

보안구역 위치정보 관리를 통한 ECM기반 전자문서유출방지 시스템 구현방안 연구

유갑상* · 조승연** · 황인태***

A Study on Implementation Method of ECM-based Electronic Document Leakage Prevention System through Security Area Location Information Management

Gab-Sang Yoo* · Seung-Yeon Cho** · In-Tae Hwang***

■ Abstract ■

The current technology drain at small and medium-sized enterprises in Korea is very serious. According to the National Intelligence Service's survey data, 69 percent of technology leaks are made through employees of small and medium-sized enterprises. A document security system was introduced to compensate for the problem. However, small and medium-sized enterprises are not doing well due to their poor environment. Therefore, it proposes a document security system suitable for small businesses by developing a location information machine learning system that automatically creates a document security Green Zone through learning, and an ECM-based electronic document leakage prevention system that manages generated Green Zone information by reflecting it into the document authority system. And step by step, propose a universal solution through cloud services..

Keyword : Technology Leakage, Document Security, Machine Running, Ecm, Electronic Documentation, Leakage Prevention System, Cloud, Algorithm

1. 서론

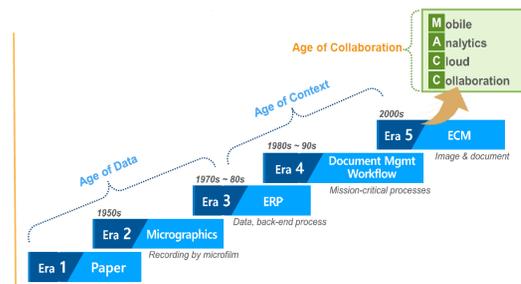
문서보안 안전 구역(Green Zone)을 자동으로 생성하는 위치정보 머신러닝 시스템과, 생성된 보안 안전 구역(Green Zone) 정보를 문서관리 권한 체계에 반영하여 관리하는 ECM(Enterprise Content Management) 기반의 전자 문서 유출 방지 시스템이다. 기존의 문서보안 관련 제품이 서비스 측면에서 사용자의 편의성 및 구축비용 측면에서 많은 문제점이 제기되고 있다. 본 연구에서는 문서보안과 관련하여 사용자에 의한 위치정보 등록 관리하는 형태로 관리하여 문서보안정책을 수립하여 문서보안에 적용하고자 한다. 기존의 문서관리 시스템은 위치정보를 각 문서마다 위치정보를 수동으로 입력하고, 부서별 위치정보를 입력하는 방식으로 활용되고 있어 수작업에 의한 처리 방식을 벗어나지 못하고 있는 실정이다. 또한 문서관리에서도 ID, Password로 사용자 확인하고, 확인된 사용자에게 설정된 권한 정보로 조회여부 결정하는 방식으로 적용되어 문서 유출 시 추적에 대한 문제점 및 사용자의 불편함이 존재하였다. 보안구역을 적용한 ECM시스템은 외부 활동 직원의 스마트 기기의 위치정보를 이용하여 자동으로 위치정보 입력/관리하도록 지원하며, 사용자 확인 및 사용자의 위치정보로 조회 가능 여부를 결정하여 보안을 고려하고, 사용자에게 설정된 권한 정보로 해당문서의 조회 여부를 결정하도록 ECM 기반의 문서관리 시스템을 구현하고, 보안 기능을 강화하여 차별화 된 서비스를 제공하고자 한다. 논문의 구성은 제 2장에서는 국내외 전자 문서관리 기술 동향 분석 차원에서 기술하며, 제 3장에서는 K-평균 알고리즘을 적용한 기존 문서관리 시스템의 개선방안을 제시하고, 제 4장에서는 ECM 기반의 문서관리 시스템의 프로토타입 구현 차원에서 기술하도록 한다.

