

# 원문정보공개 지원을 위한 민감정보 필터링 요건에 관한 연구

## A Study on Sensitive Information Filtering Requirements for Supporting Original Information Disclosure

오진관 (Jin-Kwan Oh)\*, 오세라 (Seh-La Oh)\*\*  
최광훈 (Kwang-Hoon Choi)\*\*\*, 임진희 (Jin-Hee Yim)\*\*\*\*

### 목 차

- |                                      |                     |
|--------------------------------------|---------------------|
| 1. 머리말                               | 4. 원문정보공개 프로세스의 재설계 |
| 2. 민감정보의 정의와 필터링 도구의 필요성             | 5. 맺음말              |
| 3. 기계학습 기반의 필터링 기능의 작동 방식<br>분석과 시사점 |                     |

### <초 록>

원문정보공개 서비스를 개시한 후 한해 천만 건에 가까운 전자 결재문서가 온라인을 통해 공개되고 있다. 하지만 대량의 전자결재문서를 정보공개 업무담당자가 모두 확인하여 원문정보공개 서비스를 수행하는 것은 현실적으로 불가능한 상황이다. 이에 따라 최근 일부 공공기관에서는 개인정보 필터링 도구를 활용하여 문서 생산단계에서 정형화된 개인정보를 필터링하고 있으나 비정형화된 민감정보는 관리되지 않고 있다. 본 연구에서는 원문정보공개 지원을 위해 사용 중인 필터링 도구 분석을 통해 필터링 도구의 고도화 방향을 설정하였으며, 필터링 도구 활용단계가 추가된 결재문서 본문 작성과 원문정보공개 프로세스를 재설계하였다.

주제어: 정보공개, 원문정보공개, 필터링, 개인정보, 민감정보, 기계학습

### <ABSTRACT>

Approximately 10 million electronic approval documents have been released online since the commencement of the original information disclosure service. However, it is practically impossible to carry out an original information disclosure service by confirming a large amount of electronic approval documents to all persons in charge of information disclosure. Recently, some public organizations have been using private information filtering tools to filter personal information at the stage of document production, but the management of different sensitive information has not been managed using solutions. In this study, we set up the advanced direction of the filtering tool by analyzing the filtering tool in use to support the original information disclosure, and redesigned the text of the approval document and the original information disclosure process with the use of the filtering tool.

Keywords: information disclosure, original information disclosure, filtering, personal information, sensitive information, machine learning

\* 명지대학교 기록정보과학전문대학원 기록관리학전공 박사 수료, 알엠소프트 기록연구소 선임연구원 (jinkwan5@rmssoft.co.kr), (jinkwan5@gmail.com) (제1저자)

\*\* 네모아이씨티(주) (sehla@naver.com) (공동저자)

\*\*\* 알엠소프트(주) (1004@rmssoft.co.kr) (공동저자)

\*\*\*\* 명지대학교 기록정보과학전문대학원 부교수, '인간과기억아카이브'(hmarchives.org) 아키비스트 (yimjhkr@mju.ac.kr) (교신저자)

■ 접수일: 2017년 1월 20일 ■ 최초심사일: 2017년 1월 31일 ■ 게재확정일: 2017년 2월 24일  
■ 한국기록관리학회지 17(1), 51-71, 2017. <<http://dx.doi.org/10.14404/JKSARM.2017.17.1.051>>

## 1. 머리말

우리나라는 1998년 1월부터 「공공기관의 정보공개에 관한 법률」을 시행하여 아시아에서 최초로 정보공개제도를 도입하였으며, 2014년 3월부터는 133개 기관(중앙 47, 시·도 17, 시·군·구 69)에서 원문정보공개 서비스를 개시한 후 2017년 1월 현재 12,179개 공공기관에서 원문정보공개 서비스를 진행하고 있다. 원문정보공개는 전자적 형태로 보유·관리하는 정보 중 공개대상으로 분류된 정보를 국민의 청구가 없더라도 정보통신망을 활용한 정보공개시스템 등을 통하여 공개하는 제도이다(행정자치부, 2015). 이 제도의 시행을 통해서 이전의 '청구에 의한 공개' 중심에서 공공기관이 '적극적인 정보 공개' 방식으로 전환되었다.

