공학석사 학위논문

e-Discovery에서 ESI의 수집과 보존에 관한 도구 설계 및 구현

2011년 8월

부 경 대 학 교 대 학 원

정보보호학 협동과정

김 주 영

공학석사 학위논문

e-Discovery에서 ESI의 수집과 보존에 관한 도구 설계 및 구현

지도교수 신 상 욱

이 논문을 공학석사 학위논문으로 제출함.

2011년 8월

부 경 대 학 교 대 학 원

정보보호학 협동과정

김 주 영

김주영의 공학석사 학위논문을 인준함.

2011년 8월 26일

주심 이학박사 이 경 현 (인)

위원 공학박사 송 하 주 (인)

위원 이학박사 신 상 욱 (인)

차 례

그림 차례

표 차례

Design and Implementation of ESI Collection and Preservation tools in e-Discovery

Ju Young Kim

Interdisciplinary Program of Information Security, The Graduate School,
Pukyong National University

Abstract

On December 2006, the Federal Rules of Civil Procedure (FRCP) amendments of the United States make e-Discovery compulsory. Therefore, companies must build the system to manage ESI(electronically stored information) for preparing e-Discovery.

This thesis provides not only explanation on e-Discovery but also analyze information of relevant e-Discovery projects and the e-Discovery solutions. Based on these analysis results, this thesis elicits the requirements of the collection and preservation step in e-Discovery. Also, this thesis explains the acquisition method of ESI in SNS and Smartphone.

And then this thesis designs the support tool for e-Discovery which is composed of 3 modules including Agent, Server and Manager. Especially, this tool provides the function to acquire SNS data. Finally, this thesis presents the implementation of the prototype of e-Discovery support tools.

I. 서 론

1. 연구배경

2006 년 12월 1일에 개정된 미국 연방 민사소송법은 전자적 정보의 증가와 증거로써의 중요함을 반영하여 기존 Discovery 제도에서 증거의 범위를 전자적 증거까지 포함하는 e-Discovery를 제정하였다[1]. Discovery 제도는 민사소송에 있어서 소송당사자 간의 증거를 제시하는 제도로 e-Discovery가 제정됨에 따라 전자적 증거도 법정에 제시해야 한다. 따라서 기업들은 e-Discovery에 대비하기 위해 기업 내 전자적 정보들을 관리하기 위한 시스템을 구축해야 한다. e-Discovery는 증거가 될 수 있는 정보를 수집하고 보존하는 IT 전문가들과 수집된 정보를 검토하고 분석하는 법률 전문가들이 상호 의견을 교환하며 진행해야 함으로 이를 지원해줄 e-Discovery 지원 솔루션을 도입하는 것이 바람직하다. 또한, e-Discovery는 조직 내 모든 전자적 정보를 검토해야 함으로 비용과 시간이 많이 소요되는 작업이므로 지원 도구는 반드시 필요하다.

최근 e-Discovery의 주요 판례들을 살펴보면 e-Discovery의 초기 단계인 전자적 증거 수집 및 보존 단계에서 증거물을 산출하지 못해서 패소하는 사례가 발생하고 있다. 또한, 증가하는 SNS(Social Network Service)와 스마트폰과 함께 증거로써의 SNS 데이터 및 스마트폰 데이터의 가치도 높아지고 있다. 실제로 많은 기업에서 Twitter[2], Facebook[3]를 홍보 및 마케팅 용도로 운용하고 있다. SNS가 증거가 된 대표적인 사례로는 국내에서 Twitter를 이용한 선

거법 위반 사례가 있다 [4]. 따라서 e-Discovery 지원 솔루션은 기존의 전자 문서나 전자 메일에 국한되지 않고 SNS와 같은 다양한 형태의 증거를 수집하고 보존할 수 있어야 한다.

본 논문은 e-Discovery의 절차와 관련 프로젝트에서 전자적 증거의 수집 및 분석에 관한 내용을 분석한다. 또한, 현재 국산 e-Discovery 지원 솔루션이 전무 한 가운데 e-Discovery 지원 솔루션의 국산화를 위해 e-Discovery 지원 도구를 설계하고 기존 솔루션이 지원하지 않는 SNS 데이터 및 스마트폰 데이터를 분석하여 수집하고 보존하는 방법을 설명하며 e-Discovery 지원 도구의 프로토타입을 구현한다.

2. 연구 내용 및 구성

본 논문에서는 e-Discovery 제도에 대해서 설명하고 e-Discovery 관련 프로젝트인 Electronic Discovery Reference Model(이하 EDRM)[5]과 Sedona Conference[6]의 문서들을 분석하고 시중에 출시되어 있는 e-Discovery 지원 솔루션을 분석한다. 이를 토대로 e-Discovery에서 Electronically Stored Information (이하 ESI)의 수집과 보존의 요구 사항을 도출해 본다.

현재 가장 널리 보급되어 있는 스마트폰 중 아이폰을 타켓으로 하여 아이폰의 데이터를 수집하는 방법에 대해서 설명하고 SNS의 특성을 이용하여 SNS 데이터를 수집하는 방법에 대해서 설명한다. 또한, 앞서 도출한 요구 사항과 수집 방법을 토대로 e-Discovery 지원 도구를 설계하고 프로토타입을 구현한다.

본 논문의 구성은 2장에서는 e-Discovery 제도에 대해서 설명하

고 e-Discovery 관련 프로젝트들을 소개한다. 또한, 기존의 e-Discovery 솔루션에 대해서 조사하고 대표적인 e-Discovery 판례에 대해서 설명한다. 3장에서는 EDRM과 Sedona Conference의 ESI 수집과 보존에 관한 가이드라인 문서를 분석하고 분석한 내용을 토대로 수집과 보존에 관한 요구 사항을 도출하고 4장에서는 e-Discovery 지원 도구를 설계하고 프로토타입을 구현하고 5장에서 결론을 맺는다.

II. 관련 연구

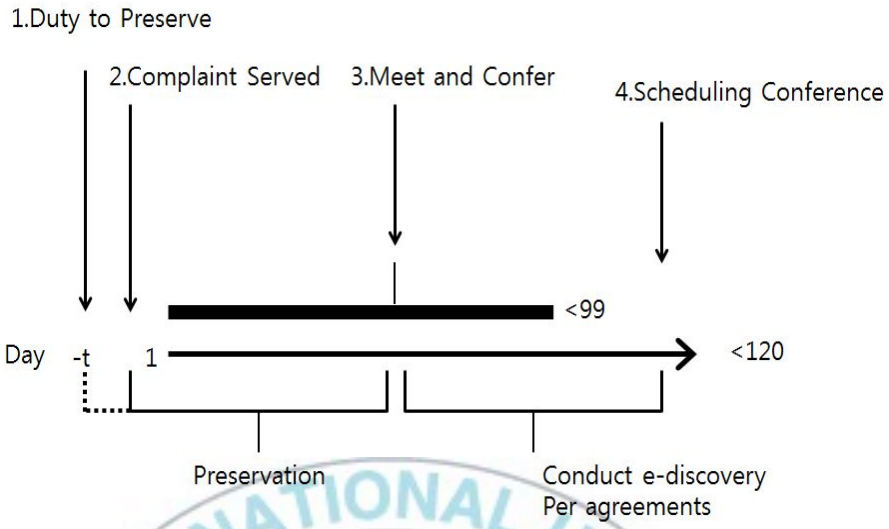
e-Discovery와 관련하여 현재 여러 프로젝트가 진행 중이며 EDRM과 Sedona Conference를 통해 기술적 절차와 가이드라인을 제정하는 작업이 진행 중이다.

본 장에서는 e-Discovery 제도에 대해서 알아보고 EDRM과 Sedona Conference에 대해서 살펴본다. 또한, 기존의 ESI 보존과 수집에 관련한 판례를 분석하고 현재 출시된 솔루션들에 대해서 알아본다.

1. e-Discovery 제도

가. e-Discovery

Discovery 제도란 소송당사자가 공판 전에 공판의 준비를 위해 법정 외에서 법정의 방법에 의하여 소송의 issue(쟁점 혹은 쟁점사실)를 명확히 하는 정보 및 증거를 공개, 수집하는 제도를 말한다 [7]. 즉 소송당사자 간에 증거를 교환하고 요구하는 제도로 재판 당사자들은 공판 이전에 증거자료들을 수집하고 보존해야 한다. 대다수의 소송은 Discovery 절차를 거치면서 소송의 쟁점이 명확해지고 증거의 윤곽이 밝혀짐에 따라 효율적인 재판이 가능해지며 상대방의 증거 유무를 판단하여 재판에서 승패소를 예측할 수 있게 된다. e-Discovery 제도는 기존 Discovery 제도에서 증거의 범위를 ESI로 확대하여 개정한 것으로 ESI를 소송개시일로부터 120일 이내에 제출할 것을 요구한다. 아래 [그림1]은 소송 진행 절차를 나타낸 것이다 [8].



[그림 1] 소송 진행 절차

나. Electronically Stored Information

ESI는 디지털 형태로 작성된 모든 정보를 의미한다. 따라서 전자문서, 전자메일, 비디오, 오디오 파일 등 다양한 형태를 보인다. 근래에는 소셜 네트워킹 서비스, 클라우딩 시스템, 스마트 디바이스 등에 포함된 ESI도 증거로써 가치가 높아짐에 따라 수집 대상이 되고 있다.

ESI는 Native File, Metadata File, Image File 파일로 분류할 수 있으며 각 파일의 특징을 요약하면 아래 <표 1>과 같다 [9].

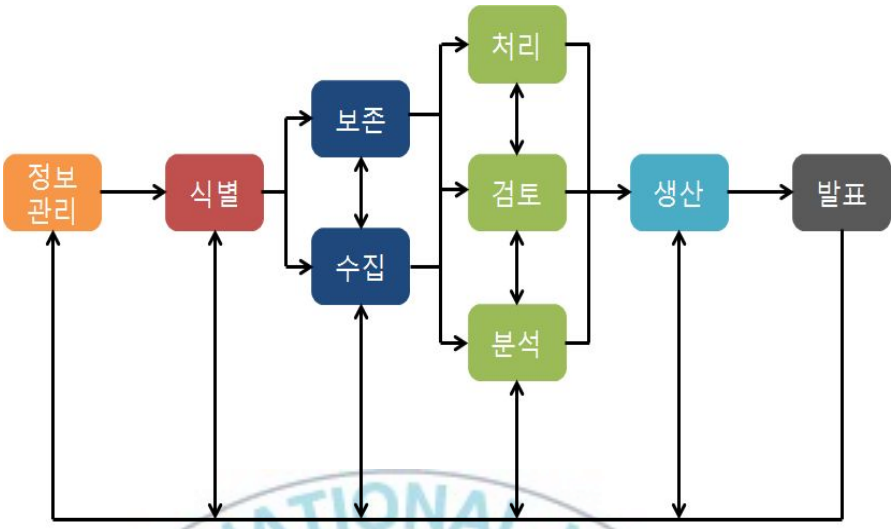
<표 1> ESI의 분류

형태	설명
Native File	응용프로그램에 의해서 생성된 파일로 스프레드시트나 워드프로세서 등을 이용해서 생성된 파일들이 여기에 속한다.
Metadata File	파일의 생성날짜, 수정날짜, 파일 크기 등 파일의 속성을 나타내는 파일로 ESI 특징을 파악하기 위한 정보가 담긴 파일
Image File	표준화된 문서파일 형태로 TIFF, PDF 등의 포맷을 가지고 있다. 출력은 가능하지만 일부 포맷들은 수정이 안 되는 특징을 가지고 있다.

2. e-Discovery 관련 프로젝트

가. Electronic Discovery Reference Model

EDRM은 미국연방민사소송법에 따른 e-Discovery의 절차와 각 절차에 대한 프레임워크 가이드를 작성하는 프로젝트이다. EDRM에서는 프레임워크뿐만 아니라 표준 Dataset 과 Information Management Reference Model 등 e-Discovery에 관련된 다양한 프로젝트를 진행 중이다. [그림 2]는 e-Discovery 단계를 EDRM에서 도식화한 것이다[5].



[그림 2] Electronic Discovery Reference Model

EDRM은 e-Discovery 제품과 솔루션을 개발하고 평가하기 위해서 각 단계별로 프레임워크 가이드를 제공한다. 단계별 프레임워크 가이드를 요약하면 <표 2>와 같다.