2. 국내외 기술동향 분석

2.1 관련기술 현황

기업에서의 정보 유출 방지와 내부 통제를 위해

문서 보안은 중요한 이슈가 되고 있다.(한국저작권위원회, 2015.) 기업의 정보 유출 방지, 통제, 그리고 효율적인 데이터 관리를 위한 문서보안은 최근 3~4년 전부터 그 중요성이 더욱 부각되고 있다. 기업에서 데이터 자체에 대한 보안을 강화하는 추세가 이어지면서 건설업, 물류업, 병원, 및 은행, 금융 유통/서비스업까지 산업계 전반으로 ‘디지털저작권관리(DRM : Digital Rights Management)’ 솔루션 채택이 이어지고 있다. 기술 시장이 글로벌화 되고, 한국의 기술 유출이 문제가 되면서 국내 기업이 보안 전담 조직을 운영한지 대략 10여 년이 되어 가고 있으며(LG전자(2002년), 대한항공(2002년), SK하이닉스(2004년), 현대중공업(2004년) 등), 최근 2~3년 사이 예산과 인력이 크게 보강되고 있다. 기업의 정보와 문서관리 방법은 ECM 기반의 Mobile, Cloud로 진화되고 있으며, 단순 파일공유에서 시작한 문서관리 개념은 ECM을 넘어 비즈니스를 위한 문서 및 콘텐츠 중앙 집중 관리 단계로 진입하였다.



[그림 1] 기업정보/문서 관련 변화

2.2 ERM 시장규모

Enterprise Resource Management에 대한 SPRi 소프트웨어 정책 연구소 조사에 따르면 국내 ERP 시장은 지속적인 성장으로 2017년에는 314,101억원에 이를 것으로 전망하고 있으며, 해외시장은 2019년에 67,168백만\$로 조사되고 있다. 이는 전자 문서 관련 시장은 국내시장은 물론, 해외시장 역시 매년 10%내외의 성장을 전망하고 있다(SPRi 소프트웨어정책연구소, 2015).

〈표 1〉 ERM 시장 규모

구 분	2017년	2019년
세계 ERM 시장규모	58,149백만 \$	67,168백만 \$
국내 ERM 시장규모	280,696억 원	314,101억 원

2.3 ECM 기술동향

ECM(Enterprise Contents Management, 기업 콘텐츠 관리) 이란 업무의 효율화와 문서보안 등을 위해 파일 형태의 각종 문서와 이미지, 영상 등을 생성부터 폐기에 이르기까지 등록, 저장, 관리, 송수신, 조회 등 관리를 가능케 하는 솔루션이다. 콘텐츠는 프로젝트 관련 문서 및 제안서, 전략이 안, 일반 사무 양식 등을 포함하며 그림, 사진, 비디오 등 전자적으로 생성되는 다양한 종류의 파일과 데이터를 포괄적으로 의미한다(구홍서, 2007).

ECM은 기업에서 존재하는 모든 정형 데이터와 비정형 데이터 정보들, 각종 양식과 팩스 등 종이문서에서 웹페이지, 전자 우편 등 전자 문서에 이르는 모든 전자 콘텐츠를 생성, 가공, 활용 그리고 폐기까지 정보의 라이프 사이클에 걸쳐 통합 관리해주는 기술이다(이선우, 2013). 더 나아가 활용적인 측면에서 ECM은 콘텐츠와 프로세스의 통합 관리(Content Centric Process Management)와 지식관리시스템(KMS), 전사적 자원관리(ERP), 비즈니스 인텔리전스(BI) 등 주요 시스템들과의 연계를 통해 보다 신속 정확한 비즈니스 의사결정을 지원함으로써 업무 효율을 향상시키는 역할을 하고 있다(심재진, 2010).