행정자치부 정보공개포털(www.open.go.kr)을 통해 2016년 한 해 동안 523만 건의 원문정보가 공개되었다. 정보공개포털을 통해 현재 원문정보공개되는 결재문서는 행정기관의 경우 국장급 이상, 공공기관의 경우 임원급 이상의 결재문서가 공개되고 있다. 또한 서울시는 2013년 10월부터 내부에서 생산된 행정정보가 자동으로 인터넷에 공개되는 서비스인 정보소통광장(opengov.seoul.go.kr)을 운영 중에 있다. 서비스 오픈 이후 2017년 1월 현재까지 정보소통광장을 통해 공개되고 있는 전자결재문서는 1,000만 건 이상이다. 전자결재문서 1건에 10장의 종이문서가 생산된다고 가정하였을 때 요구되는 보존서고 면적은 서울광장의 약 3배이며, 매년 서울시

에서 원문정보공개하는 전자결재문서를 종이 기록물로 관리한다면 매년 서울 광장 규모의 보존서고가 요구된다.<sup>1)</sup> 공개포털 사이트 운영 이후 상당히 많은 원문정보가 매일 온라인으로 공개되고 있다.

「공공기관의 정보공개에 관한 법률」 제7조에 따르면, 행정정보는 공개가 원칙이지만 제9조제1항(비공개대상정보)이 포함된 경우에는 공개 유예를 할 수 있다. 「개인정보 보호법」 제17조(개인정보의 제공)에 따라 개인정보는 정보주체의 동의 없이 개인정보가 유출되지 않아야 한다. 공개유예 해야 할 정보나 보호해야 할 개인정보를 잘 가려내는 것이 현 시점에서 중요한 과제라 할 수 있으며, 이런 문제를 해결하기 위해 일부 공공기관들은 개인정보 필터링도구를 사용하고 있다.

현재 일부 공공기관에서는 전자문서 생산 중에 개인정보의 필터링을 위한 솔루션을 사용 중이다. 이 솔루션의 구동 시점은 업무담당자가 문서 작성을 완료하고, 결재 이전에 필터링 도구를 구동하여 본문과 첨부파일의 개인정보 필터링을 진행한다. 본문과 첨부파일의 필터링이 마치면 문서관리카드에 포함된 개인정보를 팝업창에 알려준다. 하지만 현재 일부 공공기관에서 사용하고 있는 필터링 도구는 개인정보에 대한 필터링만 실행되고 있어 일반적인 패턴을 가진 개인정보 이외의 민감한 정보는 필터링 도구를 통해 걸러지지 않는 문제점이 있다. 그래서 기계학습 기반의 인공지능을 갖춘 고도화된 필터링 도구가 필요하다. 필터링 도구의 기능은

1) 산출방법은 「공공기록물 관리에 관한 법률」 시행령 별표 6(기록물관리기관의 보존시설 및 장비의 기준)의 이동식 보존서고(전자기록물) 시설기준 100,000장당 40m<sup>2</sup>에 따랐으며, 서울광장 규모는 13,207m<sup>2</sup>로 산정하여 계산하였다.

이전에 걸러지지 못한 정보를 걸러내는 것도 중요하지만, 필터링 도구를 언제 누가 어떻게 사용하는 것도 중요한 문제이다. 필터링 도구의 자동화에만 의존할 것이 아니라 업무담당자가 도구를 적절히 활용할 수 있도록 하면서 개인정보를 포함한 민감정보가 유출되지 않도록 한다는 본래의 목적을 달성할 수 있게 프로세스를 설계하는 것도 중요하다.