<표 2> EDRM 프레임워크 가이드의 단계별 수행 내용

단계	설명
Information Management (정보 관리)	소송이 시작되기 전에 조직에서 정보 관리 정책에 의해 ESI가 관리 되는 단계
Identification (식별)	소송이 예상되는 시점에서 수집해야 할 ESI의 범위를 결정하고 ESI의 데이터 맵을 작성하는 단계
Preservation (보존)	식별된 ESI를 변경 및 훼손되지 않게 보관하는 단계
Collection	네트워크 스토리지, 이동식 저장장치, SNS, 스마트

(수집)	디바이스 등 다양한 소스로부터 ESI를 수집하는 단계
Processing (처리)	중복되는 ESI를 제거하고 검토 분석을 위한 포맷으로 변환하는 단계
Analysis (분석)	수집된 ESI를 소송과 관련하여 평가하고 ESI 관련된 사건주제, 인물, 문서의 중요성 등을 작성하는 단계
Review (검토)	기밀성 데이터를 식별하고 e-Discovery 전략을 수립하기 위해 ESI를 검토하는 단계
Production (생산)	상호 식별 가능한 형태로 ESI를 산출하는 단계
Presentation (발표)	법정에서 제출된 ESI 들을 효과적으로 보일 수 있도록 발표하는 단계

나. Sedona Conference

Sedona conference는 1997년 Richard G. Braman에 의해서 설립된 법률 컨퍼런스로 반독점법, 특허와 저작권, 종합 소송의 이슈를 다룬다. 이 컨퍼런스는 변호사, 법학자, 자문위원 등으로 참여하여 7개로 나누어진 워킹그룹으로 구성되어 있으며 각 워킹그룹들은 ESI에 관한 가이드라인과 원칙을 제정하고 있다. 각 워킹그룹이 다루는 내용은 아래와 같다.

(1) WG1: Electronic Document Retention and Production

전자 문서의 생산과 및 보존을 다루는 워킹그룹으로 e-discovery에서 전자정보 관리, legal hold 등에 관한 가이드라인과 원칙을 개발한다.

(2) WG2: Protective Orders, Confidentiality & Public Access

기밀 유지 문제와 기밀 명령 보호, 공용 액세스 허용 명령 등에 관한 원칙과 모범사례를 개발한다.

(3) WG3: The Role of Economics in Antitrust

반독점법에 관한 권장사항, 모범사례를 개발한다.

(4) WG4: Working Group on the Intersection of the Patent and Antitrust Laws

반독점법에 관한 분쟁을 해결할 수 있는 가이드라인을 제정한다.

(5) WG5: The Markman Process and Claim Construction

특허 소송에서 발생하는 분쟁과 특허청구범위의 법적 해석에 대한 모범 사례를 개발한다.

(6) WG6: International Electronic Information Management, Discovery and Disclosure

ESI 관리와 e- disclosure에서 발생할 수 있는 문제를 해결하기 위해 모범 사례를 개발하고 disclosure에 관한 국내 및 국제 법률과 정책에 대해 연구한다.

(7) WG7: Sedona Canada

캐나다 실무자들을 위한 워킹 그룹이다.

3. e-Discovery 관련 주요 판례

가. Zubulake 와 UBS Warburg, (2003-2005)[10]

Zubulake가 UBS Warburg에 제기한 고용차별 소송은 대표적인 e-Discovery 판례이다. 이 소송에서 UBS Warburg은 ESI를 매니지먼트 하는 시스템을 갖추었음에도 불구하고 ESI의 보존 즉 Litigation hold 있어서 지속적인 모니터링과 개별적으로 Litigation hold 통지를 하지 않아 UBS Warburg은 Zubulake가 제시한 메일을 복구하기 위해 복구비용으로 $175,000을 소요하였다.

나. Qualcomm 과 Broadcom 의 특허 분쟁 (2007)[11]

Qualcomm과 Broadcom사의 특허 분쟁 소송은 ESI 보존의 의무를 다하지 못한 대표적인 사례라고 볼 수 있다. 당시 Qualcomm은 변호사들에 의해 ESI를 왜곡하고 Brodacom에서 제시한 200,000개의 문서와 메일을 산출하지 못해서 $8.6M를 배상하고 ESI를 왜곡한 Qualcomm 측 변호사들에게 징계를 내렸다.

다. Phillip M. Adams와 ASUS 특허 분쟁 (2009)[12]

Phillip M. Adams와 ASUS의 특허 분쟁 소송은 ASUS가 Phillip M. Adams의 특허를 침해했던 사건이다. 이 사건은 소송 중에 Phillip M. Adams는 ASUS가 고의로 증거가 되는 메일을 훼손했다고 주장했지만 ASUS측은 메일을 검색하는 시스템조차 갖추어 놓고 있지 않았다. 법원은 메일을 저장하는 중앙화된 스토리지나 검

색 매뉴얼조차 갖추어 놓지 않은 ASUS를 ESI에 보존의 의무를 다하지 못했다고 판단하여 Phillip M. Adams가 승소하게 되었다.

라. 판례 분석

위에 3가지 판례를 볼 때 e-Discovery에서 ESI의 수집과 보존은 재판의 승소를 가르는 중요한 단계임을 알 수 있다. 또한, ESI를 왜곡하는 것은 패소하는 것뿐만 아니라 변호사들에게 징계가 내려질 수 있는 행위라는 것을 알 수 있다. 따라서 ESI 보존과 변경 단계를 가이드라인을 통해 철저하게 이행하는 것이 필요하다.

4. 기존 솔루션 분석

현재 출시된 e-Discovery 솔루션들은 EDRM를 참고하여 작성된다. 현재 시중에 나와 있는 대다수의 솔루션들은 ESI의 검토와 분석을 위한 도구로써 활용되고 있으며 일부 도구들은 ERP 솔루션과 함께 제공되는 솔루션들도 있다. 현재 출시된 솔루션들의 기능적 범주는 아래 표와 같다.

<표 3> e-Discovery 솔루션 기능

항목	설명
Information Management	조직 내 정보 관리 기능으로 ERP 솔루션들이 제공함
Collection and Preservation	ESI 보존 및 수집 기능으로 일부 포렌식 도구들이 수집도구로 사용
Search	수집된 ESI로부터 필요한 ESI를 생산하는 기능

Review, Processing, Analysis, Production	ESI의 검토, 분석, 생산 기능으로 대다수의 e-Discovery 솔루션은 이 기능을 중점으로 개발되고 있음

대다수의 e-Discovery 솔루션들이 ESI 분석, 검토에 초점이 맞춰져 있으며 ESI의 수집을 위해 포렌식 도구를 병행하면서 사용하고 있다. 현재 출시된 도구들의 종류와 기능을 요약하면 아래 <표 4>와 같다.

<표 4> 현재 출시된 e-Discovery 솔루션

솔루션	설명
AccessData eDiscovery[13]	포렌식 수집 방법을 이용한 증거 수집과 다양한 검색 방법들을 지원. 실버라이트 기반의 분석도구로 편리한 UI를 제공
EnCase eDiscovery[14]	포렉식 도구를 기반으로 하는 강력한 수집 기능과 사전 데이터 수집 보존 기능들을 제공.
Nuix[15]	사전 인덱스 기능으로 검색속도가 빠름. 수집된 자료의 연관성을 그래프로 만들어 주는 기능이 특징
SourceOne e-Discovery[16]	EDRM에서 제정한 표준 XML를 지원하는 생산 기능과 자료 식별을 위한 커스터마이징 태그 기능 등을 제공함
IBM Infosphere[17]	ERP를 기반으로 하는 솔루션으로 정보관리 측면에서 강력한 성능을 가짐
Clearwell[18]	가장 많이 사용되는 e-Discovery 솔루션으로 e-Discovery에 전반적으로 필요한 기능 등을 모두 다 갖추고 있음

III. ESI의 보존과 수집에 관한 요구 사항

본 장에서는 e-Discovery의 단계중 ESI의 보존과 수집에 대해서 설명하고 EDRM과 Sedona Conference의 ESI 수집과 보존에 관한 문서들을 분석하고 요구사항을 도출해 본다. 또한, 추가적으로 증거로써의 가치가 있는 소셜 네트워크 서비스와 클라우딩 시스템 그리고 스마트 디바이스에 존재하는 ESI를 수집하는 방법에 대해서도 설명한다.

1. 보존 단계와 수집 단계 분석

EDRM에서는 현재 e-Discovery 제도 절차에 관한 프레임워크 가이드를 제공하고 있으며 현재에도 꾸준히 업데이트되고 있다. 또한, Sedona Conference에서는 ESI의 보존에 관한 가이드라인과 실례를 재정하는 작업이 진행 중에 있다. 이 절에서는 EDRM의 프레임워크 가이드 중 보존과 수집의 가이드라인을 분석하고 Sedona Conference의 여러 가이드 중에 ESI의 식별과 보존에 관한 가이드라인 문서와 보존의 의무에 관한 가이드라인 문서를 분석한다.

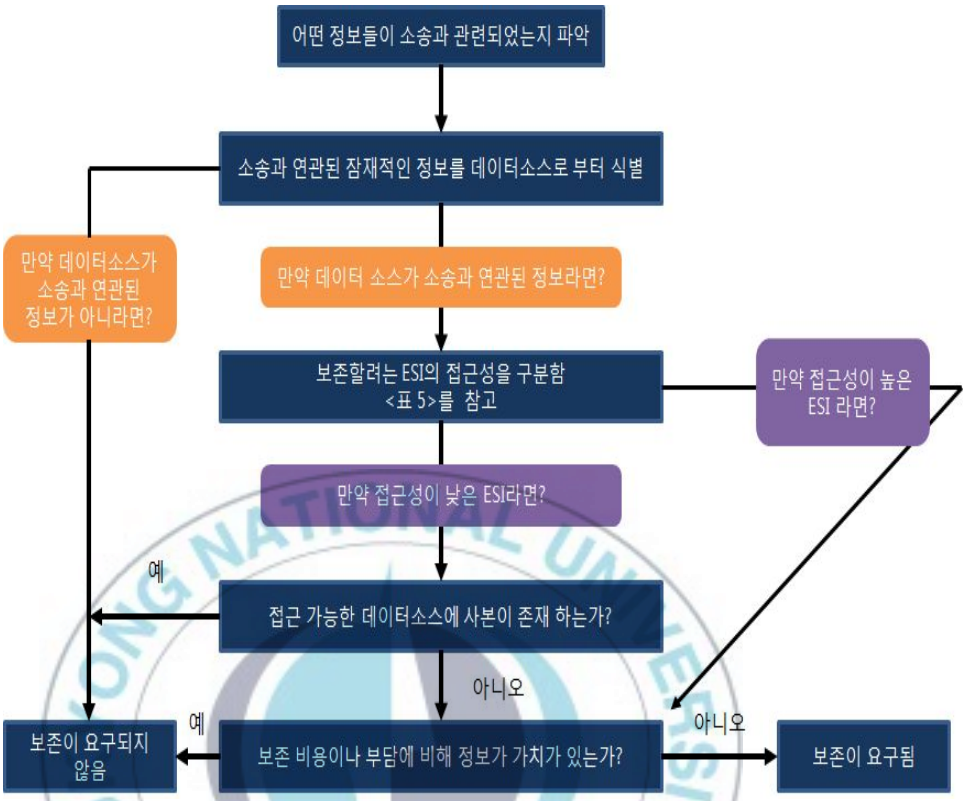
가. 보존과 수집 단계

보존 단계는 이전 단계인 ESI 식별 단계를 거쳐 소송에 연관된 ESI들을 확보하고 e-Discovery 절차가 끝날 때까지 ESI의 파괴와 훼손을 막는 단계이다. 이 단계에서는 Litigation Hold가 적용되는데 Litigation Hold는 소송이 시작되는 시점에서 조직 구성원들이 소송에 연관된 ESI를 보존해야 하는 의무를 말한다. ESI의 보존 단계에

서는 정보 관리 정책에 의한 정보 삭제부터 고의적인 삭제 시도 까지 조직 내 ESI를 변경, 파괴하는 모든 위협을 차단해야 한다. 따라서 IT 전문가들은 ESI를 보존하기 위한 정책과 도구들을 결정하고 수립된 정책에 의하여 보존 업무를 수행해야 한다. [그림 3]은 EDRM에서 보존 단계를 도식화한 것이다[19].

[그림 3] 보존 단계의 세부 절차

Sedona Conference의 THE SEDONA CONFERENCE COMMENTARY ON: Preservation, Management and Identification of Sources of Information that are Not Reasonably Accessible (2008)[20]에서는 ESI를 보존하기 위한 의사 결정 트리를 제공한다. 이 의사 결정 트리는 어떤 ESI를 보존해야 할지를 결정하기 위해 ESI의 접근성을 이용한다. 상대적으로 접근이 어려운 데이터의 경우 수집 및 보존비용을 고려하여 보존 여부가 결정된다. [그림 4]는 의사결정트리를 도식화한 것이다[20].



