2와의 거리 값은 1이고 a1과 a3과의 거리 값은 2가 된다. 각각의 프로세스는 d_{ij}^k 의 단일 값을 가지고 있다. 만약 k번째 프로세스에서 i번째 단위 업무와 j번째 단위 업무 사이에 관계가 없다면 $d_{ij}^k = 0$ 이 된다.

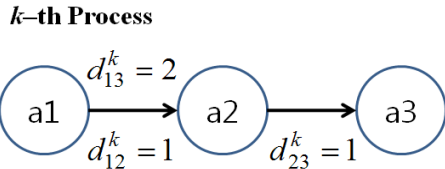


Figure 6. Example process for the APS

다음 <Table 1>은 <Figure 5>의 APS 계산 결과를 나타낸다.

APS를 이용하여 두 프로세스 모델 사이의 구조 유사성은 다음 식 (2)로부터 구할 수 있다.

$$sim_{st}(p_k, p_l) = 1 - \frac{\sum_i \sum_j |d_{ij}^k - d_{ij}^l|}{\max(\|p_k\|, \|p_l\|)}, \text{ where } k, l \in k \quad (2)$$

Table 1. Results of APS in <Figure 5>

(a)	TO						
		a1	a2	a3	a4	a5	a6
F R O M	a1	0	1/1	1/2	1/2	1/3	1/4
	a2	0	0	1/1	1/1	1/2	1/3
	a3	0	0	0	0	1/1	1/2
	a4	0	0	0	0	1/1	1/2
	a5	0	0	0	0	0	1/1
	a6	0	0	0	0	0	0

식 (2)에서 $|d_{ij}^k - d_{ij}^l|$ 는 k번째와 l번째 프로세스 사이에서 i번째 단위 업무와 j번째 단위 업무의 APS 차에 대한 절대 값을 의미한다. 그리고 $\|p_k\|$ 는 k번째 프로세스 내에서 도달 가능한 ($a_i \rightarrow a_j$) 단위 업무 쌍의 수를 의미하는 것으로 $\max(\|p_k\|, \|p_l\|)$ 는 k번째 또는 l번째 프로세스 중에서 도달 가능한 최대 단위 업무 수를 의미한다. 만약 k번째와 l번째 프로세스의 유사성 측정값이 1과 같다면 구조는 동일한 것이다. <Figure 1>(a)와 <Figure 1>(b) 두 프로세스에 대한 구조 유사성을 구할 경우 <Figure 1>(a)와 <Figure 1>(b)를 각각 p_1 과 p_2 로 두고 계산한 결과는 다음 <Table 2>와 같다.

Table 2. Results of structure similarity

(p1, p2)	TO						
		a1	a2	a3	a4	a5	a6
FROM	a1	0	0	0	0.17	0	0
	a2	0	0	0	0.5	0	0
	a3	0	0	0	0	0	0
	a4	0	0	0	0	1	0.5
	a5	0	0	0	0	0	0
	a6	0	0	0	0	0	0

$sim_{st}(p_1, p_2) = 1 - (3.17/14) = 0.77$

3.2 의미 유사성(Semantic similarity)

일반적으로 단위 업무에 적용되는 의미는 단위 업무 이름을 비롯하여 각 단위 업무에 포함된 속성 등 다양하게 존재 할 수 있다. 본 장에서는 프로세스 모델링에서는 같은 의미지만 다른 용어를 사용하므로 발생하는 문제를 해결하고 보다 현실적인 의미 유사성 측정을 위해, String 조합에 따른 유사성 측정과 더불어 각 단위 업무의 용어 사이의 연관 관계를 고려한 유사성 측정 방법을 제시한다. 이를 위해 단위 업무의 의미들을 단순히 String 관점에서 접근하여 유사성을 측정하는 String similarity와 단위 업무 각 속성 값들의 분류 및 연관 관계에 따른 유사성을 측정하는 온톨로지 기반의 Taxonomy similarity를 제

(b)	TO						
		a1	a2	a3	a4	a5	a6
F R O M	a1	0	1/1	1/2	1/3	1/3	1/4
	a2	0	0	1/1	1/2	1/2	1/3
	a3	0	0	0	1/1	1/1	1/2
	a4	0	0	0	0	0	1/1
	a5	0	0	0	0	0	1/1
	a6	0	0	0	0	0	0

시한다. 또한, 이 두 가지 방법을 고려하여 최종적으로 의미 유사성을 측정할 수 있는 모델을 제시한다.

