
기업의 효과적인 소송 대응을 위한 전자증거개시 절차 모델과 대체 기술

이태림*, 신상욱**

E-Discovery Process Model and Alternative Technologies for an Effective Litigation Response of the Company

Tae-Rim Lee*, Sang Uk Shin**

요약 본 논문에서는 미국의 전자증거개시(E-Discovery) 제도의 확산에 따라 국내 도입에 대비하여 표준화 된 업무 수행 절차 확립을 위해 선행 연구가 이루어지고 있는 EDRM(Electronic Discovery Reference Model) 및 The Sedona Conference 프로젝트에 대한 분석을 바탕으로 일반화 된 E-Discovery 프로세스와 세부 절차 별 필수 업무 사항들을 제시한다. 또한 이런 절차들이 실제 소송에 활용됨에 있어서 근본적으로 내포하고 있는 시간과 비용 문제를 해결하기 위한 대체 기술로써 기계 학습, 오픈 소스 형태의 정보 검색 라이브러리, Hadoop 기반의 대용량 데이터 분산 처리 기법 등을 소개하고, E-Discovery 프로세스 상에서의 활용 방안을 제시하여, 관련 서비스 및 솔루션을 개발하고자 하는 벤더들에게 유용한 정보를 제공한다. 이는 또한 제도적 변화에 발맞추어 업무 시스템을 재정비하고자 하는 기업들로 하여금 소송에 보다 효과적으로 대처할 수 있도록 한다.

주제어 : 소송 대응, 전자증거개시, E-Discovery, 증거개시 절차, 전자 정보 관리

Abstract In order to prepare for the introduction of the E-Discovery system from the United States and to cope with some causable changes of legal systems, we propose a general E-Discovery process and essential tasks of the each phase. The proposed process model is designed by the analysis of well-known projects such as EDRM, The Sedona Conference, which are advanced research for the standardization of E-Discovery task procedures and for the supply of guidelines to hands-on workers. In addition, Machine Learning Algorithms, Open-source libraries for the Information Retrieval and Distributed Processing technologies based on the Hadoop for big data are introduced and its application methods on the E-Discovery work scenario are proposed. All this information will be useful to vendors or people willing to develop the E-Discovery service solution. Also, it is very helpful to company owners willing to rebuild their business process and it enables people who are about to face a major lawsuit to handle a situation effectively.

Key Words : Litigation Response, Electronic Discovery, E-Discovery, E-Discovery Process, ESI Management

1. 서론

전자 문서(electronic documents)란 정보처리능력을 가진 장치에 의하여 전자적인 형태로 작성되어 송수신되거나 저장된 문서 형식의 자료로서 표준화된 것을 말한

다[3]. 최근 컴퓨터를 비롯한 다양한 디지털 장치의 보급이 원활해지고, 이와 연관된 기술들이 고도로 발달함에 따라 자동화된 정보 처리 시스템의 사용이 보편화 되면서, 업무를 비롯한 일상에서도 전자 문서의 사용이 급증하게 되었다. 특히 이러한 현상은 범죄 수사 및 사법 제

이 논문은 2009학년도 부경대학교 연구년 교수 지원사업에 의하여 연구되었음 (PS-2009-025)

*부경대학교 정보보호학(협) 박사과정

**부경대학교 IT융합응용공학과 부교수(교신저자)

논문접수: 2012년 7월 12일, 1차 수정을 거쳐, 심사완료: 2012년 8월 24일

도에도 영향을 끼쳐 민·형사 소송의 결정적인 증거가 되는 자료들이 각종 디지털 매체 내에 전자 문서의 형태로 보관되어 있는 경우가 늘어남에 따라, 이들을 확보하고 증거로 재판에 활용하기 위한 디지털 포렌식 분야에 대한 기술 개발이 활발히 진행되고 있다.

그 한 갈래로, 2006년 12월 1일 미국의 민사소송규칙 FRCP(Federal Rules of Civil Procedures) 개정으로 인해 등장한 전자증거개시(E-Discovery) 제도는 소송에 처한 당사자로 하여금 소송 쟁점과 관련된 모든 전자 정보를 수집하여 특정 기간 내에 제출하도록 강제하는 법안으로 증거로서 전자 기록 제출 및 공개에 대한 책임을 소송 당사자에게 부가하는 것이다. 해당 법안이 주로 적용되는 판례들을 살펴보면, 주로 기업 대 개인, 또는 기업 간 소송임을 고려해 볼 때, 업무를 위해 매일 많은 양의 정보를 처리하고 있는 기업들에게 모든 전자 정보를 보존하고 검색할 수 있는 시스템을 구축하여 사용해야 함을 의무화한 것으로 볼 수 있다[2]. 해당 제도가 아직 국내에서 법제화 되지는 않았으나, 최근 디지털 자료의 증거 채택과 개정된 형사소송법의 증거개시제도 도입, 디지털 증거의 중요성 인식 등 여러 상황들을 고려해 볼 때, 과도기적 단계이긴 하지만 디지털 포렌식 및 전자증거개시 제도와 관련하여 법 제도 개선과 기술 표준 마련 등 다양한 노력을 지속하고 있는 실정이다.

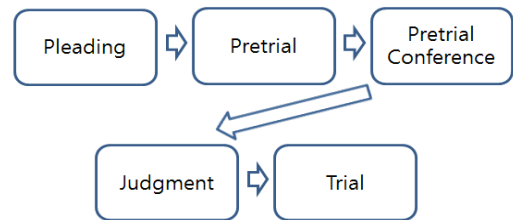
이러한 제도적인 보완이 신속히 이루어져야 하는 또 하나의 중요한 원인은 다양한 IT 컴플라이언스들의 등장이다. IT 컴플라이언스란 새로운 규제나 법안, 권고 등 각 나라 별 혹은 글로벌 감독 당국이 제시한 각종 요건을 만족시킬 수 있도록 기업의 정보 시스템과 업무 프로세스를 재정비하는 것을 의미한다[1]. 현재 미국을 포함한 해외 각 나라에는 SOX, HIPAA, GLBA, SBI386 등 해당 국가 및 기관을 대상으로 경제 활동을 펼치고자 하는 기업들이 준수해야 할 IT 컴플라이언스들이 존재하고 있으며, 이는 특히 최근 들어 세계를 무대로 활동하고 있는 글로벌 기업들이 증가함에 따라 더욱더 주목받고 있다. 또한 기업 간 이해관계를 둘러싼 분쟁과 민사 소송들이 급증하면서 기업들은 일반적인 조직 내 정보 관리를 위한 시스템들 이외에도 전자증거개시 제도에 활용 가능한 전문화된 솔루션들을 도입하여 자신들의 업무 프로세스를 개선하고자 시도하지만, 구체적인 법 규정이나 기술적 표준 등이 없는 상황에서 어떻게 대처해야 할지 막연한 입장이다.