[그림 4] ESI 보존 의사 결정 트리

ESI의 보존 결정에 있어서 ESI의 접근성은 보존 결정에 큰 영향을 미친다. 예를 들면 OCR이 적용되지 않은 이미지에서 이미지에 담겨 있는 콘텐츠를 검색하는 작업은 사람이 모든 ESI 직접 검토해야 함으로 많은 시간과 비용을 요구한다. <표 5>는 의사결정트리에서 ESI 보존을 결정하기 위해 접근성을 분류하고 그 예를 나타낸 것이다[20].

<표 5> ESI 접근성

	Factors	Examples
1	Active on-line data	하드 디스크, PDA, network storage
2	Near-line data	Robotic storage devices와 같은 광학 디스크
3	Offline storage/archives	이동식 광학 디스크나 자기테입
4	Backup tapes	순차적 접근 가능한 구조화 되지 않은 개별 문서 혹은 파일
5	Physically damaged media	손상된 CD , DVD, 하드 디스크 로 일반 드라이브로 읽을수 없는 미디어
6	Legacy media	현재 디바이스들과 호환되지 않아 접근이 불가능한 드라이브나 장치들
7	Transient complexity	웹페이지가 계속 삭제되고 추가하기 위해서 스토리지에 오버라이팅 하는 것
8	Hidden complexity	삭제된 파일을 확인 하기 위해서는 전문적지식이나 도구가 필요
9	Extraction complexity	슬랙 스페이스에서 발견한 데이터 플래그먼트를 복사하는 것
10	Preservation complexity	PC에서 생성된 캐쉬 파일과 임시 파일을 운영체제를 disabling 하지 않고 보존 하는 것
11	Search complexity	OCR이 적용되지 않은 이미지 파일
12	Dispersion complexity	여러대의 PDA 디바이스에 있는 보존되어 있는 데이터를 검토해야 함.

데이터의 축적을 최소화 하기 위해서는 정보관리 정책을 세워서 ESI를 관리할 것이 권장된다. 또한, 수집된 정보를 이동해야 할 경우 모든 접근 가능한 데이터의 보존 의무를 지켜야 하며 소송기간 중에는 보존의 의무를 충족시키기 위해서 조직 구성원 간에 적절한 협력과 효과적인 리소스 활용이 장려된다.

ESI의 보존은 소송의 승패를 좌우할 만큼 중요한 요소이다. Sedona Conference의 COMMENTARY ON LEGAL HOLDS: THE TRIGGER & THE PROCESS[21] 에서는 ESI 보존에 관한 가이드라인을 실례와 함께 제정하고 있다. 가이드라인에서는 ESI 보존에 관한 정책을 수립하고 적용하는 것은 보존의 의무를 이행하는 것을 증명할 수 있는 중요한 요소라고 언급하고 있으며 ESI 보존 및 식별 범위를 지정하기 위해서는 소송과 ESI의 연관성, 증거의 가치, 보존의 비용 등을 고려해야 한다고 설명하고 있다. 또한, 기업은 Litigation Hold 관한 문서화된 정책과 특수한 상황의 hold 구현 정책을 고려해야 하며 Litigation Hold는 정기적으로 모니터 되어야 한다고 설명하고 있다.

수집 단계는 조직 내에 존재하는 소송에 연관된 잠재적 ESI를 수집하는 단계이다. 조직 내 시스템에 존재하는 잠재적 ESI를 검토 및 분석을 위해 수집한다. 수집된 데이터들은 다시 보존 단계를 거친다. 이 단계에서는 증거의 무결성 보장과 삭제되거나 손상된 ESI를 복원하기 위해 포렌식적 기법을 활용할 수 있다. 또한, ESI는 그 형태와 존재하는 디바이스가 매우 다양함으로 ESI의 형태에 맞게 수집할 수 있어야 한다.

2. ESI 수집과 보존에 대한 요구 사항 도출

EDRM과 Sedona Conference 자료를 토대로 ESI의 수집과 보존에 관한 요구사항을 도출한다. 앞에서 언급한 대로 ESI의 수집과 보존은 재판의 승패소 여부를 결정하는 중요한 요소이므로 요구사항을 토대로 e-Discovery의 보존 및 수집 단계를 준비하는 것이 반드시 필요하다. 수집과 보존에 대한 요구 사항을 도출해보면 아래와 같다.

가. 수집 요구 사항

앞선 문서들을 토대로 ESI를 수집에 관한 요구 사항을 도출해보면 보면 다음과 같은 요구 사항을 도출해 볼 수 있다.

- 조직 내의 정보관리 정책 수립 및 적용
- ESI가 포함되는 모든 디바이스에 대해서 ESI 수집
- ESI의 특성과 형태를 고려해서 ESI 수집
- ESI산출 비용 및 접근성을 고려
- 수집된 ESI의 무결성 보장
- 포렌식적 ESI 수집 방법 적용

조직 내 ESI가 수립된 정보관리 정책에 의해 관리되고 있어야 하며 ESI가 포함될 수 있는 모든 디바이스에 관해서 ESI 수집이 이루어져야 한다. ESI를 수집하는 경우에도 ESI의 접근성과 수집 비용들이 충분히 고려돼야하며 ESI 특성상 다양한 형태의 ESI가 존재하기 때문에 이를 수집하기 위해서 여러 가지 방법들이 고려되어야 한

다. 예를 들어 전자문서나 전자메일 같은 경우 메일 검색과 보존에 관한 사항이 포함된 정보관리 정책을 적용하면 ESI 수집에 소요되는 시간과 비용을 줄일 수 있지만 그렇지 않은 경우 수집에 들어가는 비용과 시간은 ESI의 양에 비례해서 커지게 된다. 또한, 소셜 네트워크 서비스 데이터나 스마트 폰에 있는 ESI 경우 별도의 추출과정이 필요로 한다. 또한, 삭제되거나 파괴된 ESI를 수집하기 위해서 포렌식적 기법을 사용할 수도 있으며 수집된 ESI에 무결성이 보장되어야 한다.

나. 보존 요구사항

ESI를 보존하기 위한 요구 사항을 도출하면 아래 목록과 같다.

- 조직 구성원들에게 ESI 보존의무를 공지
- Litigation hold 정책 수립 및 시행
- ESI 파괴와 변조를 막기를 위한 물리적 및 소프트웨어적 ESI 보호 조치 시행
- 별도의 ESI 보관 스토리지를 구축
- Litigation Hold의 주기적인 모니터링

ESI를 보존하기 위해서는 소송에 필요한 ESI가 사전에 식별되어 있어야 하며 식별된 ESI에 대해 삭제, 변경, 파괴를 막아야 한다. Litgation hold는 ESI의 파괴와 변조를 막는 절차로 조직구성원들에게 소송이 시작됨을 공지하고 ESI를 보호할 것을 요청한다. 하지만 단순히 조직구성원들에게 Litigation hold를 공지하는 것으로 ESI를 모두 보호할 순 없다. ESI를 보호하기 위해서는 물리적인 방법과 소

프트웨어적인 방법을 동원하여 ESI를 파괴를 막을 수 있어야 한다. 예를 들어 검토와 분석을 위해 자주 열람 되는 ESI들은 별도의 스토리지에 보관하고 사본을 이용함으로써 열람에 의한 ESI 변경과 파괴를 막을 수 있어야 한다. 또한, Litigation hold는 문서화된 정책을 수립하고 주기적으로 모니터 되어야 한다.

IV. e-Discovery 지원 도구 설계 및 구현

본 장에서는 기존의 e-Discovery 지원 도구에서 지원하지 않는 몇 가지 ESI 타입에 대해서 설명하고 이를 수집하고 보존하는 방법에 대해서 설명한다. 그리고 이를 고려하여 e-Discovery 지원 도구를 설계하고 프로토타입을 구현해 본다.

1. ESI 특성에 따른 수집과 보존의 고려 사항

ESI의 종류는 디지털 디바이스에 저장된 모든 정보이므로 그 종류가 매우 다양하다. 전자메일이나 전자문서 같은 가장 일반적인 디지털 증거부터 소셜 네트워크 데이터, 스마트 디바이스 데이터 등 다양한 위치에 다양한 형태로 존재한다. 현재 출시되어 있는 e-Discovery 관련 솔루션들은 대부분 전자문서와 전자메일에 초점이 맞추어져 있으며 스마트 디바이스와 소셜 네트워크 데이터 수집 및 분석에 대한 지원은 미비한 실정이다. 본 절에서는 스마트 디바이스 OS 중 가장 널리 보급되어 있는 Apple[22]사의 iOS에서의 ESI 추출 방법을 설명하고 소셜 네트워크 서비스로 잘 알려진 Twitter와 Facebook에서의 ESI 추출 기법을 설명한다.

가. iOS

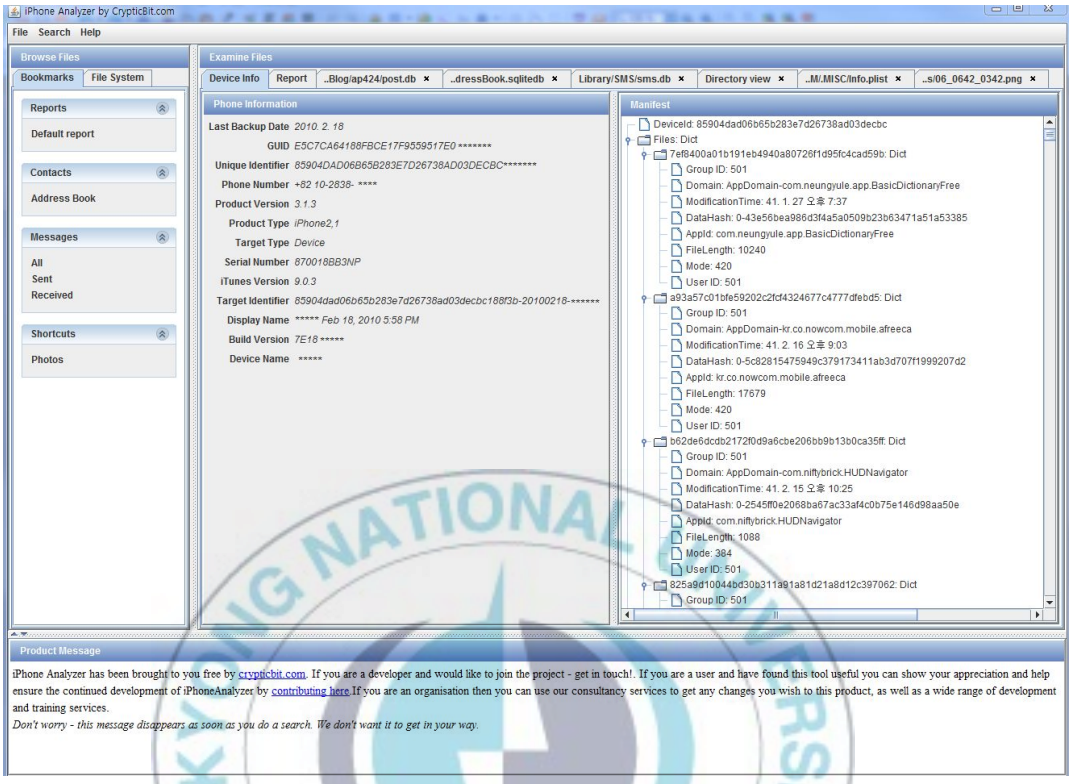
iOS는 Apple사에서 개발한 스마트 디바이스를 위한 운영체제로 아이폰, 아이팟, 아이패드에 탑재되어 있다. 특히 아이폰은 2011년 2월 현재 아이폰은 스마트폰 시장에서 점유율 2위를 기록하고 있다

[23]. 스마트폰은 다양한 기능으로 폭넓은 활용을 할 수 있으므로 기업 내에서 업무용으로도 많이 사용된다. 따라서 스마트폰 내에도 업무와 관련된 자료들이 존재하며 이는 잠재적인 디지털 증거가 될 수 있다. 아이폰의 경우 아이튠즈[24]를 통해 PC와 동기화해서 사용 할 수 있는데 이를 이용하여 아이폰의 데이터를 추출할 수 있다. 아이폰의 데이터가 아이튠즈를 통해 백업 되어 저장되는 경로는 <표 6>과 같다.