String similarity는 각 단위 업무에 포함된 String을 비교하므로 유사성을 측정할 수 있는데, 여기서 String이라 함은 단위 업무의 이름, 역할자 이름, 담당 부서, 관련된 서류 이름 등이 될 수 있다. 본 연구에서는 영어를 기반으로 하는 String을 고려한다. String은 철자들로 구성되어 있으며, 이처럼 String을 구성하는 철자들 간의 차이를 비교하므로 String similarity를 구할 수 있다. 이를 위해 다음과 같이 String Edit Distance(SED)를 정의한다.

정의 2 : String Edit Distance(SED)

str_i 와 str_j 는 a_i 와 a_j 의 단위 업무의 의미를 나타내는 두 개의 명칭을 의미하고, String x 의 길이를 $|x|$ 라고 할 경우 SED를 다음과 같이 정의한다.

$$SED^r(a_i, a_j) = \left\{ 1 - \frac{ed(\Omega^r(a_i), \Omega^r(a_j))}{\max(|\Omega^r(a_i)|, |\Omega^r(a_j)|)} \right\} \quad (3)$$

$\Omega : a_i \rightarrow str_i$ 에서 Ω 는 string 라벨에 단위 업무의 명칭을 연관시켜 주는 것이다. 예를 들면, $\Omega(a_1) = \text{“Send document”}$ 일 경우, 단위 업무 a_1 의 이름은 “Send document”를 의미한다. $ed(str_i, str_j)$ 는 str_i 에서 str_j 로 또는 이와 반대로 포함된 철자들 중 문자의 삽입, 삭제 혹은 변형을 요구하는 atomic string operations의 최대 수를 의미한다. 예를 들면, str_1 이 “Send document”이고, str_2 가 “Submit document”일 경우, $ed(str_1, str_2)$ 는 “Submit”에 기반 해서 5가 된다. r 은 전체 단위 업무 속성 중에서 해당하는 속성을 의미하며, R 은 고려된 단위 업무 전체 속성 수를 나타낸다.

앞서 정의한 SED값을 이용하여 String similarity를 측정한다. 프로세스간 단위 업무 사이의 String similarity는 각 속성의 SED 값에 고려된 단위 업무 전체 속성 수를 나누는 것으로 다음 식 (4)에 의해 측정될 수 있다.

$$sim_{sr}(a_i^k, a_i^l) = \frac{\sum^R SED^r(a_i^k, a_i^l)}{R}, \quad k \neq l \quad (4)$$

식 (4)에서 a_i^k, a_i^l 는 각각 k 와 l 번째 프로세스에서 i 번째 단위 업무를 의미한다.

Taxonomy similarity는 온톨로지를 활용하여 유사성을 측정하는 것이다. 본 연구에서는 Wu and Palmer(1994)의 측정 방법을 수정하여, 다음과 같은 Taxonomy similarity 식을 정의하였다.

$$sim_{ta}(a_i^k, a_i^l) = \frac{2 \times N_3}{(N_1 + N_2 - 2) + 2 \times N_3}, \quad k \neq l \quad (5)$$

N_1 과 N_2 는 그들의 공통 슈퍼클래스에서 각각 a_i^k 로부터 IS-A 연결의 수와 a_i^l 까지의 IS-A 연결의 수이다. N_3 는 공통 슈퍼클래스로부터 taxonomy의 루트까지 IS-A 연결들의 수이다. 예를 들면, <Figure 7> (a)에서 “Send document”와 “Submit document”의 공통 슈퍼클래스는 Task이다. 따라서 이들 속성의 특징들은 $N_1 = 1, N_2 = 1, N_3 = 3$ 이고, 식에 대입하여 계산해 보면 $sim_{ta} = 1$ 이다. <Figure 7> (b)는 $N_1 = 3, N_2 = 3, N_3 = 1$ 이며, 식에 넣어보면 $sim_{ta} = 1/3$ 이 된다.