이에 본 논문에서는 미국의 전자증거개시 제도와 이를 적용함에 있어서 표준화된 업무 수행 절차 확립을 위해 활발한 연구가 진행되고 있는 EDRM(Electronic Discovery Reference Model)[9] 및 The Sedona Conference[17] 프로젝트 등에 대한 분석을 바탕으로 일반화된 E-Discovery 프로세스와 절차 별 필수 업무 사항 등을 제시한다. 또한 이러한 절차들이 실제 소송에 활용됨에 있어서 전자증거개시 제도가 근본적으로 내포하고 있는 시간과 비용 문제를 해결하기 위해, E-Discovery 솔루션 구현 및 기업의 시스템 보완에 활용 가능한 대체 기술들을 제시하여, 기업들로 하여금 소송에 보다 효과적으로 대처할 수 있도록 한다.

2. 연구 배경

2.1 전자증거개시(E-Discovery) 제도

증거개시(Discovery) 제도란 소송당사자가 공판 전에 공판의 준비를 위해 법정 외에서 법이 정한 방법에 의하여 소송의 쟁점을 명확히 하는 정보 및 증거를 수집·공개하는 제도를 말한다. 전자증거개시 제도는 기존의 증거개시 제도에서 증거의 범위가 디지털 형태의 전자 문서로 확장되고 있는 현대의 흐름을 반영하여, 이를 ESI(Electronically Stored Information)라 총칭하고 증거 공개 대상으로 포함시킨 것이다[2]. 다음 [그림 1]은 미국 FRCP 상의 민사소송 절차를 간략하게 도식화한 것이다 [16].



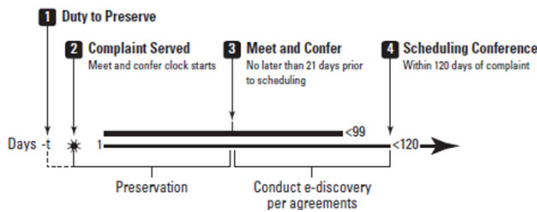
[그림 1] FRCP 상의 민사 소송 절차

원고의 소장 제출에 대한 피고의 답변서 제출 단계인 변론(Pleading)을 거치고 나면, 소송 당사자들이 상대방에서 증거의 공개를 요구하는 절차인 사전 심리(Pretrial) 단계로 들어가게 되는데, 이 단계의 핵심이 곧 전자증거개시 제도이다. 이를 통해 소송 당사자가 서로 상대방이

보유한 증거물을 공개하도록 요청하고, 사전에 확인 가능하게 하여 대등한 조건에서 소송이 진행되게 하며, 유불리를 미리 알게 하여 공판 이전에 합의의 끝내내 소송이 효율적으로 처리되게 한다.

2.2 전자증거개시(E-Discovery) 절차

다음 [그림 2]는 일반적인 소송 절차에 있어서 전자증거개시와 관련하여 법에 명시된 주요 테드라인들을 나타내고 있는 그림이다[14].



[그림 2] 전자증거개시 절차와 테드라인

- Day -t : 소송 발생 이전 기업 내 정보들에 대해 보존의 의무를 수행하는 단계
- Day 1 : 소장 접수로 인한 소송 발생
- By Day 99 : Meet-and-Confer 세션으로 소송 당사자 간 협의를 통해 공개되어야 할 증거 범위와 형식, 시간 및 비용 등을 결정하는 단계
- By Day 120 : 소송에 대한 재판 스케줄을 결정하는 단계

2.3 전자증거개시에 따른 업무 수행 시 핵심 요소

위 절차에 따르면, 실제 소송 관련 정보를 수집하고, 증거로 생산하여 제출하는데 소요할 수 있는 기간은 소송 발생 일로부터 120일 이내이다. 이는 충분한 시간처럼 여겨질 수 있으나, 여러 테드라인들을 통해 알 수 있듯이 소송 당사자 간의 주요 쟁점과 증거 제출 양식 협의, 소송 일정을 확인하는 등 많은 업무들이 포함되어 있으므로, 실제 주요 전자 정보 검색 및 증거로서 적합성 검토 등에 투자할 수 있는 시간은 길지 않음을 알 수 있다. 법이 정한 기한 내에 증거를 마련하지 못한다면 패소는 자명한 일이므로 업무 수행에 있어서 시간은 아주 중요한 요소임을 의미한다. 또한 소송은 언제든 발생 가능하며 기업이 다루는 정보의 양이 나날이 대규모화 되고 있

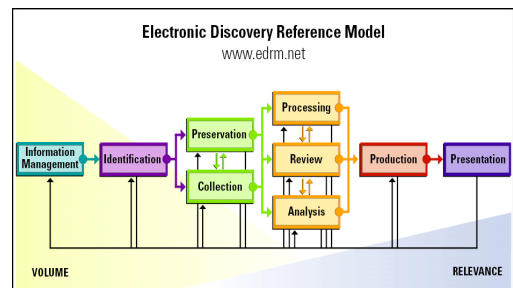
는 상황 등을 고려해보면, 소송과 상관없이 평상시에도 업무와 연관된 정보들을 체계적으로 분류하고 관리하여, 정보 검색에 소요되는 시간을 줄임으로써 신속한 대응을 가능하게 해야 한다.

더욱이 가트너(Gartner)의 조사 결과에 따르면 법률 전문가를 통한 1GB의 증거 자료 검토에 소모되는 비용이 \$18,750 정도로 대부분의 경우 관련 업무 수행에 있어서 가장 많은 비용 지출을 차지하고 있다[7]. 이는 검토가 필요한 자료의 양이 너무 많아 과도한 비용이 지출되어 오히려 소송의 패소로 인해 입게 되는 기업의 경제적인 손실보다 더 큰 손해가 발생하게 된다면, 전자증거개시 관련 업무 수행 자체가 무의미함을 의미하는 것이다.

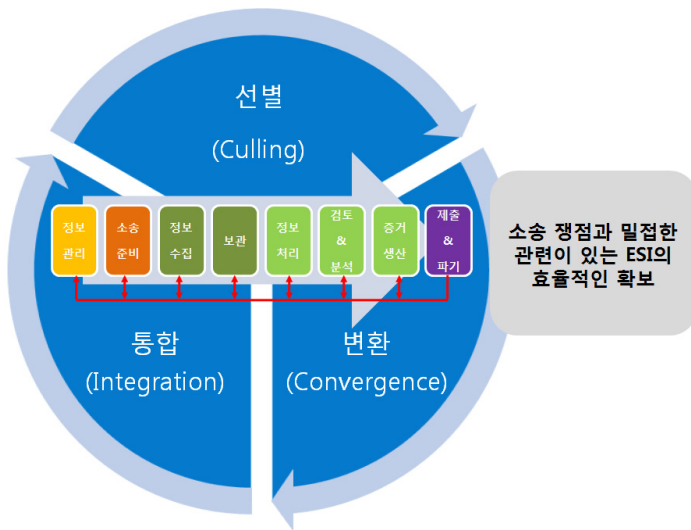
요약하자면, 기업이 제도의 변화에 부응하기 위해 전자증거개시를 위한 전문적인 솔루션 도입으로 업무 절차 재정비를 꾀하고자 할 때는, 시간과 비용 측면에 있어서 가장 효과적인 방법을 모색해야 한다.