기록관리 분야에서 필터링 도구와 관련된 연구는 현재까지 수행되지 않았다. 하지만 원문정보공개 관련 연구는 황진현, 임진희(2015)의 '체계적인 원문정보공개를 위한 정부산하공공기관 기록관리시스템 기능요건 연구', 허준석, 홍덕용(2016)의 '원문정보공개서비스를 위한 전자기록 구성요소에 관한 연구'가 있다. 황진현, 임진희(2015) 연구에서는 공공기관별 업무특성이 반영된 자체개발 기록관리시스템 구축을 통하여 기록관과 기록관리시스템에 기반을 둔 원문정보공개가 이루어질 수 있는 시스템 간 연계도 구상을 통해 원문정보공개 대응을 위한 기능요건을 제시하였고, 허준석, 홍덕용(2016) 연구에서는 원문정보공개 서비스에서 전자기록 구성요소에 따라 관리하여 서비스를 향상시킬 수 있는 요인을 밝혔다.

본 연구의 목적은 원문정보공개 지원을 위해 고도화된 필터링 도구가 갖춰야 할 기능요건 도출과 필터링 도구를 활용한 원문정보공개 프로세스를 제안하는데 있다. 연구 수행 방법은 개인정보, 민감정보 정의와 종류를 확인하기 위해 각종 법규를 분석하였고, 현재까지 개발된 필터링 도구의 기능 분석을 위해 나라장터에 등록된 18개의 필터링 솔루션 기능을 분석하여 기계학습 기반의 필터링 기능의 작동 방식을 설계하고, 원문정보공개 프로세스를 재설계하였다.

## 2. 민감정보의 정의와 필터링 도구의 필요성

### 2.1 개인정보의 정의 및 종류

우리나라 법률에 명시된 '개인정보'의 정의는 법률에 따라 약간의 차이가 있지만 그 의미는 비슷하다. 「개인정보 보호법」에서 개인정보란 살아 있는 개인에 관한 정보로서 성명, 주민등록번호 및 영상 등을 통하여 개인을 알아볼 수 있는 정보를 말한다. 「전자서명법」에서 개인정보라 함은 생존하고 있는 개인에 관한 정보로서 성명·주민등록번호 등에 의하여 당해 개인을 알아볼 수 있는 부호·문자·음성·음향·영상 및 생체특성 등에 관한 정보를 말한다. 「정보통신망 이용촉진 및 정보보호 등에 관한 법률」에서 개인정보란 개인에 관한 정보로서 성명·주민등록번호 등에 의하여 특정한 개인을 알아볼 수 있는 부호·문자·음성·음향 및 영상 등의 정보를 말한다.

법률에서 정의된 개인정보의 개념은 상당히 포괄적이라 구체적인 개인정보의 범위를 설정하는데 어려움이 있다. 개인정보 종류에 대한 보다 명확한 범위는 2015년 개정된 행정자치부 '개인정보보호 지침'에 제시되어 있는데 성명, 주소, 연락처, 소득, 학력, 성적, 직업, 전자우편, 영상, 통화내용, 신용, 부채, 인터넷 접속 IP 등 객관적 사실에 관한 정보와 개인에 대한 제3자의 의견이나 평가 등 주관적 정보를 포함하는 것을 말한다. 이밖에도 행정자치부가 운영 중인 개인정보보호 종합포털에는 개인정보의 예시를 <표 1>과 같이 제시하고 있다.