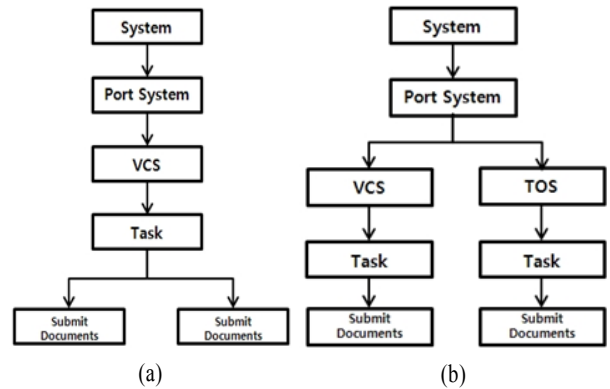


Figure 7. Taxonomy similarity measurement

String similarity와 Taxonomy similarity를 활용한 의미 유사성은 다음 식 (6)에 의해 측정될 수 있다.

$$sim_{sm}(p_k, p_l) = \frac{\sum_i^{\max(|A^k|, |A^l|)} sim_{sr}(a_i^k, a_i^l) \times sim_{ta}(a_i^k, a_i^l)}{\max(|A^k|, |A^l|)}, \quad (6)$$

$k \neq l$

식 (6)에서 $|A^k|$ 는 k 번째 프로세스에서 전체 단위 업무 수이다. $\max(|A^k|, |A^l|)$ 는 k 번째 프로세스와 l 번째 프로세스 사이의 최대 단위 업무 수를 의미한다.

3.3 프로세스 유사성(Process similarity)

본 장에서는 앞서 제시한 단위 업무의 구조 유사성과 의미 유사성 방법을 고려한 프로세스 유사성 측정방법을 제시한다. 식 (7)은 본 연구에서 제시하는 프로세스 유사성에 대한 모델로, 구조 유사성과 의미 유사성에 측정자의 선호도 가중치(α)를 부여하여 유사성을 측정하도록 하였다. α 값은 0과 1사이의 값으로 주어 질 수 있으며, 0에 가까운 값일수록 의미 유사성에 비중을 두고 1에 가까울수록 구조 유사성에 비중을 두어 프로세스 유사성을 측정하도록 정의하였다.

$$sim_{sm}(p_k, p_l) = \alpha \times sim_{st}(p_k, p_l) + (1 - \alpha) \times sim_{sm}(p_k, p_l), \quad k \neq l \quad (7)$$

4. 유사성 측정 및 비교

앞서 제 3장에서 정의한 구조 유사성과 의미 유사성을 이용하여 선박통관 프로세스를 대상으로 유사성을 측정한다. 총 10개의 유사한 선박통관 프로세스를 비교 대상으로 하였고, 각 프로세스는 8개의 단위 업무로 구성된다. 각 단위 업무는 12개의 각기 다른 이름들(create document, select document, submit document, receive document, check document, acknowledge document, reject document, hold document, approve document, modify document, temporary approve document, review document)로 구성되고, 각 단위 업무의 수행자로 항만과 관련 있는 6개 정부기관(Immigration office, Health office, Chemistry office, Marine office, Customs office, Johor Port Authority)과 6개의 선사 포함하였다. 본 장에서 비교하기 위한 선박 통관 프로세스의 그래프 모형은 다음 <Figure 8>과 같다.

기준 프로세스 P1과 각 프로세스간 구조에 따른 유사성을 측정한 결과는 다음 <Table 3>과 같다. 기준 프로세스와 구조적으로 유사할수록 유사성 값이 높는데, 구조 유사성 측면에서 P1과 가장 유사한 프로세스는 P5가 된다.