3. 보안알고리즘 적용

3.1 군집 알고리즘 분석

입력받은 데이터를 k 개의 임의의 centroid에 각 데이터로부터 각 Cluster까지의 Euclidean distance를 계산하여 가장 가까운 centroid에 데이터를 할당

〈표 2〉 ECM 분류

ECM 분류	내 용
웹 콘텐츠 관리(WCM)	콘텐츠를 웹채널을 통해 배포
문서관리(DM)	콘텐츠 생성, 팀 간 협업과 워크플로우 관리에 활용
디지털 자산 관리(DAM)	이미지, 오디오, 비디오 등의 미디어 자산관리에 사용
이미징/캡처	높은 볼륨의 콘텐츠를 종이 형태로 생성 시 분류, 저장 관리 등을 전자포맷으로 적용
기록물 관리(RM)	콘텐츠를 기록물로서 보관이 필요할 때 기록물의 보관 주기 등 라이프 사이클에 따라 관리
아카이빙	콘텐츠를 ERP, 파일시스템 등 다른 애플리케이션에 캡처 하는데 사용
디지털 산출물 관리(DOM)	콘텐츠 저장소로부터 다양한 산출물을 관리(프린트 스트림, PDF 전환)
디지털 저작물 관리(DRM)	콘텐츠 저장소에 접속할 수 있는 권리를 통제

하고, 다시 centroid 위치를 수정 및 변경을 반복하여 optimal solution을 구한다(Hartigan and Wong, 1979.) 그러나 Hard Clustering의 한 종류인 K-means Clustering은 Euclidean distance의 한계로 Clustering 결과가 원 형태로 출력되고, Hard Cluster로 경계선에서 데이터에 대해 구분이 모호하다. 그렇기 때문에 아래와 같은 한계점이 발생한다.

- 1) Cluster 개수 k 를 입력 파라미터로 지정해야 한다.
- 2) local min으로 수렴할 가능성이 있다.
- 3) outlier에 민감하다.

3.2 알고리즘 적용

스마트 기기를 통한 군집 정보의 분석 알고리즘은 다양하게 존재한다. 계층적 군집분석방법을 사용하면 사전에 K 를 정하지 않고 군집의 개수를 파악할 수 있다(Xuan et al., 2001). 하지만 연산량이 $O(n^3)$ 로 알려져 있고, local optimum으로 수렴할 가능성이 있다(Pernkopf and Bouchaffr., 2005.) 그리고 현재 사용 중인 NAS 서버의 성능이 낮아 테스트에

어려움이 있으므로 향후 클라우드 기반의 ECM 구현을 위해서는 계층적 군집분석방법을 사용하도록 한다. 본 연구에서는 K-평균 알고리즘을 적용하여 안드로이드 스마트폰을 통한 위치정보의 수집이 되도록 프로그램을 개발하여 적용하며, 사용자 스마트 기기로부터 행동 특성 정보(GPS, 가속도 센서)의 수집 기능을 개발하여 머신러닝을 위한 데이터의 전처리 단계를 완성하여 사용자의 위치정보를 수집하여 안전 구역(Green Zone)을 생성하여 ECM에 연계하도록 한다.

3.2.1 위치정보의 수집절차

위치정보의 수집을 위하여 개인 정보 보호를 위하여 스마트 기기에서 안드로이드 앱 프로그램 시작 시 위치정보 제공 동의 확인 후 프로세스 시작하여 사용자 확인 후, GPS 정보를 5분 간격으로 수집하고, 2~3번 이상 같은 조건이면 정지 상태(A)로 결정하고, 가속도 센서로부터 속도 값을 계산, 속도가 4km/s 이하이면 정지 상태(B)로 결정하도록 데이터 값을 추출한다. 추출된 정보가 정지 상태(A) and 정지 상태(B)이면 정지 상태(C)로 판정하도록 한다. 두 번째 조건으로 현재 시간이 업무시간(08:00~20:00)이면 시간 값을 추출(T)하고 정지 상태(C)일 때, GPS X, Y, T, ID 값을 서버로 전송하도록 한다. 서버에 수집된 정보를 분석하여 사용자 동적 상태 정보 추정하여 정지 상태인지, 도보 중 인지, 자동차로 이동 중인지를 구분하도록 하며, 사용자 행동 특성 정보 필터링 기능 개발하여 속도가 정지 상태만 추출하도록 하며, 안전 구역(Green Zone) 학습용 데이터 송신 기능을 개발하여 EDMS의 계정과 연동이 가능하도록 한다.