단 하나의 개인정보로 특정 개인을 알아볼 수

〈표 1〉 개인정보의 예시

유형 구분	개인정보 항목
일반정보	이름, 주민등록번호, 운전면허번호, 주소, 전화번호, 생년월일, 출생지, 본적지, 성별, 국적
가족정보	가족구성원들의 이름, 출생지, 생년월일, 주민등록번호, 직업, 전화번호
교육 및 훈련정보	학교출석사항, 최종학력, 학교성적, 기술 자격증 및 전문 면허증, 이수한 훈련 프로그램, 동아리활동, 상벌사항
병역정보	군번 및 계급, 제대유형, 주특기, 근무부대
부동산정보	소유주택, 토지, 자동차, 기타소유차량, 상점 및 건물 등
소득정보	현재 봉급액, 봉급경력, 보너스 및 수수료, 기타소득의 원천, 이자소득, 사업소득
기타 수익정보	보험 (건강, 생명 등) 가입현황, 회사의 판공비, 투자프로그램, 퇴직프로그램, 휴가, 병가
신용정보	대부잔액 및 지불상황, 저당, 신용카드, 지불연기 및 미납의 수, 임금압류 통보에 대한 기록
고용정보	현재의 고용주, 회사주소, 상급자의 이름, 직무수행평가기록, 훈련기록, 출석기록, 상벌기록, 성격 테스트결과 및 직무태도
법적정보	전과기록, 자동차 교통 위반기록, 파산 및 담보기록, 구속기록, 이혼기록, 납세기록
의료정보	가족병력기록, 과거의 의료기록, 정신질환기록, 신체장애, 혈액형, IQ, 약물테스트 등 각종 신체테스트 정보
조직정보	노조가입, 종교단체가입, 정당가입, 클럽회원
통신정보	전자우편(E-mail), 전화통화내용, 로그파일(Log file), 쿠키(Cookies)
위치정보	GPS나 휴대폰에 의한 개인의 위치정보
신체정보	지문, 홍채, DNA, 신장, 가슴둘레 등
습관 및 취미정보	흡연, 음주량, 선호하는 스포츠 및 오락, 여가활동, 비디오 대여기록, 도박성향

없을지라도 다른 정보와 용이하게 결합할 경우 당해 개인을 알아볼 수 있는 조합된 정보도 「정보통신망 이용촉진 및 정보보호 등에 관한 법률」 제2조제1항제6호에 따라 개인정보로 볼 수 있다.<sup>2)</sup> 예컨대 취미나 출신학교 등의 정보는 그 자체만으로는 개인을 식별할 수 없으나 성명과 결합하여서는 당해 개인을 알아볼 수 있는 개인 정보에 해당한다(정연수 외, 2007).

「공공기록물 관리에 관한 법률」 제3조 및 시행령 제3조에 따른 공공기관<sup>3)</sup>은 '기록물 관리 지침'을 준수하여 개인정보가 포함된 기록물을 관리하여야 한다. 개인정보가 포함된 기록물의 유형의 내용과 그 사례는 〈표 2〉와 같다.

〈표 2〉 개인정보가 포함된 기록물의 유형

구분	내용	사례
유형1	- 업무관리시스템, 전자문서 시스템에 생산·등록된 기록물에 포함된 개인정보	- 전자문서의 본문 및 붙임 파일에 포함된 개인정보
유형2	- 행정정보시스템 보유 개인정보	- 인사관리시스템의 인사 기록카드, 교육정보시스템의 학생생활기록부 등
유형3	- 서면신청, 홈페이지 등 기타방식으로 수집한 개인 정보	- 홈페이지(회원접수, 전자접수) - 오프라인 서면 신청서 등

출처: 2016년도 기록물 관리지침, p. 20.

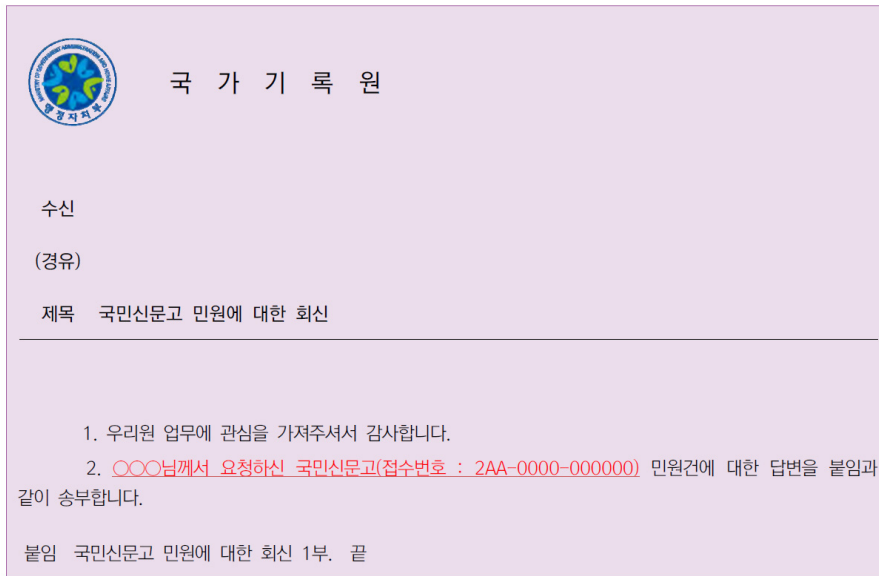
유형1의 업무관리·전자문서시스템 기록물에 포함된 개인정보는 「공공기록물 관리에 관한 법률」에 따라 관리되어야 하고, 개인정보가 포함된 기록물의 암호 설정 금지 및 기록물 공