<표 6> 아이튠즈를 통한 백업 데이터 저장 경로

항목	경로 - 윈도우 7	윈도우 XP
연락처, Wi-Fi 접속 내역, 노트, 어플리케이션 캐시	C:\Users\<계정명>\AppData\Roaming\Apple Computer\MobileSync	C:\Documents and Settings\<계정이름>\Application Data\Apple Computer\MobileSync\Backup\

아이튠즈에 저장된 데이터들은 SQLite, XML, 각종 멀티미디어 포맷으로 구성되어 있다. SQLite에는 연락처, SMS, 같은 정보들이 포함되어 있으며, 디바이스의 정보와 각종 로그 기록들은 XML 형태로 저장된다. 아래 [그림 5]는 아이폰 백업 데이터 분석 도구인 iPhone Analyzer[25]를 통해 아이폰 백업 데이터를 열람한 내용으로 디바이스 정보와 동기화 날짜 아이튠즈 버전 및 iOS의 버전을 보여 준다. 백업 데이터에는 디바이스 정보, SMS, 통화 송수신 내역, 연락처, WiFi 접속 내역, 응용프로그램 캐쉬 데이터, 이미지, 동영상 등 아이폰에 있는 모든 데이터가 저장된다. 이러한 데이터들은 잠재적인 디지털 증거들이므로 별도의 스토리지에 백업을 하거나 소프트웨어적인 방법으로 잠금을 하는 등 수정, 삭제 변경을 막는 조취를



[그림 5] iPhone Analyzer를 통한 아이폰백업 자료 확인

취해야 한다.

나. Twitter

SNS의 사용량이 증가함에 따라 SNS가 디지털 증거로써의 가치가 높아지고 있다. 그 예로 앞서 언급한 국내에서는 Twitter를 이용하여 선거법을 위반한 사례를 들 수 있다. 이 사례를 통해 Twitter가 법정에서 증거로써의 효력을 미친다고 볼 수 있다. 따라서 SNS 데이터도 법적 증거가 될 수 있다. Twitter는 마이크로 블로그의 일종으로 140자의 tweet을 상대방에게 보여주는 소셜 네트워크 서비스의 일종이다. 대다수 기업들은 Twitter를 마케팅 수단으로 이용하고 있

으며 특히 DELL은 국가별로 33개의 Twitter 계정을 운영하고 있다.

Twitter는 오픈 API를 이용해 사용자가 별도의 클라이언트를 만들어서 사용하거나 모바일 환경에서 사용하기 위해 별도로 정보를 가공할 수 있다. 이러한 오픈 API의 특성을 이용하면 Twitter의 tweet 데이터를 수집할 수 있다.

Twitter에서 수집해야 할 주요 데이터로는 tweet 데이터와 DM(Direct Message)이 있다. tweet 데이터의 경우 사용자가 protect 설정을 하지 않는 경우 별도의 로그인 과정 없이 Twitter 계정만 알고 있으면 수집 가능하나 DM은 반드시 로그인해야만 확인 할 수 있다. protect는 자신의 tweet을 팔로우들에게만 보여주게 하는 옵션으로 protect 된 내용을 확인 하기 위해선 로그인 과정을 거쳐야 한다. Twitter API를 사용하기 위해서는 사전에 Twitter 홈페이지에서 API 사용에 관한 등록을 거쳐야 한다. Twitter의 사용자 인증방식은 OAthu를 이용하며 C++,C#등 다양한 라이브러리를 지원한다. tweet 데이터를 수집, 보존하기 위해서는 먼저 tweet 데이터를 스토리지에 저장 가능한 형태로 가공해야 한다. 아래 <표 7>은 OAthu의 라이브러리인 libauth를 이용하여 tweet를 XML로 추출한 것이다 [26].

<표 7> XML로 변환한 Tweet 데이터

```
<?xml version="1.0" encoding="UTF-8" ?>
- <statuses type="array">
- <status>
<created_at>Mon Jan 10 02:43:47 +0000 2011</created_at>
<id>24295307709779968</id>
<text>TEST tweet</text>
<source>web</source>
```

```
  <truncated>false</truncated>
  <favorited>false</favorited>
  <in_reply_to_status_id />
  <in_reply_to_user_id />
  <in_reply_to_screen_name />
  <retweet_count>0</retweet_count>
<retweeted>false</retweeted>
</status>
  </statuses>
```

XML 데이터를 분석해 보면 tweet 내용, 사용자 아이디, 리트윗 여부, 스크린 네임, 생성한 날짜 등 tweet의 내용이 모두 포함되어 있다. 수집한 tweet 데이터는 별도의 ESI 스토리지에 보관하고 수정 및 삭제가 되지 않도록 주기적으로 모니터링 되어야 한다.

다. Facebook

Facebook은 프로필 기반의 소셜 네트워크 서비스로 최근 가입자가 5억 명을 넘어섰다. Facebook도 Twitter와 마찬가지로 삼성, HP 등 여러 기업에서 계정을 운용하고 있다. Facebook은 자기의 WALL을 통해 사용자가 메시지를 남기거나 멀티미디어 파일을 업로드 할 수 있으며 다른 사용자들이 WALL에 코멘트를 남기거나 추천하는 형태로 사용된다. Facebook에서도 수집해야 할 주요 데이터로는 WALL 데이터와 쪽지기능인 Message 데이터를 수집해야 한다. Facebook도 Twitter와 마찬가지로 오픈 API를 통해 별도의 클라이언트를 만들어 사용할 수 있다. 이를 사용하기 위해서는 등록 과정을 거쳐야 한다. 기본적으로 API는 PHP로 제공이 되고 있으며 C#, 자바스크립 등을 지원한다. 아래 <표 8>은 PHP용 API를 이용하여 사

용자 정보를 조회한 결과이다.

<표 8> Facebook 유저 데이터

id	100001990531531
name	Sns EDiscovery
first_name	Sns
last_name	EDiscovery
link	http://www.facebook.com/profile.php?id=100001990531531
hometown	[object Object]
location	[object Object]
gender	male
email	jubo****@gmail.com
timezone	9
locale	ko_KR
verified	true
updated_time	2011-01-10T09:01:31+0000

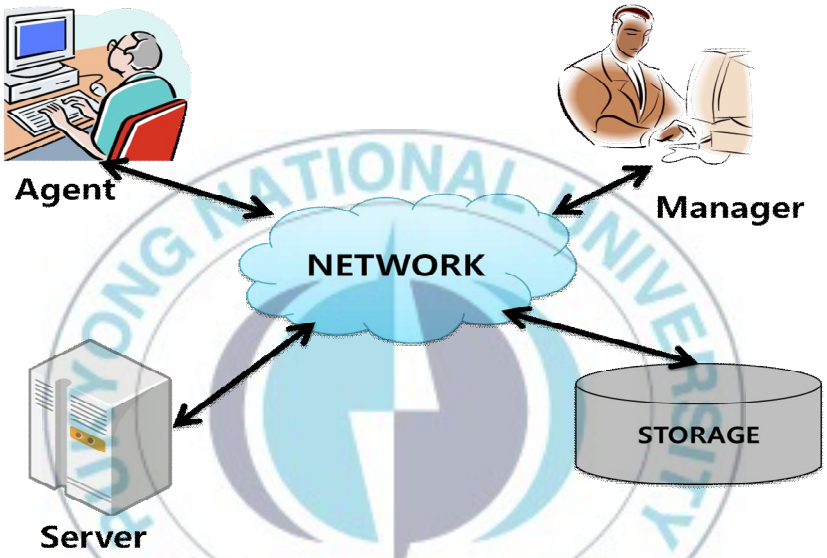
2. e-Discovery 지원 도구 설계

현재 시중에 나와 있는 e-Discovery 지원 도구들은 모두 외산 솔루션들만 존재하며 SNS 데이터를 수집하는 e-Discovery 지원 솔루션들은 전무한 실정이다. 본 논문에서는 e-Discovery 지원 도구를 설계하여 e-Discovery 지원 도구 국산화의 기반을 만들고 SNS를 수집하는 기능을 추가하여 ESI 수집 기능을 확장해 보았다.

본 논문에서 설계한 e-Discovery 지원 도구는 ESI의 수집과 보존에 초점이 맞추어져 있으며 향후 분석과 검토 단계를 수행하는 것을 고려하여 전체 시스템을 구성하였다. 또한, 기존의 지원 도구에서는 지원하지 않는 SNS 데이터 수집과 iPhone 데이터 수집을 고려하여 설계 하였다.

가. 전체 시스템 구성

e-Discovery 지원도구는 Agent, Server, Manager의 3가지 모듈로 구성된다. 아래 그림 [그림 6]은 본 논문에서 설계한 e-Discovery 지원도구의 전체 시스템 구성도를 나타낸다.



[그림 6] e-Discovery 지원도구 전체 시스템 구성도

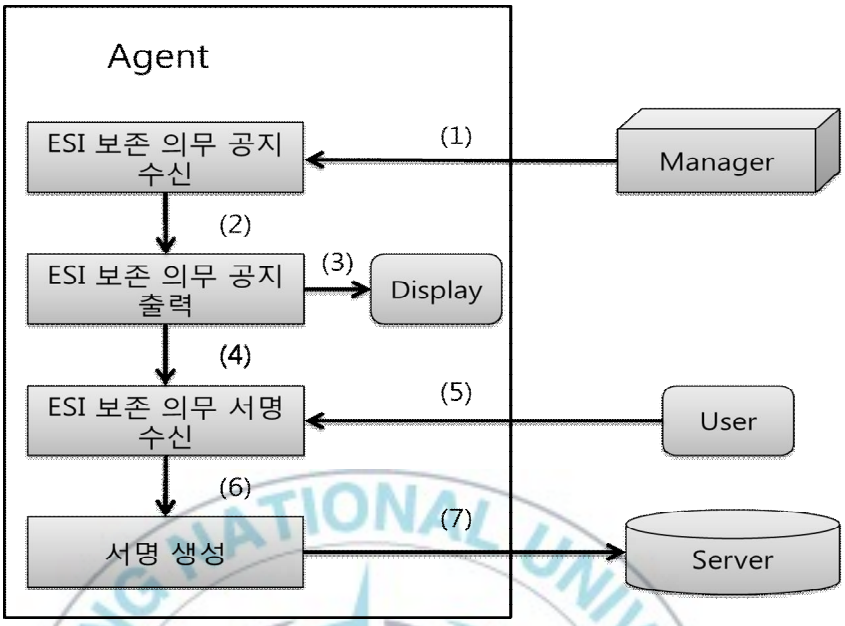
Agent는 PC에 설치되어 Manager로부터 수신받은 명령을 수행하고 Server는 각종 ESI를 스토리지에 저장하는 역할을 한다. Manager는 ESI 보존과 수집에 관한 각종 정책을 수립하고 수립한 정책들을 Agent에게 적용한다. 3가지 모듈은 TCP/IP를 통해서 통신하게 되며 무결성 보장을 위해 모든 패킷은 암호화되어 전송된다.

나. Agent

Agent는 ESI를 수집하고 보존하는데 그 역할이 초점이 맞추어져 있다. 따라서 PC에서 ESI를 검색하기 위해서는 파일의 이름뿐만 아니라 확장자, 생성날짜, 본문 내용도 검색할 수 있어야 한다. 파일 시스템에 있는 모든 ESI에 대해 앞서 언급한 검색 옵션들을 적용하여 검색을 하기 위해선 많은 시간이 소모되므로 사전에 ESI에 대한 인덱스를 생성하는 것이 필요하다. 따라서 인덱스 정책을 통해 사전 인덱스를 생성해야 하며 사전 인덱스 정책엔 앞서 언급한 검색 옵션들을 모두 포함해야 한다. Agent가 가지는 기능은 아래 목록과 같다.

- ESI 보존 의무 공지
- Litigation hold 기능
- 실시간 ESI 검색 및 ESI 전송 기능
- SNS 계정 수집 기능

ESI 보존 의무 공지는 Manager로부터 수신받은 보존의 의무 메시지를 출력한다. [그림 7]은 ESI 보존 의무공지 기능의 흐름도를 나타낸 것이다.



[그림 7] ESI 보존 의무 공지 기능 흐름도

ESI 보존 의무 공지 기능은 소송에서 ESI 보존에 대한 의무를 공지하고 조직구성원들이 인지함으로써 ESI 보존하는 기능으로 흐름도의 각 단계에서 실행되는 내용은 아래 리스트와 같다.

(1) Manager로부터 ESI 보존 의무 공지를 수신한다.