[그림 2] 사용자 위치정보 수집 조건 예시

3.2.2 Green Zone정보 적용

안전 구역(Green Zone) 정보가 추가된 전자 문서관리 시스템의 개발은 기존 전자 문서 권한 대상(사람, 부서, 직위)과 범위(읽기, 쓰기, 목록 보기, 폴더 생성, 삭제)에 안전 구역(Green Zone) 정보를 추가하여 DB 테이블을 생성하게 되며, 권한 대상과 안전 구역(Green Zone) 정보 관계설정 기능을 개발하여 적용한다. 여기에서 안드로이드 프로그램의 개발은 위치 기반 전자 문서의 뷰어 시스템을 개발하여 적용하고, 안전 구역(Green Zone) 위치 확인 요청 기능을 개발하여 적용한다.

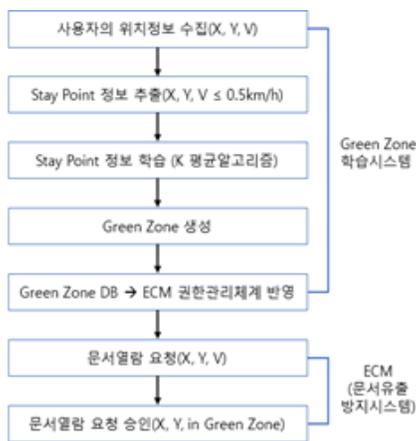
- 전자 문서 열람 요청 및 문서 Viewing 기능 개발
- 안전 구역(Green Zone) 밖일 경우 상위 권한자에게 열람 요청 기능 개발
- 안전 구역(Green Zone) 추정 정보와 전자 문서 권한 관리체계와 연동 프로토콜 개발(X, Y, t, user, DocNumber)

사용자의 행동 특성 정보의 수집은 안드로이드 기반의 프로그램으로 사용자의 스마트 기기에서 사용자의 행동 특성 정보(GPS, 가속도 센서) 수집의 기능을 개발하여 적용하게 되며, 응용프로그램의 조건은 아래와 같다.

- ① GPS는 5분 간격으로 추출, 2~3번 이상 같은 조건이면 정지 상태(A)로 결정
- ② 가속도 Sensor 속도 값 계산, 속도가 4km/s 이하이면 정지 상태(B)로 결정
- ③ 정지 상태(A) and 정지 상태(B)이면 정지 상태(C)로 판정
- ④ 현재 시간이 업무시간(08:00~20:00)이면 시간 값을 추출(T)
- ⑤ 정지 상태(C)일 때, GPS X, Y, T, ID 값을 서버로 전송

서버에서는 수집정보를 분석하여 사용자의 정지 상태 여부 및 동적 상태 정보를 판단하고, 안전 구역(Green Zone) 정보와 비교하여 보안문서 검색을 위한 사용자 권한부여에 대한 승인 여부를 판단한다. 여기에서 행동 특성 정보 필터링 기능을 통한 속도가 정지 상태만 추출이 가능하도록 하여 안전

구역(Green Zone) 학습용 데이터 송신 기능을 개발하여 적용하고, 문서관리 시스템과의 계정 연동이 가능 하도록 연계 모듈을 개발하여 적용한다. 안전 구역(Green Zone) 정보가 추가된 전자 문서관리 시스템의 개발은 문서 조회(열람) 요청 시 사용자의 위치정보 수집 및 전송 모듈이 요구되며, 행동 특성 정보(GPS, 가속도 센서) 수집을 통한 사용자의 정지 상태 여부를 판단하도록 한다.