Table 3. Result of structure similarity between P1 and other processes

	P2	P3	P4	P5	P6	P7	P8	P9	P10
P1	0.494	0.663	0.588	0.734	0.683	0.645	0.605	0.500	0.589

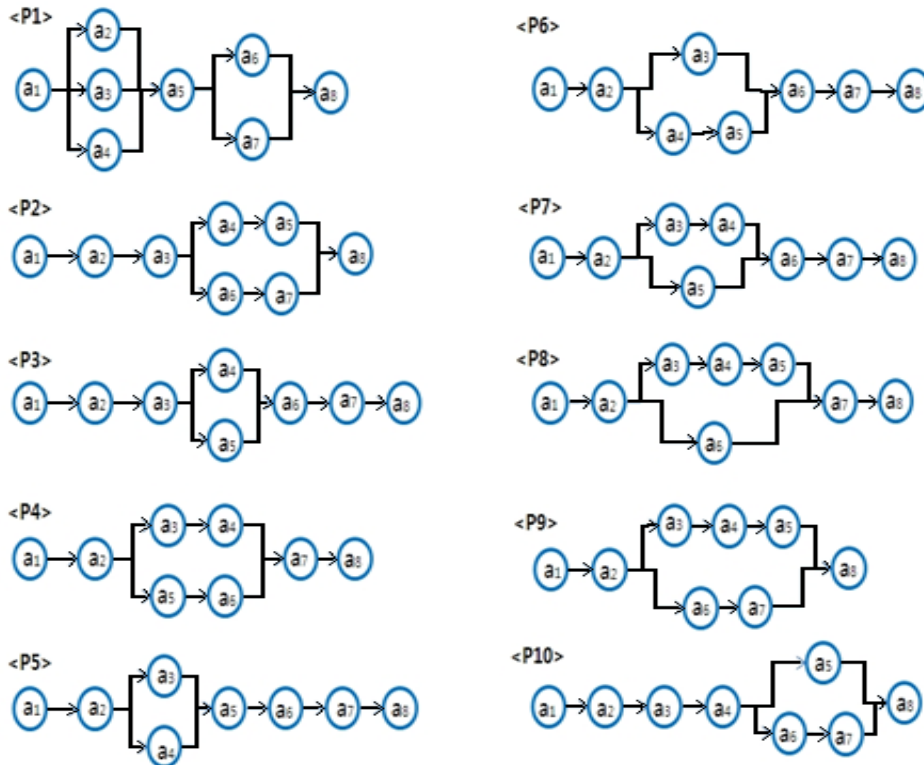


Figure 8. Graph structure of the vessel clearance process

다음으로 의미 유사성을 측정하기 위해 프로세스에서 단위 업무의 이름을 고려하여 P1과 다른 프로세스와의 String similarity 값을 측정된 결과는 다음 <Table 4>와 같다. String similarity와 더불어 Taxonomy similarity를 측정하기 위해 다음 <Figure 9>와 같이 실제 항만 정보를 활용하여 만든 선박통관 시스템의 항만 온톨로지를 이용하여 측정하였다. 이를 바탕으로 측정된 Taxonomy similarity 값을 구하면 <Table 5>와 같고, String similarity와 Taxonomy similarity를 고려한 의미 유사성 값은 <Table 6>과 같다. 의미 유사성 측면에서는 P1과 가장 유사한 프로세스는 P10이 된다. 이는 단지 구조 유사성만 고려한 모델링이나, 의미 유사성만을 고려한 모델링은 각기 다른 프로세스가 유사성이 가장 높게 나왔음을 알 수 있다. 즉 모델링 수행자는 단지 구조 유사성만 또는 의미 유사성만을 활용한 모델링보다는 구조 유사성과 의미 유사성을 모두 고려한 프로세스 모델링이 보다 많은 정보를 디자이너에게 제공하여 보다 유용한 모델링을 수행을 가능하게 한다. 구조 유사성과 의미 유사성을 고려한 프로세스 유사성을 구하기 위해 α 값을 0부터 1까지 달리 하여 P1과의 유사성 순위를 분석하였고, 결과는 <Table 7>과 같다.