2.4 EDRM과 The Sedona Conference

2005년 5월에 만들어진 EDRM(Electronic Discovery Reference Model)은 미국의 FRCP에서 명시하고 있는 요구 사항들을 효과적으로 준수하기 위해 전자증거개시 절차를 표준화 하고 절차 별 업무 수행을 위한 가이드라인을 제공하고 있으며, 60개 이상의 관련 기관들의 협의를 거쳐 참조 모델 개선을 위한 연구가 진행되고 있다[9]. 현재 많은 상용화 제품들이 이를 기반으로 개발되었거나 진행 중에 있는데, 해당 모델의 주요 목적이 관련 벤더들로 하여금 E-Discovery 제품과 서비스 개발에 활용 가능한 프레임워크를 제공하는데 치우치는 경향이 있기 때문이다. 세부 절차는 아래 그림에서 나타난 것처럼 총 9개의 단계로 구분되어 지며, 각 절차 별 상세들은 전문화된 솔루션이 수행하는 기능으로 봐도 무방하다.



[그림 3] Electronic Discovery Reference Model



[그림 4] E-Discovery 업무 절차 모델

1997년 Richard G. Braman에 의해서 설립된 The Sedona Conference는 반독점 법, 특허와 저작권, 복잡한 소송의 쟁점들을 다루는 컨퍼런스이며, 7개의 워킹 그룹으로 나누어져 있다. 각 워킹 그룹들은 전자 정보들에 관한 다양한 이슈들을 다루며, 변호사 및 법률 전문가, 자문 위원 등으로 구성되어 있다. 이는 주요 목적이 소송에 대한 효과적인 대응 절차를 마련하는 것이므로, EDRM에 비해 보다 포괄적이며 법적인 관점에서 전자증거개시 제도에 대한 가이드라인 및 원칙들을 제정하며, 솔루션이나 도구적 관점에서 가치 있는 증거 생산을 위한 기술적인 요구조건은 물론 소송 당사자의 입장에서 준수해야 할 의무, 권고사항 등을 기술한다[17].

3. 전자증거개시 제도의 효과적인 대응을 위한 업무 수행 절차 모델

본 절에서는 앞서 소개한 2가지 프로젝트에 대한 분석 내용을 기반으로, 각 자료들 간에 존재하고 있는 전자증거개시 제도에 대한 미묘한 관점의 차이들을 접목하여 보다 일반화된 E-Discovery 업무 수행 절차를 제시하고, 절차 별 필수적인 업무 사항들을 나열한다. 다음 [그림 4]는 새로운 형태의 E-Discovery 업무 절차 모델을 나타낸다.

3.1 E-Discovery 업무 수행 절차 모델 정립 방법과 의미

- 궁극적인 목적 : 원 내부의 세부적인 절차가 지향하는 E-Discovery 업무의 최종 목표로서, 우측 사각형에 배치하였다. 1차적인 목표는 소송에서의 승소를 위해 쟁점과 관련성이 높은 ESI들을 빠짐없이 확보하는 것이며, 2차적인 목표로써 과도한 시간과 비용의 소모를 방지하는 것이 중요한 요소임으로 ‘효율적인 확보’라는 표현을 사용하였다.

- 화살표로 둘러싼 원 : E-Discovery 업무 중 핵심이 되는 3가지 요소를 나타내며, 세부 절차가 이루어지는 모든 과정들의 기반이자 특정한 순서 없이 전반적인 모든 절차에 걸쳐 지속적으로 수행되는 사항이므로 순환 형 화살표를 활용하였다. 다음 <표 1>은 3가지 요소에 대한 설명을 요약한 것이다.

- 원 내부의 연속 블록 프로세스 다이어그램 : 증거 생산을 위해 E-Discovery 관점에서 수행되어야 할 일련의 절차들을 연속적으로 나열한 것이며, 아래쪽의 화살표들은 세부 절차 간 업무 회귀 경로를 나타내기 위한 것이다. 이는 각 절차마다 업무 수행 결과에 대한 평가를 통해 현재 진행 중인 작업의 적합성을 지속적으로 진단하여, 잘못된 작업 진행에 대한 시간적, 비용적 피해를 최소화하기 위한 것이다. 다음 <표 2>는 다이어그램의 중심을 구성하고 있는 연속된 블록 내의 절차 별 필수적인 업무 내용들을 요약한 것이다.

〈표 1〉 E-Discovery 업무의 핵심 3요소

선별 (Culling)	<ul style="list-style-type: none"> · 기업이 다루는 정보의 양은 방대하지만, 이들 중 소송에 직접적으로 연관되어 증거로써 가치 있는 정보의 양은 일부분임 · 기업의 전체 데이터베이스로부터 검색을 통해 연관성 정보를 수집하는 작업에서부터 비용 절감을 위해 적합성 검토 대상을 연관성 정도에 따라 추려내는 작업, 검토 결과에 따라 증거로 생산할 ESI들을 결정하는 작업 모두 이에 해당한다고 볼 수 있음
통합 (Integration)	<ul style="list-style-type: none"> · 전자 문서의 형태들은 다양하며, 어떠한 절차에서도 이들을 누락 없이 다룰 수 있어야 할 필요성이 존재함 · 특히 소송 쟁점과의 연관성과 증거로써의 가치는 전자 문서의 형태보다 내용에 따라 달라지는 것이기에, 이들 간 상호 검토 및 중요도 비교를 위해서는 정보 통합의 개념이 필요함
변환 (Convergence)	<ul style="list-style-type: none"> · 절차에 따라 효과적인 업무 수행을 위해서는 각 단계 별로 명확한 결과물을 정의하고, 이를 다음 단계의 작업 대상으로 활용하는 것이 중요하므로, 정보에 대한 형태 변환 작업이 필수적임 · 또한 소송 당사자 간 협의를 통해 결정된 제출 증거물의 전자 문서 포맷에 맞게 확보한 정보들을 변환할 수 있어야 함