(2) 수신받은 ESI 보존 의무공지 기능을 실행한다.

(3) 수신받은 ESI 보존 의무공지를 사용자에게 출력한다.

(4) 사용자가 ESI 보존 의무를 충분히 인지했음을 확인하기 위해 응답서명을 받는다.

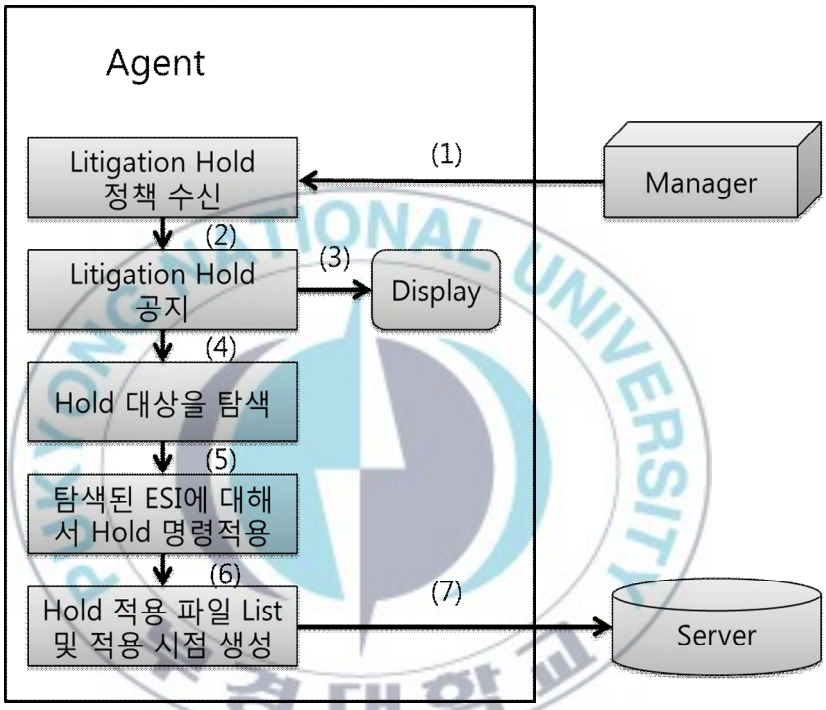
(5) User가 응답 서명을 Agent에게 입력하도록 한다.

(6) 응답 서명을 Server로 전송하기 위해 생성한다.

(7) 생성된 응답 서명을 Server로 전달한다. 생성된 응답 서명은 Manager가 확인하여 각 Agent에 대해 ESI 보존 공지의

인지 여부를 파악하는데 사용된다.

Litigation Hold 정책 수신 기능은 소송이 예측되는 시점으로부터 ESI를 보호하기 위한 기능이다. [그림 8]은 Litigation Hold 정책의 흐름도를 나타낸 것이다.



[그림 8] Litigation Hold 기능 흐름도

Litigation Hold 기능은 Manager로부터 받은 Litigation Hold 정책을 토대로 Litigation Hold 기능을 실행한다. Litigation Hold 정책에는 Hold 할 ESI에 관한 키워드(파일명, 본문내용 등을 포함), 생성 날짜, 확장자 등을 포함한다. Litigation Hold가 실행된 기록은 로그로 생성하여 Server로 전송한다. 로그에는 Hold 가 적용된 시점과 Hold 대상들이 기록되며 Hold를 해제한 시점도 기록한다. 흐름도의 각 단계에서 실행되는 내용은 아래 리스트와 같다.

(1) Manager로부터 Litigation Hold 정책을 수신한다.

(2) Litigation Hold 공지 기능을 실행한다.

(3) Litigation Hold 공지를 사용자에게 출력한다.

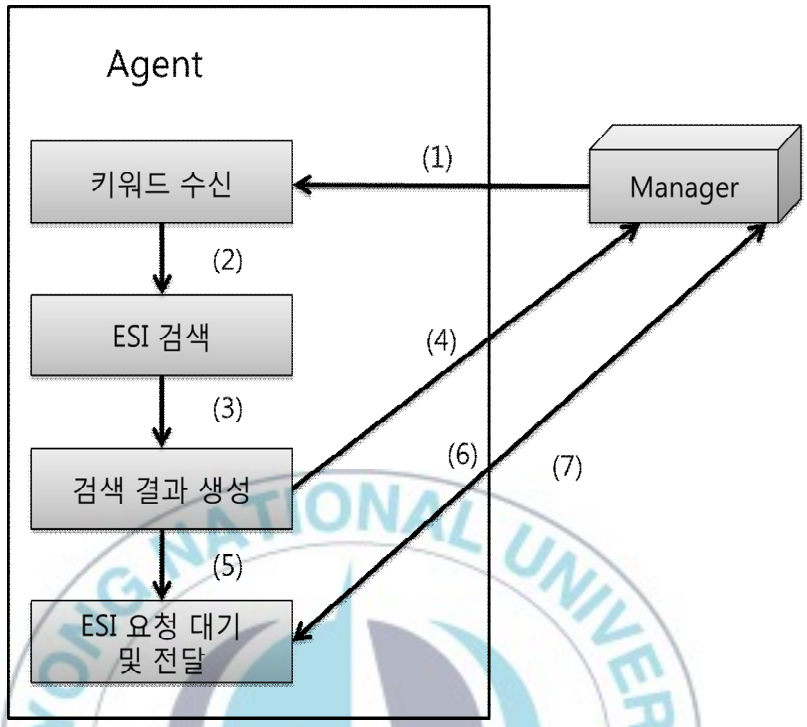
(4) Hold 할 ESI를 탐색한다.

(5) 탐색 된 ESI에 대해서 Hold 명령을 적용한다. Hold 명령이 내려진 ESI는 삭제 및 변경이 되지 않도록 소프트웨어적 조치가 취해진다.

(6) Litigation Hold 명령에 대한 로그를 생성한다. 로그에는 Hold가 적용된 ESI 리스트들과 적용시점 등이 기록된다.

(7) 생성된 Litigation Hold 로그를 서버로 전송한다. Manager는 서버에 있는 로그를 토대로 각 Agent에 대한 Litigation Hold 내역을 확인할 수 있다.

Manager가 ESI 검토가 필요한 경우 Agent로부터 ESI를 직접 전송받아 확인할 필요가 있다. 따라서 Agent는 Manager에게 ESI를 직접 송신할 수 있는 ESI 전송 기능이 필요하다. ESI 전송 기능을 흐름도로 나타내면 [그림 9]와 같다.



[그림 9] 실시간 ESI 검색 및 ESI 전송 기능 흐름도

Manager로부터 수신받은 키워드를 토대로 ESI를 검색한다. ESI는 검색은 사전에 색인 된 자료 외의 자료도 검색해야 하기 때문에 인덱스를 거치지 않고 ESI를 검색한다. 실시간 ESI 검색 및 ESI 전송 기능의 흐름도의 각 단계에서 실행되는 내용은 아래 리스트와 같다.

(1) Manager로부터 검색에 사용할 키워드를 수신받는다.

(2) Agent는 해당 파일을 색인을 거치지 않고 검색한다.

(3) Manager로 전달하기 위해 검색 결과를 생성한다.

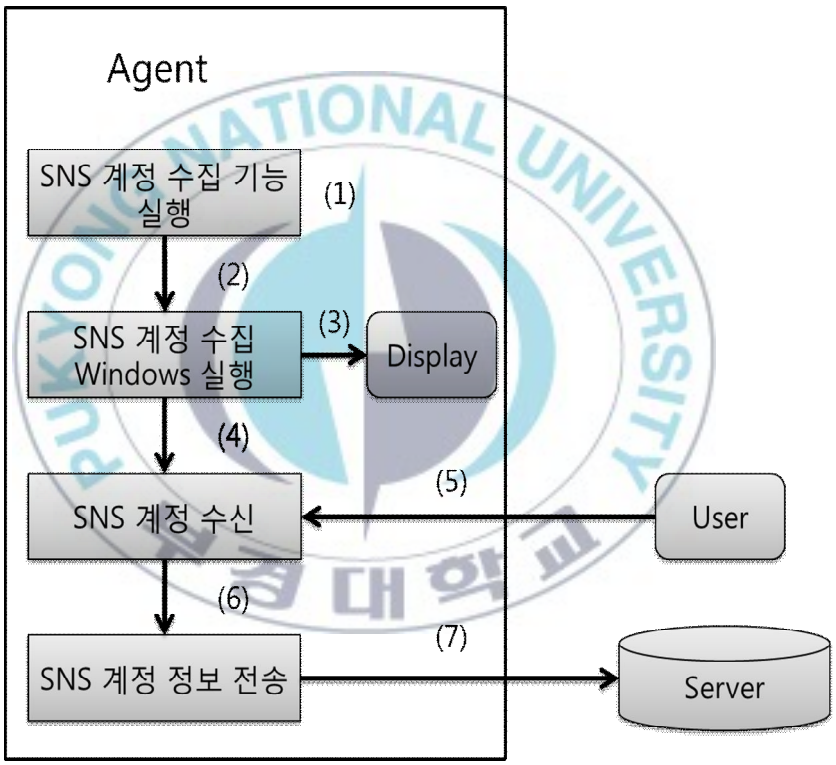
(4) 생성된 검색 결과를 Manager로 전달한다.

(5) Manager가 검색 결과를 토대로 필요한 ESI를 요청할 때 까지 대기한다.

(6) Manager로부터 ESI 요청 통보를 수신 한다.

(7) Manager가 요청한 ESI를 전송한다.

SNS 계정 수집 기능은 조직 구성원들에게서 수집한 SNS 계정을 토대로 ESI를 수집하여 서버로 전송하는 기능이다. 전송된 SNS 계정은 Manager를 통해서 SNS 데이터를 XML 형태로 추출하여 저장한다. [그림 10]은 SNS 계정 수집 기능의 흐름도를 보여준다.



[그림 10] SNS 계정 정보 수집 기능 흐름도

SNS 계정 정보는 Server로 전송되며 Manager는 수집된 Server로부터 수신받은 SNS 계정 정보를 토대로 SNS 데이터를 수집한다. 현재는 Twitter와 Facebook에 대해서만 SNS 데이터를 수집한다. 각 단계에서 실행되는 내용은 아래 리스트와 같다.

(1) SNS 계정 수집 기능을 실행한다.

(2) SNS 계정 입력 폼을 생성한다.

(3) SNS 계정 입력 폼을 사용자에게 출력한다.

(4) 사용자로부터 SNS 계정을 입력받을 때까지 대기한다.

(5) 사용자로부터 SNS 계정을 입력받는다.

(6) SNS 계정을 Server로 전달하기 위해 준비한다.

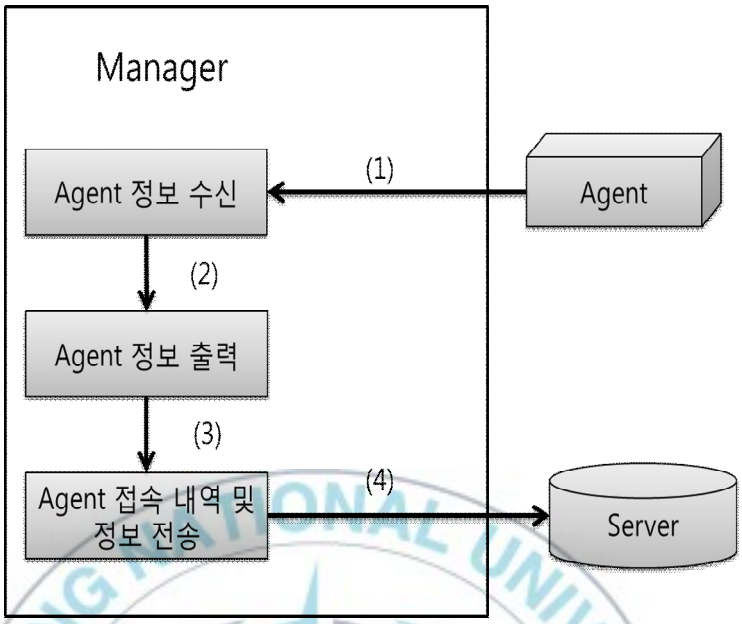
(7) SNS 계정을 Server로 전송한다.

다. Manager

Manager는 Agent에게 적용할 각종 정책을 수립하고 Agent를 관리하는 역할을 한다. Manager는 Server를 통하여 Agent의 각종 내역들을 확인하며 필요할 경우 각각의 Agent에게 명령을 내릴 수 있다. Manager의 주요기능은 아래 리스트와 같다.