<Table 7>의 결과를 통해 구조와 의미 유사성에 동일한 선호도 가중치($\alpha = 0.5$)를 부여한 경우 P1과 가장 유사한 프로세스는 P5가 됨을 알 수 있다. 또한, α 값이 0에 가까울수록 의미 유사성이 높은 프로세스가 선택되고 1에 가까울수록 구조 유사성이 높은 프로세스가 선택됨을 알 수 있다.

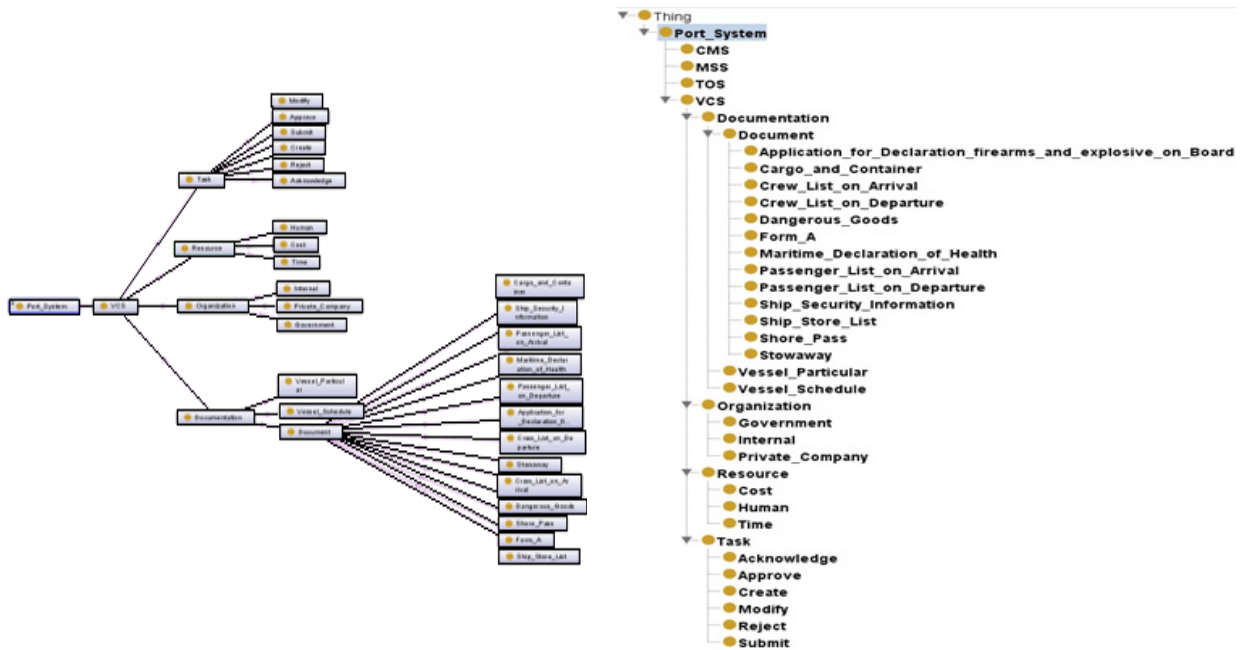


Figure 9. Ontology of vessel clearance system

Table 4. Result of string similarity between P1 and other processes

	a1	a2	a3	a4	a5	a6	a7	a8
P1&P2	1.00	1.00	0.67	0.63	0.45	1.00	0.45	1.00
P1&P3	1.00	1.00	0.67	0.56	0.56	1.00	0.50	1.00
P1&P4	1.00	1.00	0.56	0.60	0.56	0.60	0.50	1.00
P1&P5	1.00	1.00	0.63	0.56	0.56	1.00	0.50	1.00
P1&P6	1.00	1.00	0.67	0.60	0.56	0.64	0.50	1.00
P1&P7	1.00	1.00	0.60	0.60	0.56	0.60	0.50	1.00
P1&P8	1.00	1.00	0.60	0.60	0.56	0.67	0.50	1.00
P1&P9	1.00	1.00	0.60	0.36	0.56	0.60	1.00	1.00
P1&P10	1.00	1.00	1.00	1.00	0.45	1.00	0.50	1.00