〈표 2〉 절차 별 주요 목표에 따른 업무 요약

정보 관리	<ul style="list-style-type: none"> · 소송이 발생하기 이전의 단계로, 기업이 운영되고 있는 평상 시 사내 정보들에 대하여 자체적인 관리 정책을 적용하고, 유지하는 작업이 이루어짐
소송 준비	<ul style="list-style-type: none"> · 소송이 발생하게 되면, 법률 전문가와 함께 소장의 내용을 검토하고 주요 쟁점을 파악하여, 향후 대응 전략을 수립하는 작업이 이루어짐 · E-Discovery 작업이 요청될 경우 소송 당사자 간 협의를 통해 확보해야 할 정보들에 대한 식별 작업을 수행
정보 수집	<ul style="list-style-type: none"> · 관리되고 있는 기업 내 데이터들에 대하여 식별된 소송 쟁점에 따라 검색 작업을 수행하여, 관련 정보를 확보
보관	<ul style="list-style-type: none"> · 확보한 정보들의 원본성을 보장하기 위해 무결성 검증과 향후 파기 및 분실에 대한 안전성을 확보하기 위한 작업을 수행
정보 처리	<ul style="list-style-type: none"> · 수집한 정보들을 소송 쟁점과의 적합성 검토를 위해 변환하거나 선별하는 작업을 수행
검토 및 분석	<ul style="list-style-type: none"> · 기업 고유의 기밀 정보나 특권 등과 관련된 정보들을 가려냄 · 소송 쟁점과의 적합성에 대한 검토 · 검토 결과에 대한 업무 평가 분석
증거 생산	<ul style="list-style-type: none"> · 협의된 증거 생산 포맷에 맞게 정보를 가공하는 작업을 수행
제출 및 파기	<ul style="list-style-type: none"> · 법이 정한 방식에 따라 증거를 공개하는 단계, 사용하지 않은 정보나 사용이 끝난 정보들을 파기

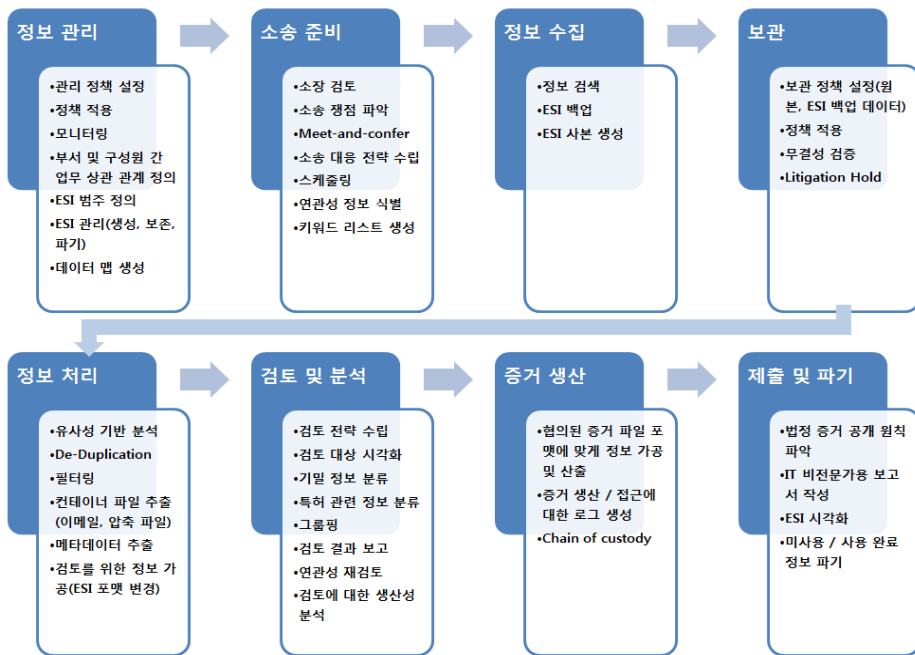
3.2 기존 E-Discovery 모델과의 차별성

현재 E-Discovery 분야의 수많은 전문가 및 연구 단체, 관련 제품 및 서비스 생산을 주도하고 있는 전문 업체들은 E-Discovery 전반에 걸친 모든 필수적인 요소들을 확일화 된 표준으로 마련하는 것이 불가능하다는 것에 동의하고 있다. 이는 소송의 내용에 따라 요구되는 법적 대응 절차들이 너무나 다양하며, 해당 분야에 활용 가능한 관련 기술들이 매우 빠르게 변화하며, 발전하고 있기 때문이다. 이에 E-Discovery 관련 시장에서 각기 다른 역할을 담당하고 있는 주요 구성원 및 단체들은 자신의 이익에 맞게 경쟁적으로 차별화 된 표준화를 진행하고 있으며, 해당 견해에 따른 절차 모델들을 제시하고 있는 상황이다.

이러한 관점에서 볼 때, 앞서 [그림 3]을 통해 소개된 EDRM은 E-Discovery 업무 수행을 지원하는 서비스 솔루션의 필수 기능 및 절차들을 표준화하는 성향이 강하기에, 관련 기술 및 제품을 개발하여 공급하는 기업들이 선호하는 모델이라 할 수 있다. 반면 이러한 제품들의 소비자가 되는 일반 기업 또는 Law Firm, 컨설팅 업체 등의 법률 전문가 단체들은 사람의 입장에서 조직 내 업무 처리 절차들을 유지하며, 보다 소송에 효과적으로 대처할 수 있는 표준화 된 방안을 마련하는 것이 더욱 중요하다고 여기고 있으며, 지원 솔루션들의 역할은 부수적인 것으로 생각하고 있다. 이런 성향이 강하게 반영되어 나타나는 성과가 The Sedona Conference라 할 수 있다.

하지만 E-Discovery 관련 표준화의 선두 주자로 여겨지고 있는 두 단체의 절차 모델에 있어서 공통적으로 결여되어 있는 요소는 E-Discovery 업무 수행의 궁극적인 목표와 업무 효율성 향상을 위해 반드시 선행되어야 하는 핵심 요소들이다. 특히 통상적인 소송 대응 업무 절차 이외에, 전체 E-Discovery 수행에 대한 효율성 평가에 있어서 아주 밀접한 연관성을 가지고 있으며, 실제 소송에 처한 당사자 입장에서 가장 중요하게 생각하는 요소인 시간과 비용 문제를 고려한 주요 사항들이 모델에 반영되어야 한다. 이는 업무를 수행하는 주체가 사람이냐 자동화 된 솔루션이냐 혹은 표준화에 반영되는 성향이 개발자 위주냐 소비자 위주냐를 넘어서, 최근 E-Discovery 관련 법적 요구사항의 변화 추세와 기술 연구 동향을 반영한 최적화 모델이라 할 수 있다.