- Agent 접속 관리 기능
- 보존의무공지 기능
- Litigation Hold 정책 수립 및 명령 전송 기능
- ESI 검색 및 수신 기능
- SNS 데이터 수집 기능

Agent 접속 관리 기능은 각각의 Agent의 정보를 Manager로 전송하고 Agent 접속 여부를 확인하는 기능이다. [그림 11]은 Agent 접속 관리 기능의 흐름도를 보여준다.



[그림 11] Agent 접속 관리 기능 흐름도

Agent 접속 기능은 Agent 의 접속 여부를 파악하고 접속 한 Agent의 IP, 접속한 시간, 사용자, MAC address 등을 수신한다. 수신한 정보는 로그로 남겨 Server로 전송한다. 아래 리스트는 흐름도의 각 단계가 실행하는 내용이다.

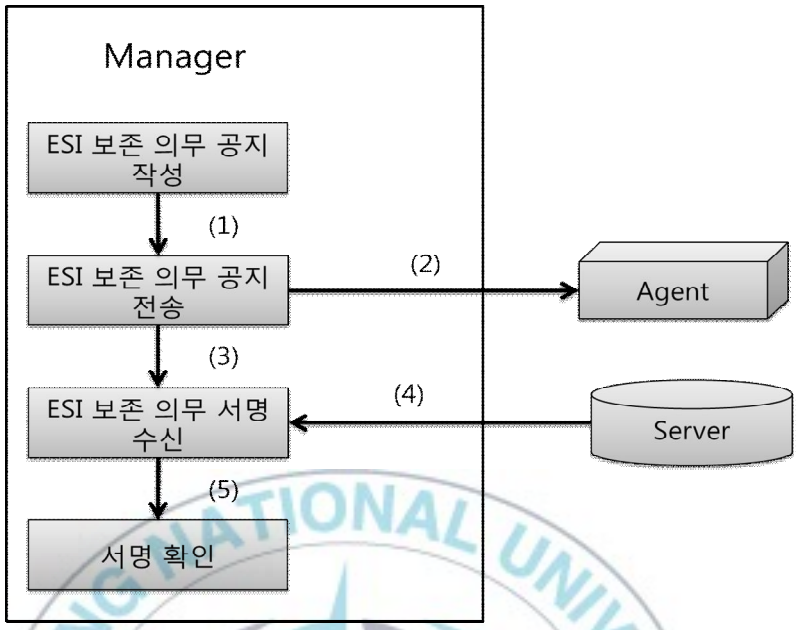
(1) Agent로부터 IP, MAC address, OS, 사용자, Agent ID,

(2) 수신받은 정보를 출력한다.

(3) 수신받은 정보를 로그로 생성한다.

(4) 생성된 로그를 Server로 전달한다.

보존 의무 공지 작성 기능은 Manager가 Agent에게 보존의 의무를 공지하는 기능이다. Manager는 Agent에게 보존의 의무를 전달하고 사용자가 공지를 인지했는지 확인한다. [그림 12]는 보존 의무 공지 작성 기능의 흐름도를 나타낸 것이다.



[그림 12] ESI 보존의무공지 기능 흐름도

Manager는 Server에서 수신받은 Agent의 보존 공지 확인 서명을 토대로 Agent의 공지 확인 여부를 판단한다. 흐름도의 각 단계에서 실행하는 내용은 아래 리스트와 같다.

(1) Agent로 보낼 ESI 의무공지를 작성한다.

(2) 작성한 공지를 Agent로 전송한다.

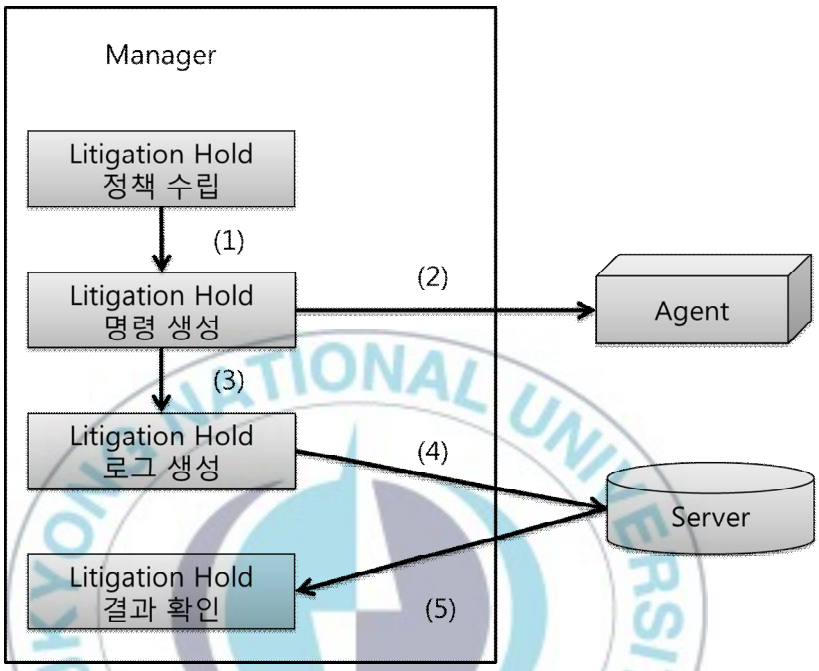
(3) Server로부터 확인 서명을 수신 받을 때까지 대기한다.

(4) Server로부터 확인 서명을 수신받는다.

(5) Agent의 서명을 확인한다. 서명하지 않은 Agent를 파악 하고 메시지를 보내는 등의 조치를 취한다.

Litigation Hold 정책 수립 및 명령 전송 기능은 ESI를 보호하기 위해 Litigation Hold 정책을 수립하고 수립한 정책을 Agent에게 전

달하는 기능이다. [그림 13]은 Litigation Hold 정책 수립 및 명령 전송 기능의 흐름도이다.



[그림 13] Litigation Hold 정책 수립 및 전송 기능 흐름도

Litigation Hold 정책은 앞서 언급했듯이 파일 확장자, 키워드, 생성날짜 등을 토대로 수립되며 사전에 생성된 색인을 통해서 Hold 대상을 지정한다. 수립된 Litigation Hold 정책은 Agent로 전송되며 Litigation Hold 명령은 로그로 남겨져 Server로 저장된다. 로그는 Litigation Hold가 적용된 Agent 리스트, Litigation 정책 등이 포함된다. 흐름도에서 각 단계가 실행하는 내용은 아래 리스트와 같다.

(1) Litigation Hold 정책을 수립한다.

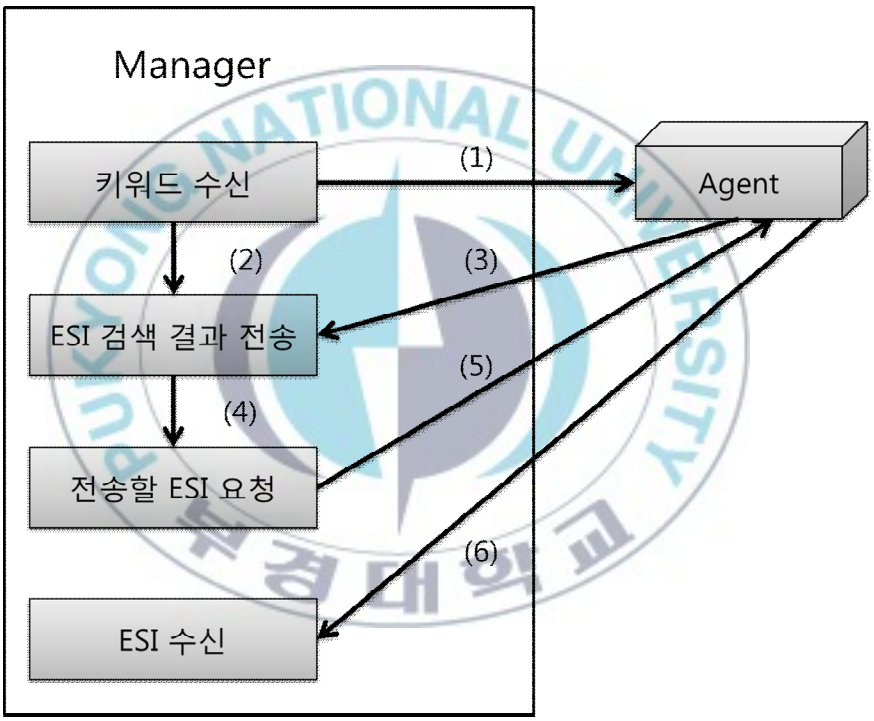
(2) 수립한 정책을 Agent에게 전달한다.

(3) Litigation Hold 로그를 생성한다.

(4) 생성된 로그를 Server로 전송한다.

(5) Server로부터 Litigation Hold 결과를 통보받는다.

ESI 검색 및 수신 기능은 Agent에 사전에 색인되지 않은 데이터를 검색하고 검토에 필요한 ESI를 직접 수신하는 기능이다. ESI 검색 및 수신 기능의 흐름도는 [그림 14]와 같다.



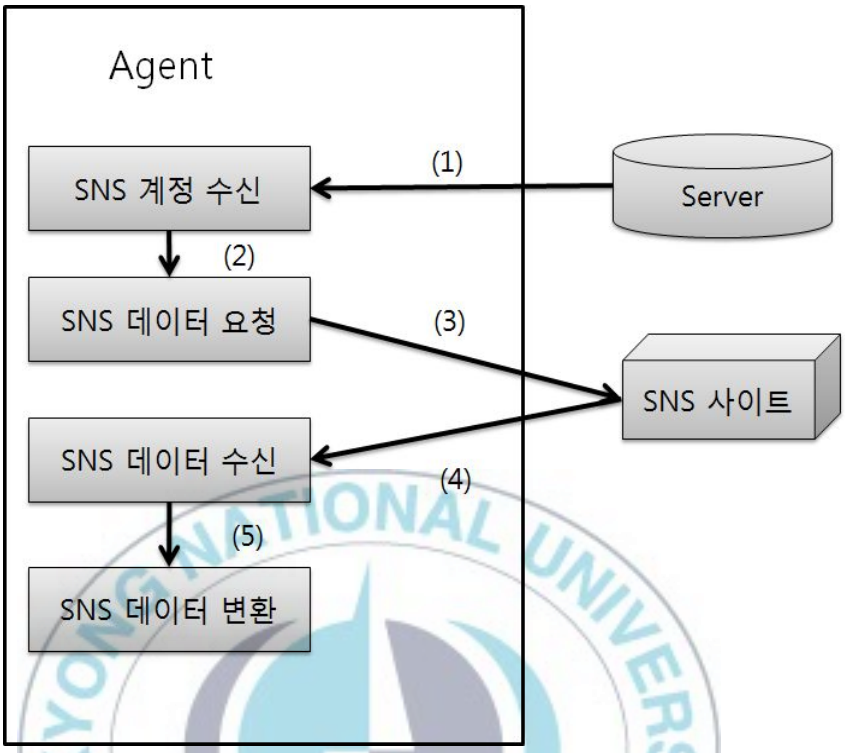
[그림 14] 실시간 ESI 검색 및 ESI 수신 기능 흐름도

실시간 ESI 검색 및 ESI 수신 기능은 Manager가 Agent로 키워드를 전달하고 Agent는 키워드를 토대로 ESI의 검색 결과를 Manager에게 전달해 주게 된다. Manager는 검색 결과를 토대로 필요한 ESI를 Agent에게 수신 요청을 보내고 Agent는 해당 ESI를 전

송한다. 흐름도의 각 단계별 내용은 아래 리스트와 같다.

(1) Agent에게 키워드를 전달한다.
(2) Agent로부터 ESI 검색 결과를 전송받을 때까지 대기한다.
(3) Agent로부터 ESI 검색 결과를 전송받는다.
(4) 검색 결과를 토대로 필요한 ESI를 선별한다.
(5) 선별한 ESI를 Agent에게 요청한다.
(6) Agent로부터 ESI를 수신한다.

SNS 데이터 수집 기능은 Agent로부터 수집한 SNS 계정을 토대로 SNS 데이터를 수집하는 기능이다. 이 기능은 Twitter와 Facebook을 타켓으로 설계하였으며 Twitter의 tweet 데이터와 Facebook의 WALL 데이터를 수집한다. 이 기능의 흐름도는 아래 [그림 15]와 같다.