[그림 3] 안전 구역(Green Zone)정보 적용절차

3.3 위치기반 학습시스템

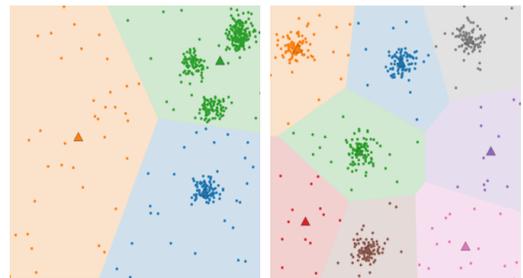
안전 구역(Green Zone) 위치 기반 학습 시스템 개발은 입력된 위치 값(X, Y, V)에서 Stay Point 추출 기능 개발하여 적용하고, 추출된 Stay Point 값을 통해 안전 구역(Green Zone) 생성 위치 기반 머신러닝 기능을 개발하여 K-평균 알고리즘에서 K 값을 자동 계산하는 모듈을 개발하여 위치 기반 학습 시스템을 완성하도록 한다.



[그림 4] 안전 구역(Green Zone) 생성 프로세스

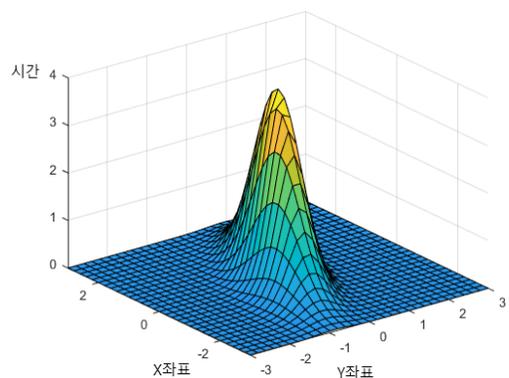
안전 구역(Green Zone) 학습 방법의 처리 절차는 다음과 같다.

- ① 사용자의 스마트폰에서 가속도 센서, GPS 값을 일정 간격으로 수집(가속도 X, Y, Z 값, GPS 위도, 경도 값)
- ② 가속도 연산 후 0~4km/h 속도인 위도, 경도 값 추출, 데이터 전송
- ③ ②에서 수신된 데이터로 K-평균 알고리즘으로 K개의 클러스터를 산출
- ④ 산출된 K개의 클러스터가 예비 안전 구역(Green Zone) 중심점임.
- ⑤ 예비 안전 구역(Green Zone) 중심점 반경 50m 영역으로 추정 후 거러치, 협력점 주소와 일치하면 안전 구역(Green Zone)으로 설정
- ⑥ 안전 구역(Green Zone)에 시간 정보를 Layer를 추가하여 안전 구역(Green Zone) 등급을 산출



K가 작을 때 K가 클 때

[그림 5] K-평균 알고리즘의 예시



[그림 6] 정규분포 분석 및 안전 구역(Green Zone)추정

모바일 시스템 모듈 개발, 두 번째 모바일 시스템으로부터 전송된 사용자의 행동 특성 정보를 파싱하고 (x, y) 위치, (t) 시간 값으로 정규분포 분석을 하여 중심 값을 구하여 Green Zone을 추론하고 생성하는 엔진 개발, 세 번째 Green Zone 정보 관리. 즉, 기존 전자 문서 접근 권한에 Green Zone 정보가 추가된 전자 문서 관리 시스템(ECM)을 개발 영역으로 구분하여 개발하도록 한다.

4.3 알고리즘의 적용

현재 국내의 ECM에 대한 인식은 부족한 상태이며, 전자 문서관리 시스템, 이미징 시스템, 기업의 문서 처리, 프로세스 워크플로우 기술 등을 보유 기업으로 위치기반 서비스, 머신러닝에 대한 전문 기술/인력이 부재인 상태이다. 부족한 전문 기술인 위치기반 서비스, 머신러닝에 대한 공동 기술 개발 및 기술 가이드를 통해 전자 문서 관련 기술에 위치 기반, 머신러닝 등 융합 기술을 축적하여 4차 혁명을 준비하고자 하며, 알고리즘 구성역역과 문서 열람 권한부여 영역을 구분하여 솔루션을 적용하도록 한다.