이에 본 논문의 절차 모델에서는 점차 다양하게 나타나는 데이터 포맷들에 대해 효과적인 처리가 가능하게



[그림 5] E-Discovery 업무 절차 별 필수적인 기능 요소

하는 정보 통합 기술, E-Discovery 업무 수행에 있어서 가장 많은 비용 소모를 초래하고 있는 소송 쟁점과의 적합성 검토 작업의 효율 향상을 위한 정보 선별 기술, 급변하는 법적 요구사항들에 능동적으로 대처할 수 있게 하는 정보 변환 기술을 전체 E-Discovery 절차의 기본 요소로 배치하고, 각 세부 절차에서는 사람과 자동화 된 솔루션의 역할을 함께 포함할 수 있는 포괄적인 용어들을 사용하였다. 다음 <표 3>은 각 기관 별 모델과 본 논문에서 제시한 모델의 차이점을 요약한 것이다.

<표 3> E-Discovery 업무 모델 간의 차이점

	EDRM	Sedona	제안 모델
모델	전 과정 도식화한 모델 존재	업무 과정 별 부분적 도식화	전 과정 도식화한 모델 존재
주요 대상	E-Discovery 서비스 및 솔루션 개발자, 벤더	E-Discovery 업무 수행 인력, 소송 관련자	E-Discovery 관련자 모두 포함
관점	E-Discovery 업무는 도구 의존도가 높음, 도구의 역할이 중요	업무 수행 인력의 역할이 중요	관련자 모두에게 유용한 정보를 제공하는 모델
성향	우수한 성능의 지원 도구 개발을 위한 가이드라인	효과적인 업무 수행을 위한 가이드라인	공급적인 목표와 업무 효율성 향상을 위한 핵심 요소 제시

3.3 절차 별 필수적인 업무에 따른 기능 요소

<표 2>에서 제시한 E-Discovery 업무 절차 별 수행 내용에 따라 각각의 목표를 달성하기 위해 필수적으로 적용되어야 하는 기능적 요소들은 [그림 5]와 같다.

· 소송 발생 이전(Day -t) : 기업이 정상적으로 운영되고 있는 평상시의 정보 관리 단계에 해당하는 시기로 자사에 적합한 관리 정책 설정이 우선되어야 하며, 이는 구체적인 정보 관리 방법을 구현해야 하는 부서 및 해당 분야 전문가와 실제 기업의 업무를 담당하고 있는 각 부서 직원들 간의 상호 작용을 통해 수행되어야 한다. 이를 통해 부서 및 구성원 별 역할과 업무 상관관계를 재정립하고, 정보가 현재 업무와 관련된 가치 이외에 잠재적인 가치 또한 평가하여 관리할 수 있도록 기업 내 다양한 정보들에 대한 데이터 맵 생성이 필요하다. 설정된 정책이 적용되면, 준수 여부에 대한 지속적인 모니터링이 이루어져야 하며, 효율적인 관리와 향후 신속한 열람이 가능하도록 ESI들의 범주 정의가 수반되어야 한다. 주요 관리 업무는 Information Lifecycle에 따라 생성된 ESI들을 정책에 지정된 기간 동안 보존하고, 파기하며 해당 작업들에 대한 기록을 유지하는 것이다[4].

· 소송 발생(Day 1 ~ 99) : 소송 준비 단계로 변호인과 같은 법률 전문가의 도움을 받아 소장에 대한 내용 검토

를 통해 쟁점 파악이 선행되어야 한다. 그 후, Meet-and-Confer 세션을 진행하게 되는데, 이는 소장 접수 후 99일 이내에 이루어져야 하며 소송 상대방과의 협의를 통해 쟁점을 명확히 하고 제출해야 하는 증거 범위, 포맷 등을 결정한다. 해당 기간에는 소송 대응 전략 및 E-Discovery 수행 계획 수립이 동반되며, 협의된 내용에 따라 이는 언제든지 변경 가능해야 한다. 또한 주요 쟁점에 따라 확보해야 하는 연관성 정보들을 식별하고, 정보 수집을 위한 ESI 검색에 활용할 수 있도록 키워드 리스트를 생성해야 한다.

· E-Discovery 업무 수행(Day 1 ~ 120) : 정보 수집 단계에서는 식별된 소송 쟁점과 생성한 키워드 리스트에 따라 기업 내 정보들에 대해 검색 작업 수행하고, 확보한 ESI들을 백업하며 검토 및 분석 단계에 활용할 수 있도록 사본을 생성해야 한다. 이는 전자 문서가 변경이 용이하다는 특징이 있기 때문에, 의도하지 않은 정보 변경이나 파기를 방지하기 위함이며, 증거로써 가치를 충분히 인정받기 위한 보관 절차와도 관련이 있다. 보관 단계에서는 해쉬 기법 등을 이용한 무결성 검증 작업이 수행되며, 특히 Litigation Hold를 위한 절차들이 진행된다. Litigation Hold란 조식이 향후 소송이 예측되거나 실제로 발생했을 때, 소송과 연관된 모든 형태의 정보들을 보존해야 하는 당사자의 증거 보존 의무로, 이를 위한 구체적인 방법들은 차이가 있을 수 있지만 대부분 우연 또는 고의적으로 발생 가능한 정보 파괴, 수정에 대한 조직 내 외부의 위협으로부터 정보들을 적극적으로 보호하는 조치들이 전부 포함될 수 있다[11]. 보관 단계를 통해 안전성이 확보된 이후, 정보 처리 단계에서는 검토 및 분석 작업을 위해 정보들을 가공하거나 추출하는 작업들과 서두에 언급한 것처럼 비용 지출이 막대한 검토 작업의 대상을 줄이기 위해 중복된 ESI들을 제거하고, 우선 검토되어야 하는 대상들을 분류하는 작업들이 주를 이룬다. 이를 위해 유사성 기반 데이터 분석, 데이터 중복 제거(De-duplication), 순위화 된 검색 결과에 따른 필터링 기법들이 활용되며, 데이터 특징에 따라 통합된 형태의 검토 대상 포맷 생성을 위해 컨테이너 파일 추출과 메타데이터 추출 작업 등이 필수적이다[14]. 검토 및 분석 단계에서는 다양한 ESI들에 대해 동시적인 연관성 분석이 가능하도록 하는 시각화 기술이 적용되어야 하며, 공개되어서는 안 될 기업 고유의 기밀 정보나 특히 관련 정보 등을 분류해 내는 것이 주요 임무이다. 또한 소송 쟁점과

의 연관성 검토를 통해 최소의 비용 수준에서 최적의 증거물을 생산하여 제출할 수 있도록 적합성을 기준으로 한 ESI 그룹핑과 순위화 된 검토 결과 생성이 이루어져야 한다. 이러한 기준들은 검토 작업이 수행되기 이전 전략 및 계획 수립을 통해 사전 결정되어야 하며, 최종 증거 생산 이전 검토 결과에 대한 생산성 분석을 통해 효율을 평가하고 원하는 수준에 미치지 못할 경우 잘못된 부분을 파악하여 신속한 재작업이 이루어질 수 있도록 한다. 마지막 증거 생산 단계에서는 Meet-and-Confer 세션 중 협의된 증거 파일 포맷에 맞게 정보를 가공하여 증거로 생산해내고, 해당 작업 및 증거 자료 접근에 대한 로그를 기록해야 한다. 이는 생산한 결과물에 대해 증거로써 가치를 보다 인정받기 위한 증거물 연계 보관(Chain of custody) 원칙에 따른 의무사항이다.