[그림 15] SNS 계정 수집 기능

SNS 계정 수집 기능은 SNS의 오픈 API의 특성을 이용하여 수집한다. 수집 된 SNS 데이트는 검토 및 보관 가능한 형태로 변환해야 한다. 대부분의 웹 베이스의 SNS 서비스는 XML을 지원함으로 XML 형태로 변환하여 저장하게 된다. 흐름도에서 각 단계에서 실행하는 내용은 아래 리스트와 같다.

(1) Server로부터 SNS 계정을 수신 받는다.

(2) 수집된 SNS 계정을 토대로 SNS 데이터를 요청하기 위해 준비한다.

(3) 수집된 계정을 토대로 오픈 API를 통해 SNS 데이터를 요청

한다.

(4) SNS 사이트로부터 SNS 데이터를 수신받는다.

(5) 수신받은 SNS 데이터를 XML 형태로 변환한다.

라. Server

Server는 Manager와 Agent의 연결해주는 역할을 수행한다. Server의 주요 기능은 각 기능을 실행할 때 생성되는 로그를 저장하고 저장된 로그를 Manager가 검토할 수 있게 한다. Server는 데이터베이스 시스템과 함께 동작한다. Server가 가지는 주요 기능들은 아래 목록과 같다.

- Litigation Hold 정책 저장 기능
- 보존의무공지 저장 기능
- Agent 정보 저장 기능
- SNS 계정 저장 기능

Server는 Agent와 Manager가 동작할 때마다 모든 이벤트를 기록하게 된다. 기록된 로그는 XML 타입으로 저장되며 데이터베이스에 삽입되어 조회할 수 있어야 한다.

3. e-Discovery 지원 도구 구현

앞서 도출한 요구사항과 설계한 내용을 토대로 e-Discovery 지원 도구의 일부 기능을 포함하는 프로토타입을 구현해 보았다. 구현 환경은 운영체제는 Windows 7을 이용하였으며 Visual Studio 2010과 C#,

.Netframework 3.5를 이용했으며 Windows Search SDK 3.x를 이용하여 구현하였다. Agent와 Manager 모듈을 구현하였으며 색인 기능, Litigation Hold 기능, SNS 계정수집 기능을 직접 구현한다.

가. Agent 색인 기능 구현

대다수의 e-Discovery 지원 도구들은 프로그램 설치와 함께 색인 작업을 시작한다. 시스템 내 모든 문서를 색인하기 위해서는 많은 시간을 소요하기 때문에 별도의 인덱스 정책을 수립하여 해당하는 ESI에 한해서 색인하는 것이 일반적이다. 앞서 설명했듯이 ESI 색인 정책은 파일 확장자, 키워드, 파일 생성날짜 등으로 보통 지정하게 된다. 이러한 색인 정책은 몇 가지 문제점을 가지게 된다. 첫 번째로 모든 ESI에 대해서 색인 할 경우 많은 시간을 소요하게 된다. ESI는 사용자 PC 안에 있는 모든 파일이 그 대상이 될 수도 있기 때문에 소송이 예측되거나 시작되는 시점에서 색인을 생성해야 할 경우 많은 시간을 소요하게 된다. 두 번째로는 색인 자료 밖에 있는 데이터를 검색하기가 어려워진다. 색인 정책에 의해 색인되지 않은 자료를 검색하기 위해서는 전체 시스템을 모두 검색해야 함으로 검색에 소요되는 시간이 길어지게 된다. 세 번째로 새로운 파일이 생성될 경우 색인 작업을 다시 수행해야 한다. 이러한 문제점을 해결하기 위해서는 Windows에 내장된 검색 기능을 사용하는 방법이 대안이 될 수가 있다.

Windows 7과 Vista Windows Search가 설치되어 있어서 전체 시스템을 색인한다. 따라서 별도의 색인 프로그램을 구동하지 않아도 색인 결과를 얻을 수 있다. Windows Search의 색인 결과는 아래 <표 9>와 같은 디렉터리에 존재한다.

<표 9> Windows 버전별 색인 경로

운영체제	경로
Windows Vista	%Profiles%\All Users\Application Data\Microsoft\Search\Data\Applications\Windows\Windows.edb
Windows 7	%ProgramData%\Microsoft\Search\Data\Application\Windows\Windows.edb

색인 파일은 다양한 형태의 파일을 색인한다. pdf, docx, pptx, txt 등의 문서 파일을 지원할 뿐만 아니라 zip, xml 형태의 자료로 색인한다. 색인에는 파일 포맷, 생성날짜뿐만 아니라 파일의 내용도 색인을 함으로 보다 높은 정확한 검색결과를 얻을 수 있다. 윈도우 색인 파일은 확장자는 edb로 데이터베이스 파일이다. 따라서 검색하기 위해서는 질의어를 사용해야 하며 따라서 다양한 옵션으로 검색할 수 있다. 예를 들어 한 문단에서 동시에 존재하는 두 개의 키워드를 포함하는 문서들도 찾아낼 수가 있다. Windows Search 4.0 기능을 사용하기 위해서는 Windows Search SDK 3.x가 필요하다. 색인 파일이 데이터베이스 파일 형태이므로 OLE DB를 이용하여 색인에 접근하여 질의어를 전달하여 검색 결과를 얻는다.

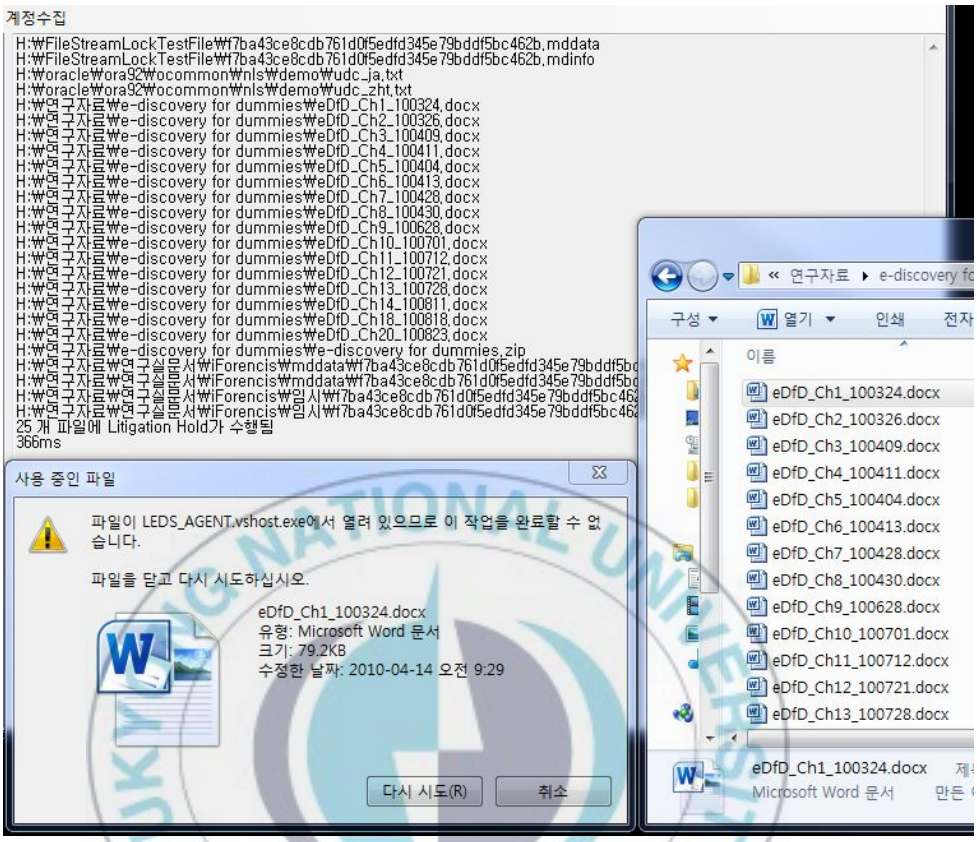
본 프로토타입은 Manager에 Agent가 접속하면 Manager는 Agent에게 질의어를 전달하고 Agent는 질의어로 색인을 검색한 뒤에 검색결과를 Manager에게 전달한다. Manager는 Agent로부터 수신받은 결과를 토대로 Agent에 어떠한 파일이 있는지 확인한다. 그림[16]은 Manager로부터 EDRM 이라는 키워드를 수신받은 Agent의 결과를 Manager에서 확인한 것이다.

[그림 16] 색인 검색 결과

Agent의 색인 크기는 약 860MB로 보통 230,000개의 파일의 인덱스는 800메가바이트의 크기를 차지한다[27]. EDRM을 검색할 경우 164개의 파일이 검색되고 1,642ms가 걸린 반면 e-Discovery를 검색할 경우 680여개의 파일이 검색되고 약 260ms 가 소요된다. 이는 파일 색인이 검색되는 파일에 유형과 내용에 따라 조금씩 차이를 보이기 때문이다.

나. Litigation Hold 기능 구현

Litigation Hold 기능은 ESI 삭제 및 변경을 막는 기능이다. Litigation Hold 기능을 구현하기 위해서 파일을 특정 프로세스가 오픈하여 다른 파일이 접근하지 못하도록 하는 방법을 사용한다. 프로세서에 의해 파일이 오픈이 되어 있으면 파일이 메모리에 적재된 상태가 되므로 하드디스크에 있는 파일의 삭제 및 변경이 불가능하게 된다. Manager는 접속된 Agent에게 Litigation Hold 명령을 내리게 되고 Agent는 해당 키워드가 존재하는 ESI를 검색하고 검색된 파일에 대해서 Litigation Hold 명령을 수행한다. 검색은 원래 파일 시스템에 있는 모든 디렉터리와 파일을 각각 조회해야 하지만 앞서 언급한 Windows Search를 이용하면 검색에 소요되는 시간을 줄일 수 있다. Windows Search는 새롭게 파일이 생성된 경우 색인에 바로 반영하여 이를 검색할 수 있도록 해준다. 그림 [17]은 키워드에 의해 검색된 파일들에 Litigation Hold가 수행되고 있는 것을 보여준다. 25개의 파일을 Lock을 실행했는데 총 366ms 가 소요되었다. Litigation Hold가 수행되고 있는 중에는 파일의 삭제 변경이 불가능하며 파일을 삭제할 경우 "사용 중인 파일"이라는 메시지가 출력된다.