[그림 10] 문서 인증 절차

알고리즘 구성 영역은 정지 상태의 위치정보 (Stay Point) 학습을 통해 K-평균 알고리즘을 적용한 보안 안전 구역(Green Zone)을 생성하는 모듈을 구성하여 적용하게 된다. 문서 열람 권한부여 영역에서는 생성된 보안 안전 구역(Green Zone) 정보를 전자 문서 유출 방지 시스템의 문서 권한 관리에 적용하는 NAS 기반의 서버에 사용자의 이동정보를 수신하여 정지 상태의 위치정보(Stay Point) 학습, 사용자의 문서 열람 요청 처리 및 열람 기능을 제공하는 사용자 시스템을 개발하여 적용한다.

4.4 프로그램 구현 및 테스트

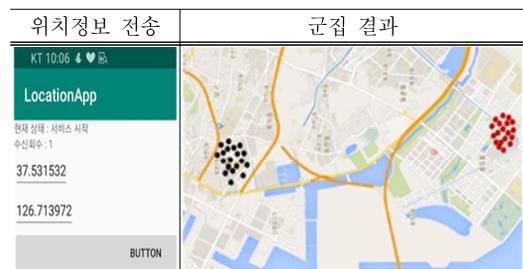
알고리즘 적용을 위하여 스마트 기기에서 위치정보 추출을 위한 Android app을 구현하여 사용자의 문서 열람 요청 시, 위치정보를 NAS server에 위치정보를 전송하고, NAS server에서는 군집 알고리즘을 구동하여 Green Zone 값을 생성하도록 구현하여 시험을 실시하였으며, 지역은 인천 갯벌로 및 낙섬동로에서 위치정보 값을 생성하도록 테스트 하였을 때, 성공적인 Green Zone 식별이 가능하다는 것을 확인하였다. 본 테스트에서는 가속도 센서를 이용한 속도 측정 대신 GPS를 이용한 속도 측정을 사용하였다.

<표 3> 위도, 경도 데이터셋(data-set)

구분	위도	경도
인천 갯벌로	37.381655	126.653950
	37.382260	126.652577
	37.382703	126.653135

인천 낙섬동로	37.455446	126.644039
	37.454394	126.644887
	37.454505	126.642376

<표 4> 프로그램 시험 결과



여기에서 수집된 정보는 K-means 알고리즘을 통하여 클러스터링 결과가 도출되게 되며, 이러한 결과는 ECM 서버에 사용자 권한관리 테이블과 비교하여 안전구역 여부를 판단하고 사용자의 아이디를 통한 문서의 열람이 가능하도록 지원된다.

4.5 성능평가 방안

본 연구는 “위치 정보를 포함하는 보안 문서 관리 시스템 및 그의 제어 방법”, “위치 기반 문서 관리 클라우드 시스템”은 보안문서를 생성할 때 사용자의 위치정보를 포함하는 방법에 관한 것으로 개발 관련 제품의 신뢰성은 초기 설치 상태에서 얼마나 빠른 시간 안에 안전 구역(Green Zone)을 생성할 수 있느냐에 달려있다. 이에 대한 신뢰성은 공인 기관에서 신뢰를 얻기에는 시험기간, 방법 등 구체적인 방안이 마련되지 않아 자체 기준으로 시험으로 수행 예정이며, 성능 평가 기준은 다음과 같다.

1. K-평균 알고리즘에서 K 값을 결정 성공률
2. 클러스터링 결과값(위치)과 실제 위치 간 거리 오차
3. Green Zone 확인 요청 결과 응답시간
4. Green Zone 확인 성공률
5. 기능 구현 정확성

보안문서 생성 시 위치정보를 포함한다는 것과 문서 열람을 요청할 때 열람 가능한 위치인지 여부를 확인하는 것과는 기술적으로 다른 분야이고 위치정보 학습을 통해 Green Zone을 생성하는 구성 요소가 없어 특히 가능성이 높은 것으로 판단하고 있으나 통상의 기술자가 쉽게 응용할 수 있다는 심사관의 거절논리에 대응하기 위한 방안은 과제를 수행 시 검토하여 회피할 예정이다.