· 증거 공개(Day 120 이후) : 제출 및 파기 단계로 법정 증거 공개 원칙 파악하고, 공개 임무를 주로 수행하는 법률 전문가들이 IT 비전문가임을 고려하여 그들의 쉬운 이해를 돕기 위해 보고서 작성 및 ESI 시각화 작업이 수행된다. 또한 증거로 생산되었지만 사용하지 않았거나, 사용이 완료된 정보들을 파기하고 원본 데이터들에 대해서는 소송 관련 기록과 함께 정보 관리 정책에 따라 이를 보존해야 한다.

4. E-Discovery 업무의 시간 및 비용 절감을 위한 대체 기술 분석

본 절에서는 전자증거개시 제도가 근본적으로 내포하고 있는 시간과 비용 문제를 해결하여 기업들로 하여금 소송에 보다 효과적으로 대처할 수 있도록 하기 위해, 기존에 E-Discovery 솔루션 구현 및 기업의 시스템 보원에 사용되던 기술들을 대체하여 새롭게 활용 가능한 기술들을 제시한다.

4.1 E-Discovery의 과거와 현재

과거의 E-Discovery 서비스 및 솔루션의 주요 목적은 사람에게 의한, 즉 법률 전문가에 의한 소송 쟁점과의 적합성 검토 작업 준비 단계로써, 키워드에 근거하여 문서를 식별하고 수집하는 역할만을 수행하였으며, 단순한 정보 검색 분야 기술에 대한 의존도가 높았다. 기업이 처리하는 정보의 양이 급증하고 있음을 고려해 볼 때, 사용자의

정보 요구에 부합하는 데이터를 신속하게 검색하는 기법의 사용은 필수불가결한 요소일 수밖에 없지만, 최근 연구 동향들을 살펴보면 키워드를 기반으로 한 단순한 검색 기법 활용은 그 한계가 분명하며, 특히 검색 용어를 선정함에 있어서 법률 전문가로 하여금 큰 부담을 안겨주는 결과를 초래하고 있다. 정보 검색 분야에서 기법에 대한 성능 평가 시 사용되는 2가지 요소인 재현율(recall)과 정확율(precision)을 고려해 볼 때, 이는 재현율이 아주 낮은 검색 결과를 만들어 낼 수 있기 때문이다. 일부 전문가들의 견해에 따르면, ‘증거물 제출에 있어서 관련성이 떨어지는 증거물을 많이 제출하는 경우보다 증거물 자체가 부족한 경우 소송에 있어서 더욱 불리하게 작용할 수 있다’라고 주장하는데, 이는 다시 말해 재현율의 중요성을 강조하는 의미라 할 수 있다[13]. 이에 ‘Advanced Search’라 불리는 동의어, 시소러스 등을 활용한 쿼리(query) 확장 기법들이 주목을 받았으나, 이는 검색 결과물을 증가시키게 됨으로 검토 업무 관점에서 볼 때, 그 시간과 비용을 엄청나게 증가시키는 좋지 못한 결과를 초래할 수 있다. 결국 재현율과 정확율에 대한 트레이드오프에 대해서는 그 적정선을 결정하는 것이 쉽지 않은 문제이긴 하지만, 상황에 따라 이를 효과적으로 조율할 수 있는 기법이 필요한 것은 분명한 사실이다.

최근 TREC의 Legal Track이나 E-Discovery 관련 컨퍼런스에서 주목받고 있는 방법은 자동화 된 기계 학습(machine learning)을 활용한 텍스트 마이닝(text mining) 기반의 적합성 분류(relevance classification) 기법이다 [13][18]. 또한 Hadoop과 같은 Map-Reduce 기반의 분산 처리를 이용한 인덱스 기법 및 문서 처리 기술 등을 활용하여 속도적인 측면에서 검색 작업의 성능을 높이고 있으며, 인덱스 정보의 확장을 이용한 문서 태깅(tagging) 기술을 적용하여 보다 손쉽게 사용자의 정보 요구를 만족시킬 수 있게 하는 결과 도출 방법들이 각광을 받고 있다.

4.2 기계 학습(Machine Learning)

기계 학습은 인공 지능의 한 분야로, 컴퓨터가 학습할 수 있도록 하는 알고리즘과 기술을 개발하는 분야를 말한다[3]. 정보 검색 분야에서는 전자 문서 관리를 위한 문서 분류 기술 연구가 활발히 진행되고 있는데, 특히 웹과 같은 대규모 검색 시스템에서는 새롭게 추가되는 웹 페이지들을 자동으로 수집·분류하여 사용자 질의에 대한

검색 대상 범위를 축소함으로써 성능 향상을 꾀하고, 검색 결과에 대한 사용자 만족도를 높이기 위해 분야 별로 분류된 결과를 제공하는 등의 다양한 기술들이 개발되어 사용되고 있다. 이러한 기능들은 E-Discovery의 절차를 진행하는 과정에 있어서 다음과 같이 다양한 부분에 활용될 수 있다[6].