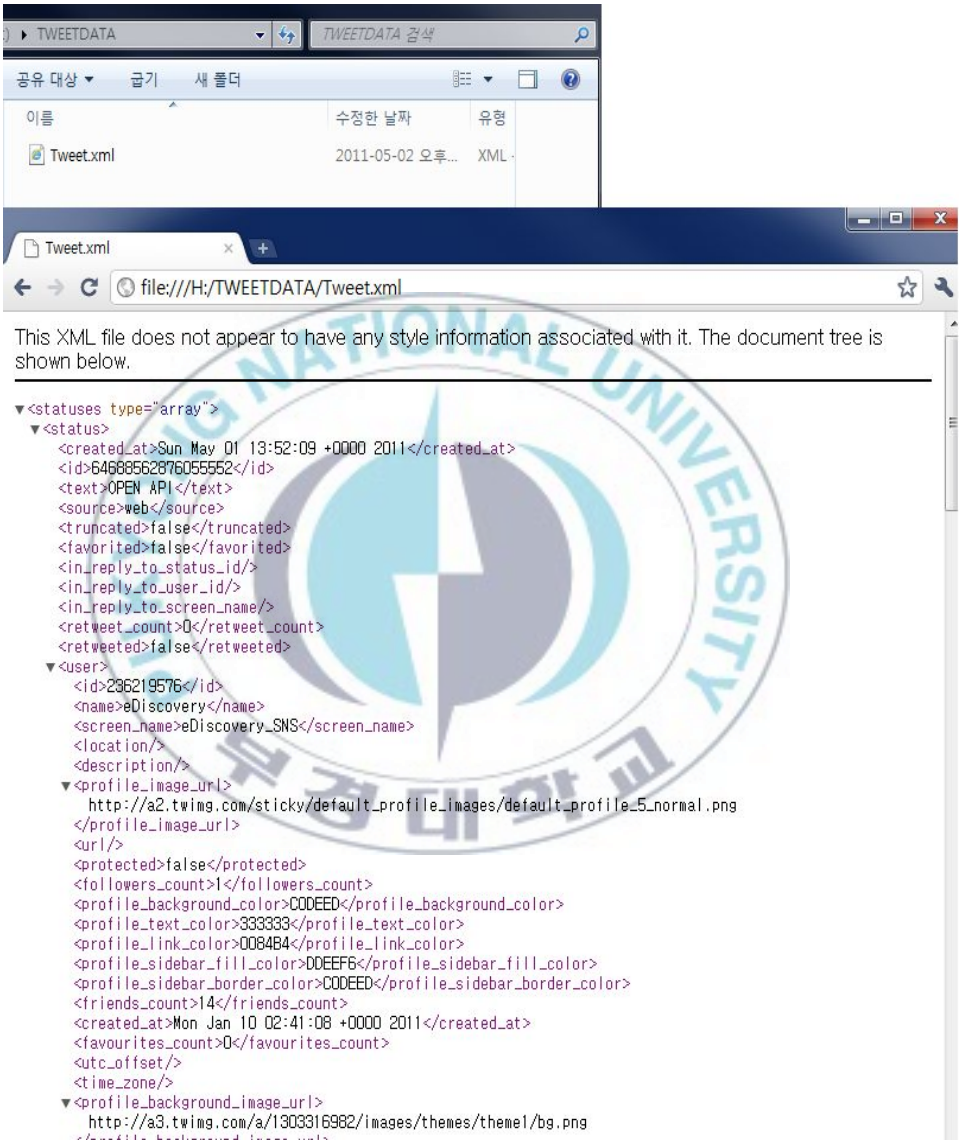


[그림 17] Litigation Hold 수행

다. SNS 계정 수집 기능

SNS 계정 수집 기능은 조직 내 SNS 계정을 수집해서 SNS 내용을 XML 형태로 저장하는 기능이다. 본 프로토타입에서는 Twitter의 데이터를 수집하는 기능을 구현해 보았다. Twitter는 프로텍터 된 계정이 아니면 별도의 로그인과정 없이 Twitter의 내용을 확인할 수 있다. SNS 서비스 특성상 다양한 클라이언트를 제공하므로 오픈 API를 제공한다. Twitter도 오픈 API를 제공함으로 이를 이용하여 tweet 데이터를 수집할 수 있다. Twitter의 OPEN API중 타임라인을 불러오는 API를 이용

하여 tweet 데이터를 불러와 XML로 저장한다. 그림[18]은 저장된 XML 데이터와 그 내용을 보여준다.



[그림 18] Tweet 데이터 수집 및 저장

4. 분석

앞서 개발한 프로토타입을 분석해보면 기존 솔루션보다 몇 가지 이점을 가지게 된다.

첫 번째로 문서 색인에 걸리는 시간을 줄일 수가 있다. 기존 솔루션들은 각 에이전트를 설치하고 매니저로부터 색인 정책을 수신받아 색인을 시작하게 된다. 하지만 본 프로토타입의 색인 기능은 Windows에 미리 색인되어 있는 정보를 이용함으로 색인 생성에 걸리는 시간을 단축할 수 있다. 또한, 새로운 ESI가 생성되더라도 바로 색인에 반영되므로 새롭게 색인을 할 필요가 없다. 단 이러한 색인 기능은 Windows 시스템에만 적용이 된다.

두 번째로 SNS수집 기능이다. 이 기능은 ESI보존 요구 사항 중 “ESI의 특성과 형태를 고려해서 수집” 항목을 충족하게 하려는 기능으로써 기존 e-Discovery 도구에서는 찾아보기 어려운 기능이다. 하지만 SNS의 사용량이 늘어나면서 SNS데이터가 증거로써 중요도가 높아지고 있으므로 SNS데이터를 수집하는 기능은 활용성이 높다. 또한, 수집된 SNS데이터를 XML 형태로 변환해서 저장하기 때문에 분석 및 저장하는데 편리하다.

세 번째로 PC에 Litigation Hold가 동작중에도 PC를 사용할 수 있다. 이 기능은 ESI보존 요구 사항 중 “조직 구성원들에게 ESI 보존의무를 공지”, “Litigation hold 정책 수립 및 시행”, “ESI 파괴와 변조를 막기를 위한 물리적 및 소프트웨어적 ESI 보호 조치 시행”을 충족하기 위한 기능이다. 본 프로토타입에서는 Litigation Hold가 시스템 전체에 내려지지 않고 개별 파일에만 적용되므로 Litigation Hold가 적용되어 있더라도 PC를 이용한 업무가 가능하다는 장점이 있다.

V. 결론

본 논문에서는 개정된 미국 연방 민사 소송법의 e-Discovery 제도와 e-Discovery에서의 핵심인 ESI 대해서 설명했다. 또한, e-Discovery 관련 프로젝트인 EDRM과 Sedona Conference에 대해서 설명하고 e-Discovery와 관련된 주요 판례들을 분석하였다.

EDRM 과 Sedona Conference의 ESI 수집과 보존에 관한 문서들을 분석하여 요구 사항을 도출하고 대표적인 스마트폰인 아이폰의 데이터를 수집하는 하는 방법과 Twitter와 Facebook의 데이터를 수집하는 방법에 대해서 연구하였다. 그리고 Agent, Manager, Server로 구성된 e-Discovery 지원 도구를 설계하면서 기존의 도구에서 지원하지 않는 SNS 데이터를 수집하는 기능을 포함하였다.

기존의 솔루션이 가지고 있는 색인 문제를 보완하기 위해서 Windows Search를 이용한 원격 색인 조회 기능과 파일 삭제와 변경을 막는 Litigation Hold 기능 및 tweet 데이터를 수집하는 SNS 계정 수집 기능을 가진 프로토타입을 구현하였다. 개선된 색인 기능과 Litigation Hold 기능의 구현으로 ESI를 보다 빠르게 조회 할 수 있으며 안전하게 보존할 수 있다. 또한, Twitter 자료의 수집기능으로 ESI 수집 범위를 넓혔다.

향후 e-Discovery 보존과 수집에 관한 요구 사항을 계속 도출하고 도출한 요구 상태를 토대로 e-Discovery 지원도구설계를 확장하고 ESI의 효율적인 분석과 검토를 위해 ESI의 Classification과 의미기반 검색, 연관성 검색 과 같은 분야를 지속적으로 연구 할 필요가 있다.

참고 문헌

[1] FRCP Committee Print, available, http://www.uscourts.gov/uscourts/RulesAndPolicies/rules/cv2008.pdf, 2008

[2] Twitter, http://www.twitter.com

[3] Facebook, http://www.facebook.com

[4] 한국 일보 2010년 4월 30일, http://news.hankooki.com/lpage/society/201004/h2010043022155722000.htm

[5] Electronic Discovery Reference Model, http://www.edrm.net

[6] Sedona Conference, http://www.thesedonaconference.org/

[7] 이창훈, 백승조, 김태완, 임종인, "E-Discovery를 위한 디지털 증거 전송시스템에 대한 연구", Korea Institute of Information Security and Cryptology, Vol. 18, No. 5, pp. 171-180, 2008

[8] Linda Volonino and Ian J. Redpath, e-Discovery For Dummies, Wiley Publishing, Inc., Indianapolis, Indiana, pp.18, 2010

[9] 김영수, 신상욱, 홍도원, “ESI 관점에서의 e-Discovery”, 정보통신산업진흥원 주간 기술동향,2010

[10] Linda Volonino and Ian J. Redpath, e-Discovery For Dummies, Wiley Publishing, Inc., Indianapolis, Indiana, pp.309, 2010

[11] Linda Volonino and Ian J. Redpath, e-Discovery For Dummies, Wiley Publishing, Inc., Indianapolis, Indiana, pp.310, 2010

[12] Linda Volonino and Ian J. Redpath, e-Discovery For Dummies, Wiley Publishing, Inc., Indianapolis, Indiana, pp.50, 2010

[13] AccessData eDiscovery, http://accessdata.com/

[14] EnCase eDiscovery, http://www.guidancesoftware.com

[15] Nuix, http://www.nuix.com/

[16] SourceOne e-Discovery, http://korea.emc.com/

[17] IBM Infosphere, http://www.ibm.com/software/data/infosphere/

[18] Clearwell, http://www.clearwellsystems.com/

[19] EDRM Farmework Guides, http://edrm.net/resources/guides/edrm-framework-guides/preservation

[20] Sedona Confernce WG1: Electronic Document Retention and Production, http://www.thesedonaconference.org/content/miscFiles/publications_html?grp=wgs110

[21] Sedona Conference WG1: COMMENTARY ON LEGAL HOLDS: THE TRIGGER & THE PROCESS, http://www.thesedonaconference.org/content/miscFiles/publications_html?grp=wgs110

[22] Apple, http://www.apple.co.kr

[23] 지식 경제부, http://www.mke.go.kr

[24] iTunes, http://itunes.com/

[25] iPhone Analyzer, http://sourceforge.net/projects/iPhoneanalyzer

[26] OAuth library functions, http://liboauth.sourceforge.net/

[27] 윈도 검색,
http://http://ko.wikipedia.org/wiki/%EC%9C%88%EB%8F%84_%EA%B2%80%EC%83%89

감사의 글

학부 3학년 때 처음으로 연구실 문을 두드렸을 때가 엊그제 같은데 벌써 4년 가까이 연구실에서 생활해 왔습니다. 연구실 생활을 하면서 너무나도 고마운 사람들을 많이 만났습니다. 덕분에 석사과정까지 무사히 마칠 수 있었던 것 같아 감사의 말을 전해드리려 합니다.

먼저 연구실에 들어오고 싶다고 불쑥 찾아온 저를 받아 주신 지도교수님 신상욱 교수님께 감사드립니다. 학부 시절부터 대학원 졸업까지 많이 모자라고 늘 실수투성인 저를 항상 자상하게 지도해주시고 배려해주신 덕분에 많은 것을 알게 되고 얻게 되었습니다. 앞으로 교수님의 가르침을 잊지 않고 훌륭한 인재가 되도록 하겠습니다. 다음으로 바쁘신 일정에도 불구하고 좀 더 좋은 논문 완성을 위해 조언을 아끼시지 않으셨던 이경현 교수님과 논문의 세세한 부분 까지 신경을 써서 논문을 지도해 주신 송하주 교수님께도 감사드립니다.

그리고 저를 가족처럼 대해준 우리 LACUC 멤버들, 언제나 저를 따뜻하게 대해주신 사모님, 큰 형 같은 태훈 선배, 소중한 인턴 경험을 만들어 주신 진흥선배, 친 누나처럼 챙겨 주셨던 수완 선배, 알고 보니 동아리 선배 였던 도희 선배, 센스쟁이 태식선배, 석사 생활 내내 저를 담금질 해주셨던 태림 선배 모두들 진심으로 감사드립니다. 그리고 학부 시절 모르는 게 있으면 언제나 친절하게 가르쳐 주신 종탁 선배와 경민 선배에게 감사드립니다. 잘생긴 준용이, 연구실 초 울트라 하이퍼 트리플 코어 중원이와 헌이, 연구실 유일한 준회원 혜진이, 듬직한 혁이, 연구실 막내 형주, 수빈이 모두 같이 연구실 생활에서 정말 즐거웠습니다.

한참 고생을 하고 있는 입학 동기 은혜와 중기, 그리고 나를 믿고 따라와서 고생하고 있는 본민이 모두모두 고맙습니다.

얼떨결에 들어간 N.A.N 에 저를 따뜻하게 받아주신 재현선배, 지훈선배, 실력으로 모범을 보여주신 기현선배, 꼭 성공하리라 의심치 않는 찬두선배, N.A.N 동아리 안방 마님 희경이 모두 감사 드립니다. 통영에서 부산으로 와 고생하고 있는 병훈이, 진수, 영어의 자신감을 불어넣게 해준 훌륭한 영어 선생님 소연이, 그리고 균덩어리 회장님 석균햄, 광호햄, 은영이 누나, 정두햄, 성우햄 모두 감사합니다.

뜬금 없는 방문에도 늘 반갑게 맞아 주신 작은 외삼촌, 외숙모, 언제나 좋은 이야기 많이 해주신 큰외삼촌, 늘 맛있는 음식 많이 해주시는 큰외숙모, 늘 인자하신 이모, 이모부 모두 감사합니다. 자주 찾아 뵙지 못아늘 죄송한 할아버지, 할머니 손주의 손주 보실 때까지 오래 오래 건강하세요. 이외에 제가 미처 언급하지 못한 고마운 분들께도 감사의 마을 전합니다.

까칠 하지만 누구보다도 오빠 생각해주는 동생과 대학원을 마칠때까지 누구보다도 든든한 지원군이 되어 주시고 늘 믿어 주신 부모님에게 논문을 바칩니다.

amor, amour, kærlighed, αγάπη, liefde, любовь , ভালবাসা , cinta, rakkaus, grá yygtmh