2) 서울중앙지방법원 2007.02.08. 선고 2006가합33062 판결.

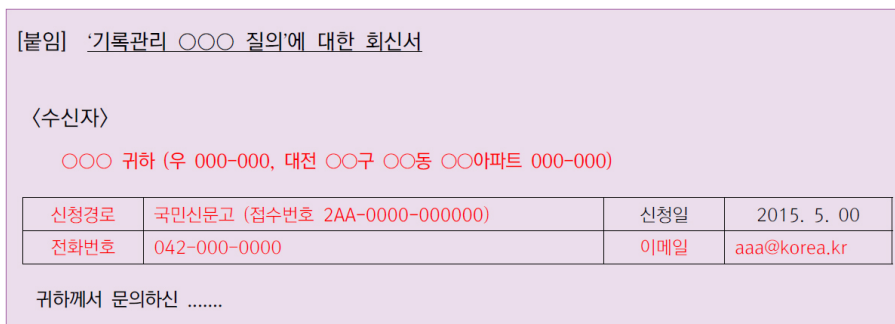
3) 공공기관의 범주는 국가기관, 지방자치단체, 「공공기관의 운영에 관한 법률」 제4조에 따른 공공기관, 「지방공기업법」에 따른 지방공사와 지방공단, 특별법에 따라 설립된 특수법인(「지방문화원진흥법」에 의한 문화원 및 특별법에 의하여 설립된 조합·협회 제외), 「초·중등교육법」 및 「고등교육법」, 그 밖에 다른 법률에 따라 설립된 각급 학교를 말함

개 시 개인정보를 식별하지 못하도록 조치하여야 한다. 업무관리시스템<sup>4)</sup>과 전자문서시스템<sup>5)</sup>은 「공공기록물 관리에 관한 법률」 시행령 제3조에 따른 전자기록 생산시스템이다. 이들 시

스템의 본문(기안문) 및 붙임파일에 포함된 개인정보는 기록관리 대상이며 개인정보가 포함되어 업무관리시스템에 등록·관리되는 기록물 사례는 <그림 1>~<그림 2>와 같다.



<그림 1> 개인정보가 포함되어 업무관리시스템에 등록·관리되는 기록물 사례(본문)



<그림 2> 개인정보가 포함되어 업무관리시스템에 등록·관리되는 기록물 사례(붙임파일)

- 4) “업무관리시스템”이란 행정(공공)기관이 업무처리의 전 과정을 과제관리카드 및 문서관리카드 등을 이용하여 전자적으로 관리하는 시스템(행정 효율과 협업 촉진에 관한 규정 제3조)
- 5) “전자문서시스템”이란 문서의 기안·검토·협조·결재·등록·시행·분류·편철·보관·보존·이관·접수·배부·공람·검색·활용 등 모든 처리절차가 전자적으로 처리되는 시스템(행정 효율과 협업 촉진에 관한 규정 제3조)

유형2의 행정정보시스템 보유 개인정보는 기록물의 생산부터 보존기간까지 보존·관리되어야 하며, 개인정보 보유기간이 경과하였으나 해당 개인정보가 포함된 기록물의 보존기간이 만료되지 않은 경우 개인정보를 파기하지 않도록 주의하여야 한다. 행정정보시스템의 개인정보 사례는 <표 3>과 같다.