먼저 정보 관리 단계에서는 사전에 정의한 ESI 범주에 따라 학습 자료를 생성하고, 자동화 된 분류를 수행 가능하게 하여 관리 작업에 대한 효율을 높일 수 있다. 이는 정보 수집 단계에서 검색 기법을 구현할 때에 추가적으로 활용 가능하며, 앞서 언급한 정보 검색 분야에서의 활용 사례가 그대로 적용되어 검색 성능 향상을 꾀할 수 있게 된다. 특히 주목해야 할 부분은 검토 절차에 대한 활용도이다. E-Discovery에서 말하는 문서 분류는 주로 소송 쟁점과의 적합성을 기준으로 유효한지 아닌지(responsive or not)에 대한 2진 분류를 의미한다. 단순한 작업처럼 보일 수 있지만, 많은 실험들을 살펴보면 사람, 즉 법률 전문가에 의한 문서 검토는 시간과 비용 측면에서 치명적인 단점들이 존재하며, 적합성 여부에 대한 평가 결과가 일정하지 않게 나타난다는 보고들이 있다[5]. 어느 정도의 소규모 ESI들에 대해서만 사람이 검토하여 적합과 부적합 문서에 대한 학습 자료를 생성하고, 나머지 ESI들에 대해서는 기계 학습 기법들을 적용한다면, 이러한 문제점들을 효과적으로 해결할 수 있을 것이다. 또한 기계 학습 기법에서 분류를 위해 사용되는 문서 간 유사도(likelihood) 수치 정보[10]를 적절한 기준 설정을 통해 증거 생산 대상으로서 ESI 선별 시 활용한다면, 상황에 맞게 재현율과 정확율의 트레이드오프를 조절할 수 있다. 예를 들어, 최종적으로 증거를 생산하기 위해 사람에게 의한 재검토 작업을 수행할 때, 시간이 충분하지 못하거나 비용이 부담스러운 경우 유사도 수치 기준을 높여서 정확률이 높은 문서만을 선별함으로써 검토 시간 및 비용 절감을 꾀할 수 있게 되고, 반대의 경우라면 재현율을 높여 많은 ESI들을 심도 깊게 검토함으로써 결정적인 증거 확보를 위해 재판에 대한 준비를 보다 철저하게 진행할 수 있게 된다. 다음 <표 4>은 3장에서 제시한 절차 모델 상에서 기계 학습 기술이 적용될 수 있는 각 단계별 기능 요소와 기대 효과를 간략하게 설명한 것이다.

〈표 4〉 기계 학습 적용 대상 기능 요소 및 기대 효과

단계	적용 대상 기능	기대 효과
정보 관리	- ESI 관리 - 데이터 맵 생성	· 기계에 의한 자동화 된 ESI 분류는 사람에게 의한 작업 의존도를 낮추어 인력 및 비용을 절감할 수 있음 · 분류 업무 처리 속도가 빨라짐에 따라 전반적인 ESI 관리 작업의 효율이 증가함
정보 수집	- 정보 검색	· 소송 이슈에 따라 선정된 검색 키워드와 ESI 분류 범주 간의 유사성 비교를 통해 검색 대상 범위를 제한함으로써 작업 속도를 빠르게 함
검토 및 분석	- 기밀 정보 분류 - 특허 분류 - 그룹핑 - 연관성 제검토	· 기밀 또는 특허 관련 ESI를 분류하기 위한 범주를 정의하고, 정보 관리 단계에서 미리 분류함으로써, 해당 범주의 ESI들이 검색되었을 때 별도의 검토 작업 없이 제외시킬 수 있게 됨 · 소송 이슈와의 유사도에 따라 ESI들을 그룹핑 하고, 증거 생산 시 비용적 한계로 인해 우선 생산 되어야 하는 ESI들을 선별하고자 할 때, 유사도를 연관성 정보로 활용 가능함 · 해당 효과들로 인해 검토 및 분석 작업에 소요되는 시간과 비용을 절감할 수 있음

4.3 고성능 검색 엔진

최근 정보 검색 기술 개발이 활발히 진행됨에 따라 고성능의 인덱싱 및 검색 기능을 제공하는 오픈 소스 형태의 라이브러리들이 다양하게 등장하고 있다. 이러한 기법들은 대규모 ESI에 검색 작업이 주를 이루는 E-Discovery 업무를 수행함에 있어서 아주 유용하게 활용될 수 있다. 정보 관리 단계에서는 데이터 맵 생성 과정에서 발생하는 다양한 부가 정보들을 이용하여, 인덱스 정보 확장을 통한 문서 태깅 기법들을 적용한다면 보다 효과적인 관리 업무 수행이 가능할 것이다. 또한 정보 수집 단계에서 이러한 태깅 정보를 활용한 쿼리 확장 기법이 4.1절에서 등장한 advanced search 기술들과 함께 구현된다면, 비약적인 검색 성능 향상을 가져올 수 있다. 특히 가중치를 이용한 질의 수행과 검색 결과 정렬 방식은 유사도를 활용하는 기계 학습 방식과 비슷하게 활용 가능하므로, 검토 단계 이전의 정보 처리 과정에서 상황에 적합한 검토 대상 선별을 위한 방법으로 아주 적합하다. 대표적인 고성능 정보 검색 라이브러리로는 Apache 프로젝트 중 하나인 Java 기반의 Lucene이 있으며, 확장

가능한 색인 및 질의 생성, 다국어 및 다양한 문서 포맷 지원, 필터링 및 순위화 등의 고급 검색 기법들이 대체 기술로써 활용 가능하다.[8][15].

4.4 대용량 데이터를 위한 분산 처리 기법

전자 문서 검색을 위한 인덱스 생성이나 기계 학습 또는 가중치 질의 처리를 위한 용어 별 벡터화(vectorization)는 컴퓨터와 같이 빠른 계산 성능을 가진 장치들을 활용해도 많은 시간이 소요되는 작업들이다. 더욱이 연일 화제가 되고 있는 빅 데이터 시대의 도래를 고려해볼 때, 많은 양의 기업 데이터들에 대해 위와 같은 작업을 특별한 접근 방법 없이 그대로 적용한다는 것은 실용성 측면에 있어서 아주 심각한 문제를 발생 시킬 수 있다. 이러한 상황 속에서 대안으로 떠오르고 있는 기법 중 하나가 MapReduce를 활용한 Hadoop 기반의 분산 처리 기술이다[20].

MapReduce는 구글(google)에서 개발한 소프트웨어 프레임워크이며, 이는 페타 바이트(peta byte) 이상의 대용량 데이터를 신뢰할 수 없는 컴퓨터로 구성된 클러스터 환경에서 병렬 처리를 지원하기 위해서 개발되었다. 이 프레임워크는 함수형 프로그래밍에서 일반적으로 사용되는 Map과 Reduce라는 함수 기반으로 구성되는데, Map 함수는 흩어져 있는 데이터를 (Key, Value)의 형태로 연관성 있는 데이터 분류로 묶는 작업을 수행하며, Reduce 함수는 Map 함수의 결과 중 중복 데이터를 제거하고 사용자가 원하는 데이터를 추출하는 작업을 수행한다[19]. E-Discovery 업무를 절차에 따라 수행함에 있어

〈표 5〉 분산 처리 적용 대상 기능 요소 및 기대 효과

단계	적용 대상 기능	기대 효과
정보 관리	- ESI 관리	· 대용량 데이터에 대한 인덱싱 작업을 분산 처리하여, 소요 시간 및 비용을 절감
정보 수집	- 정보 검색	· 동일한 쿼리를 대상 시스템 별로 분할하여 전달하고, 그 결과만을 취합하는 형태로 작업을 수행하여 검색 속도 및 효율성이 향상 됨
정보 처리	- 유사성 분석 - De-Duplication	· 해당 기능 요소 구현의 핵심은 유사성 비교를 위한 ESI 벡터화이며, 이는 기계 학습 적용 가능 대상임
검토 및 분석	- 기밀 정보 분류 - 특허 분류 - 그룹핑 - 연관성 제검토	· 분산 처리 기반으로 기계 학습을 구현하여 기법 자체가 제공하는 효과들 외에도 부가적인 작업 속도 향상을 꾀할 수 있음

서도 기업 내 데이터베이스의 규모와 각 절차상에서 필요한 입력과 출력 데이터 크기, 주요 작업 등을 고려하여 분산 처리가 필요한 부분에 적절하게 해당 프레임워크를 적용한다면, 시간 및 비용 측면에서의 예상 가능한 문제 해결은 물론 업무 효율 향상에도 큰 도움이 될 것이다. 다음 <표 5>는 업무 절차 모델 상에서 분산 처리 기술이 적용될 수 있는 각 단계 별 기능 요소와 기대 효과를 간략하게 설명한 것이다.