5. 결 론

ECM 기반의 문서관리 서비스는 자체적인 시스템 구축을 통한 방안과 클라우드 서비스와 게이트웨이 구성을 통한 서비스 방안이 다루어질 수 있으나, 본 연구에서는 간단한 프로토타입 모델을 통하여 서비스를 형상화하고, 기존 상용화된 ECM 기반의 서비스와 연계방안을 고려하도록 한다. 여기에서 구현되는 프로토타입 모델은 스마트 기기를 통하여 안전 구역(Green Zone)의 추출이 가능

하도록 하였으며, NAS 서버에서는 문서의 권한 관리 및 문서 열람 관련 정보를 관리하도록 구성하여 보안이 강화된 ECM 서비스를 완성하도록 한다. 향후 추가 연구영역은 Green Zone 정보 관리를 추가하여 Green Zone을 추가, 수정, 삭제할 수 있는 기능을 개발하여 ECM 과 연동이 가능하도록 구성하고, 사용자가 외부에서 스마트 기기로부터 문서 열람을 요청할 때 기존의 부서, 보안 등급 등 문서 접근 권한 인증 프로세스에 Green Zone에 위치하고 있는지 여부를 추가로 인증하는 프로세스를 추가하여 적용하며, 사용자 인증이 적합할 경우 요청한 문서를 검색하여 MS-Office, 한글 등을 PDF로 변환하고 스마트 기기로부터 전송하여 열람할 수 있도록 추가 기능을 개발하여 사용자의 편의성을 높이도록 한다.

참고문헌

- 구홍서, “전사 콘텐츠 관리를 위한 접근방법 연구”, 한국산업정보학회 논문지, 2007.
- 심재진, “문서혁신을 위한 기업콘텐츠관리(ECM)시스템 도입 전략에 관한 연구”, 단국대학교 정보미디어대학원 IT학과 석사학위 논문, 2010.
- 이선우, “ECM 시스템의 성능 개선 및 평가 방안에 관한 연구”, 동의대학교 대학원, 박사학위 논문, 2013.
- 한국저작권위원회, “국내 문서 보안 기술 현황 : DRM과 RFID 보안용지”, [뉴스레터] 5월 4주, 2015.
- Hartigan, J.A. and M.A. Wong, “Algorithm AS 136 : A K-Means Clustering Algorithm”, *Journal of the Royal Statistical Society, Series C (Applied Statistics)*, Vol.28, No.1, 1979, 100-108.
- Pernkopf, F. and D. Bouchaffra, “Genetic-based EM algorithm for learning Gaussian mixture models”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol.27,

No.8, 2005, 1344-1348.
SPRi 소프트웨어정책연구소, “SW산업 연간보고서”,
2015.
Xuan, G., W. Zhang, and P. Chai, “EM algorithms

of Gaussian mixture model and hidden Mar-
kov model”, *Proceedings 2001 International
Conference on Image Processing*(Cat. No.
01CH37205), 2001.

◆ About the Authors ◆



유 갑 상 (gsyoo21@empas.com)

연세대학교 (공학석사), Ural State Univ.(이학박사), 동국대 정보통신학과 객원교수, 현재는 청운대학교 컴퓨터공학과 부교수로 재직 중이다. 주요 연구 분야는 SW공학, DBMS, AI 등이다.



조 승 연 (csy1018@naver.com)

청운대학교 컴퓨터공학과 (공학석사), 현재는 (주)드림메디컬그룹 IT팀 팀장으로 재직 중이다. 주요 관심 분야는 SW공학, DBMS, 빅데이터 등이다.



황 인 태 (hwang9029@naver.com)

청운대학교 컴퓨터공학과 (공학사), 주요 관심 분야는 알고리즘, 서버개발 등이다.