유형3의 홈페이지, 서면신청 등 기타방식으로 수집한 개인정보는 전자기록생산시스템에 등록하여 관리하여야 하며, 개인정보 보유기간이 경과하였으나 해당 개인정보가 포함된 기록물의 보존기간이 만료되지 않은 경우 개인정보를 파기하지 않도록 주의하여야 한다. 홈페이지를 통해 수집한 개인정보가 포함된 기록물 중 업무관리시스템과 전자문서시스템과 연계되어

등록·관리되는 경우는 유형1, 행정정보시스템과 연계되어 등록·관리되는 경우 유형2의 관리 절차에 따라야 한다.

### 2.2 민감정보의 정의 및 종류

2011년 9월부터 「개인정보 보호법」이 본격 시행됨에 따라 기존의 개인정보보호와 관련하여 언급되지 않았던 새로운 개념인 '민감정보' 용어를 사용하게 되었다. 「개인정보 보호법」 제23조에서 '민감정보'란 사상·신념, 노동조합·정당의 가입·탈퇴, 정치적 견해, 건강, 성생활 등에 관한 정보, 그 밖에 정보주체의 사생활을 현저히 침해할 우려가 있는 개인정보로서 대통령령으로 정하는 정보<sup>6)</sup>를 말한다. 이러한 민감정보는 객

<표 3> 개인정보 보유 행정정보시스템 사례

시스템명	세부 내역	운영주체	관련법령
자동차민원행정종합정보시스템	차량의 등록과 말소 변경, 이전, 검사 등 전반적인 자동차 민원업무 서비스 제공	교통안전공단	자동차관리법
국민신문고	민원신청 및 국민제안 접수 처리 정보 등	국민권의위원회	부패방지 및 국민권익위원회의 설치와 운영에 관한 법률
건설행정정보시스템	지방자치단체 건설업 등록 행정업무 처리 및 건설업체정보	국토연구원	건설산업기본법 시행령
외국인등록정보시스템	시·군·구에서 수행하고 있는 외국인 체류 관련업무를 지원	법무부	출입국관리법
가족관계등록부시스템	주민 조사 및 가족 관계 등록 정보 등	법원행정처	가족관계등록법
사회복지통합관리망	개인별·가구별 복지대상자 관리 및 복지 급여 및 서비스 지원, 부정수급 관리 등 복지 행정 정보 등	보건복지부	사회복지사업법 및 개별 복지사업법
신여권통합정보관리시스템	여권 신청·접수 및 심사, 발급여권 교부 정보 등	외교부	여권법
공직윤리종합정보시스템	공직자 재산등록 및 심사, 금융거래 및 부동산 정보 등	인사혁신처	공직자윤리법
상훈포털시스템	훈격, 성명, 소속, 훈격, 수여일, 포상명 등 정부포상 정보	행정자치부	상훈법
인감전산정보시스템	증명청에서의 인감증명(외국인포함)의 발급 및 발급 대장의 관리 정보	행정자치부	인감증명법
자치단체통합인터넷원서접수시스템	시험시행계획관리, 원서접수, 접수증 및 응시표, 합격자 정보 등	행정자치부	지방공무원 임용령
주민등록관리시스템	전국 지자체별(시·도/시·군·구/읍·면·동) 인구, 세대 정보	행정자치부	주민등록법
통합정보공개시스템	정보공개청구 및 결정통지 정보 등	행정자치부	공공기관의 정보공개에 관한 법률
통합지방세정보시스템	지방세 대민 서비스(신고·납부) 정보	행정자치부	지방세법

6) 「개인정보 보호법」 시행령 제18조(민감정보의 범위)에 따르면 대통령령이 정하는 정보란 유전자검사 등의 결과로 얻어진 유전정보, 「형의 실효 등에 관한 법률」 제2조제5호에 따른 범죄경력 자료에 해당하는 정보를 말함.

관적인 신상정보를 포함하는 개인정보 항목과 비교하여 구체적이지 못하지만 침해나 유출시 정보주체의 사생활에 보다 큰 영향을 미칠 수 있기 때문에 엄격하게 관리해야 할 것으로 판단된다. 본 논문에서는 민감정보의 의미는 개인정보 의미를 모두 포함하고, 구체적으로 정의하지 못하지만 침해나 유출시 정보주체의 사생활에 큰 영향을 미칠 수 있는 모든 정보의 의미로 사용하였다.