5. 결론

본 논문은 미국 전자증거개시 제도의 확산과 국내 도입에 대비하여 표준화된 E-Discovery 업무 수행 절차 확립을 위해 선행 연구된 프로젝트들에 대한 분석을 바탕으로 일반화된 E-Discovery 프로세스와 절차 별 필수 업무 사항 등을 제시하였다. 이는 각기 다른 선행 연구 결과들의 미묘한 견해 차이를 조정하고 최근 법적 요구사항들의 변화 추세와 관련 기술 연구 동향을 반영하여 결여된 부분을 새롭게 포함 시킨 것으로, 소송에 대한 효과적인 대응 방법을 알고자 하는 기업 혹은 소송 당사자들은 물론 해당 분야에서 전문적인 서비스 및 솔루션 개발을 시도하려는 벤더들 모두에게 유용한 정보가 될 것으로 기대 된다. 또한 이러한 절차들이 실제 소송에 활용됨에 있어서 전자증거개시 제도가 근본적으로 내포하고 있는 시간과 비용 문제를 해결하기 위해, E-Discovery 솔루션 구현 및 기업의 시스템 재정비에 활용 가능한 대체 기술들을 제시하였다. 이는 관련 분야 기술 발전 및 한계 극복을 위한 적극적인 대안으로써 충분히 활용될 수 있을 것이라 전망한다.

참 고 문 헌

- [1] 송종혁 (2006), 컴플라이언스, 기업 경쟁력 강화 기회 :비즈니스 5대 원칙 구현 계기, Network times, 통권 159호, 210-213.
- [2] 이창훈, 백승조, 김태완, 임종인 (2008), E-Discovery 를 위한 디지털 증거 전송시스템에 대한 연구, Korea Institute of Information Security and Cryptology, Vol. 18, No. 5, 171-180.
- [3] 한국기록학회 (2008), 기록학 용어 사전, ISBN:978-897-6967-18-3, 역사비평사.
- [4] Adam I. Cogen, G. Edward Kalbaugh (2010), ESI Handbook: Sources, Technology and Process, Aspen Publishers.
- [5] Antonio Kyriakopoulou (2008), Text Classification Aided by Clustering: a Literature Review, Tools in Artificial Intelligence, Paula Fritzsche(Ed.), ISBN: 978-953-7619-03-9, InTech.
- [6] Christopher D. Manning, Prabhakar Raghavan, Hinrich Schutze (2008), Introduction to Information Retrieval, Cambridge University Press, New York.
- [7] Debra Logan (2008), Using the Electronic Discovery Reference Model to Process, Review and Analyze Digital Evidence, Gartner Research, ID Number: G00159094, July 11.
- [8] Erik Hatcher, Otis Gospodnetic, Mike McCandless (2009), Lucene in Action, Manning Publications.
- [9] EDRM Framework Guides, The Electronic Discovery Reference Model, <http://edm.net/resources/guides/edm-framework-guides>
- [10] Fabrizio Sebastiani (2002), Machine Learning in Automated Text Categorization, Journal of ACM Computing Surveys, Volume 34 Issue 1, 1-47, DOI:10.1145/505282.505283.
- [11] Jack Wiles (2007), TechnoSecurity's Guide to E-Discovery and Digital Forensic, Syngress, Elsevier.
- [12] Kenneth J. Withers (2006), Electronically Stored Information: The December 2006 Amendments to the Federal Rules of Civil Procedure, Northwestern Journal of Technology and Intellectual Property, Vol.4(2), 171.
- [13] Learning Task Guidelines, TREC 2011 Legal Track, <http://plg.uwaterloo.ca/~gvcormac/legal11/treclegal11.pdf>
- [14] Linda Volonino, Ian J. Redpath (2010), E-Discovery For Dummies, Wiley, Indianapolis Indiana.
- [15] Lucene Features, Apache Lucene, <http://lucene.apache.org/core/>
- [16] The committee on the judiciary house of representatives, "Federal Rules of Civil Procedures,"

- United States Court, Apr. 2010, <http://www.uscourts.gov/uscourts/RulesAndPolicies/rules/2010%20Rules/Civil%20Procedure.pdf>
- [17] The Sedona Conference Publications, The Sedona Principles after the Federal Amendments, Aug. 2007. <https://thesedonaconference.org/publication/The%20Sedona%20Principles>
- [18] Thomas BARNETT, Svetlana GODJEVAC, Jean-Michel RENDERS, Caroline PRIVAULT, John SCHNEIDER, Robert WICKSTROM (2009), Machine Learning Classification for Document Review, Workshop DESI at ICAIL 2009 (12th International Conference on Artificial Intelligence & Law), Barcelone, Spain, June 8.
- [19] Tom White (2009), Hadoop: The Definitive Guide 1st Edition, O'Reilly Media.
- [20] What is Apache Hadoop, Apache Hadoop, <http://hadoop.apache.org>

이 태 림



- 2008년 2월 : 부경대학교 컴퓨터멀티미디어학과(학사)
- 2010년 2월 : 부경대학교 정보보호학(협)(석사)
- 2010년 3월~현재 : 부경대학교 정보보호학(협)(박사수료)
- 관심분야 : E-Discovery, 디지털 포렌식, 정보 보호

· E-Mail : taeri@pknu.ac.kr

신 상 욱



- 1995년 2월 : 부경대학교 전자계산학과(학사)
- 1997년 2월 : 부경대학교 전자계산학과(석사)
- 2000년 2월 : 부경대학교 전자계산학과(박사)
- 2000년 4월~2003년 8월 : 한국전자통신연구원 선임연구원

- 2003년 9월~현재 : 부경대학교 IT융합응용공학과 부교수
- 관심분야 : E-Discovery, 디지털 포렌식, 모바일 네트워크 보안, 암호 프로토콜
- E-Mail : shinsu@pknu.ac.kr