Table 5. Result of taxonomy similarity between P1 and other processes

	a1	a2	a3	a4	a5	a6	a7	a8
P1&P2	1.00	0.80	0.80	0.50	0.80	0.80	0.80	1.00
P1&P3	1.00	1.00	1.00	0.50	0.80	0.80	0.80	1.00
P1&P4	1.00	1.00	0.50	0.50	0.80	0.50	0.80	1.00
P1&P5	1.00	1.00	0.50	0.80	1.00	1.00	1.00	1.00
P1&P6	1.00	0.80	0.80	0.80	0.50	0.80	0.80	1.00
P1&P7	1.00	0.80	0.50	0.50	0.80	0.50	0.80	1.00
P1&P8	1.00	0.80	0.50	0.50	0.50	0.80	0.80	1.00
P1&P9	1.00	0.80	0.50	0.50	0.80	0.80	0.80	1.00
P1&P10	1.00	0.80	0.80	0.80	0.80	0.80	0.80	1.00

Table 6. Result of semantic similarity between P1 and other processes

	a1	a2	a3	a4	a5	a6	a7	a8	Semantic similarity
P1&P2	1.00	0.80	0.53	0.31	0.36	0.80	0.36	1.00	0.65
P1&P3	1.00	1.00	0.67	0.28	0.45	0.80	0.40	1.00	0.70
P1&P4	1.00	1.00	0.28	0.30	0.45	0.30	0.40	1.00	0.59
P1&P5	1.00	1.00	0.31	0.45	0.56	1.00	0.50	1.00	0.73
P1&P6	1.00	0.80	0.53	0.48	0.28	0.51	0.40	1.00	0.63
P1&P7	1.00	0.80	0.30	0.30	0.45	0.30	0.40	1.00	0.57
P1&P8	1.00	0.80	0.30	0.30	0.28	0.53	0.40	1.00	0.58
P1&P9	1.00	0.80	0.30	0.18	0.45	0.48	0.80	1.00	0.63
P1&P10	1.00	0.80	0.80	0.80	0.36	0.80	0.40	1.00	0.75

5. 결론

본 연구에서는 항만 커뮤니티 시스템(PCS : Port Community System)의 선박 통관 프로세스(VCP : Vessel Clearance Process)를 대상으로 구조(structure) 유사성과 의미(semantics) 유사성을 고려한 프로세스 유사성을 측정하는 방법을 제시하였다. 이는 BPM 기법을 적용하기 위한 항만 시스템에서 프로세스 모델링을 수행할 때 구조적이고 의미적으로 유사한 프로세스를 찾아줌으로써, 기존의 유사한 프로세스로부터 새로운 프로세스를 쉽게 모델링할 수 있도록 도움을 주는데 초점을 맞추고 있다. 본 논문에서 제시하는 프로세스 유사성은 항만마다 상이한 용어를 적용 하므로 동일한 구조와 의미임에도 전혀 다른 프로

Table 7. Ranking of process similarity based on the α

순위	$\alpha = 0$	$\alpha = 0.1$	$\alpha = 0.2$	$\alpha = 0.3$	$\alpha = 0.4$	$\alpha = 0.5$	$\alpha = 0.6$	$\alpha = 0.7$	$\alpha = 0.8$	$\alpha = 0.9$	$\alpha = 1.0$
1	P10	P10	P5	P5	P5	P5	P5	P5	P5	P5	P5
2	P5	P5	P10	P10	P3	P3	P3	P3	P6	P6	P6
3	P3	P3	P3	P3	P10	P10	P6	P6	P3	P3	P3
4	P2	P6	P6	P6	P6	P6	P10	P10	P7	P7	P7
5	P6	P2	P2	P2	P7	P7	P7	P7	P10	P10	P8
6	P9	P9	P9	P7	P4	P8	P8	P8	P8	P8	P10
7	P4	P4	P4	P4	P8	P4	P4	P4	P4	P4	P4
8	P8	P8	P7	P9	P2	P2	P2	P2	P9	P9	P9
9	P7	P7	P8	P8	P9	P9	P9	P9	P2	P2	P2

세스로 인식되는 기존의 유사성 측정 방법을 보완하기 위해 선박 통관 프로세스를 대상으로 온톨로지 기반의 의미 유사성을 측정하는 방법을 제시하였다. 또한, 구조적 유사성과 의미적 유사성에 각각 다른 선호도 가중치를 부여하므로, 프로세스 디자이너의 의도를 반영한 프로세스 유사성 측정 방법을 제시하였다.

본 연구는 다음과 같은 장점을 지닌다. 첫째, 프로세스 디자이너가 각 항만마다 각기 다른 프로세스와 특성(property)을 고려하여 모델링을 수행할 때, 프로세스의 구조와 특성을 재사용할 수 있어서 모델링에 도움을 줄 수 있다. 둘째, 항만 도메인 지식이 없어 프로세스 모델링 시 혼돈을 줄 수 있는 용어의 차이를 온톨로지를 이용한 의미 유사성 측정을 통해 혼돈을 줄일 수 있다. 셋째, 의미 유사성 측정 방법에서 단위 업무 수행자(performer), 서류의 이름(name) 등을 고려하므로 구조적인 것뿐만 아니라 의미적인 부분을 고려하여 보다 많은 프로세스 정보를 제공할 수 있을 것으로 기대한다.

참고문헌

- Rhee, S. H., Cho, N. W., and Bae, H. (2007), A More Comprehensive Approach to Enhancing Business Process Efficiency, *Human Interface and the Management of Information*, Lecture Notes in CS, **4558**, 955-964.
- Smith, H. and Fingar, P. (2003), Business Process Management : the Third Wave, *Meghan-Kiffer*, Tampa, 2003.
- Jung, J., Bae, J., and Kang, S. (2007), workflow clustering methodology using structural similarity metrics, *Journal of the Korea Institute of Industrial Engineers*, **33**(1), 99-109.
- Bae, J., Liu, L., Caverlee, J., Zhang, L. J., and Bae, H. (2007), Development of Distance Measures for Process Mining, Discovery and Integration, *International Journal of Web Services Research*, **4**(4), 1-17.
- Jung, J., Bae, J., and Liu, L. (2009), Hierarchical Clustering of Business Process Models, *International Journal of Innovative Computing, Information and Control*, **5**(12), 4501-4511.
- Ehrig, M., Koschmider, A. and Oberweis, A. (2007), Measuring Similarity between Semantic Business Process Models, *Proceedings of the fourth Asia-Pacific conference on Conceptual modeling (APCCM 2007)*, Australia, **67**, 71-80.
- Gruber, T. R. (1993), A Transition Approach to Portable Ontology Specification, *Knowledge Acquisition*, **5**(2), 199-220.
- WfMC. (1999), Terminology and Glossary, WfMC-TC-1011, Workflow Management Coalition, Hampshire, United Kingdom.
- Shapiro, R. (2002), A Comparison of XPD, BPML and BPEL4WS, *Cape Visions*, White paper, 2002.
- Yahya, B. N., Bae, H., Bae, J. (2009), Process Design Selection using Proximity Score Measurement, 3rd International Workshop on Collaborative Business Processes(CBP), Germany.
- Wu, Z. and Palmer, M. (1994), Verb semantics and lexical selection, *Processing 32rd Annual Meeting of the Associations for Computational Linguistics*, 133-138.