민감정보의 처리는 원칙적으로 금지되며, 다음과 같은 경우에는 예외적으로 처리가 가능하

다. 첫째, 정보주체의 별도 동의가 있는 경우에는 다른 개인정보의 처리와 분리하여 민감정보 처리에 대해 정보주체가 이를 명확히 인지하고 명시적으로 자신의 동의 의사를 밝힌 경우이다. 둘째, 다른 법률에서 명시적으로 민감정보 처리를 요구하거나 허용하는 경우이다. 또한 다른 법령에서 민감정보의 처리를 구체적으로 언급하고 있거나 해석 상 요구되는 경우도 포함된다.<sup>7)</sup>

지방자치단체인 ○○구에서 도시개발과 관련하여 정의하고 있는 민감정보의 종류는 <표 4>

<표 4> 도시개발관련 민감정보의 종류

유형	민감정보
감사일반	- 완료되지 않은 감사내용 - 개인정보(성명, 주민등록번호, 급여내역 등)
주거환경개선	- 주거환경개선 사업 업무에 대한 정보 - 관계자 및 이주대상자 개인정보 - 법인의 영업과 관련된 정보 - 부동산 매매에 영향을 미칠 수 있는 정보
공공용지	- 공공용지 취득자의 개인정보, 부지 소유자들의 재산정보
국·공유재산 감정평가	- 감정평가사 법인정보 - 토지소유자의 개인정보 및 재산정보, 부동산정보
국·공유재산 관리	- 개인의 성명과 주소 - 재산이 담긴 점보와 법인의 정보 - 부동산 매매에 영향을 끼칠 수 있는 정보
국·공유재산 등기	- 성명과 주소 등의 개인정보와 재산정보 - 부동산 매매에 영향을 미칠 수 있는 정보
제·개정 및 폐지	- 개정이나 폐지 논의 중으로 의사결정과정에 있는 법률 내용 및 의견
재개발보상	- 토지 및 건물소유자의 개인정보 - 보상에 따른 재산과 관련된 정보 - 부동산매매에 영향을 미칠 수 있는 정보
도로재개발	- 재개발 사업 계획(안) 정보 - 재개발 사업 지역과 개인에 대한 정보 - 도로 확장 예정지에 대한 부동산정보
실시계획 및 허가	- 도시계획에 따른 시설배치도등과 내부 업무 협의 과정에 대한 정보 - 개인과 법인의 정보 - 부동산 매매에 영향을 미칠 수 있는 정보
주택재개발	- 재개발 구역 거주자 개인정보 - 계획(안), 재개발 사업 관련 지시, 부동산 매매에 영향을 미칠 수 있는 정보
토지수용	- 토지 수용에 따른 개인의 성명과 주소, 재산을 추측할 수 있는 정보 - 부동산 매매에 영향을 미칠 수 있는 정보

7) 「개인정보 보호법」 제23조(민감정보의 처리 제한)

와 같다. 민감정보의 종류에는 일반적인 패턴을 지닌 개인정보 유형부터 특정한 패턴을 가지지 못하지만 침해나 유출시 큰 영향을 미칠 수 있는 정보 유형이 있다. ○○구에서 정의한 민감 정보 종류는 5년 주기 공개재분류 과정에서 비공개 대상정보를 정리한 후 민감정보를 정의하였으며, 이들 정보가 포함된 문서의 경우는 비공개 또는 부분공개 처리하여 기록을 관리하고 있다.

### 2.3 필터링 도구의 기능과 필요성

2017년 1월 현재 나라장터 종합쇼핑몰(<http://shopping.g2b.go.kr>)에 개인정보보호관련 보안 소프트웨어를 등록하여 판매하는 업체는 18개 이다. 시판 중인 필터링 솔루션의 종류에는 PC 내 존재하는 개인정보가 포함된 문서를 검색하고, 개인정보 포함여부를 실시간 알림으로 개인

정보 노출 위험성을 발견하는 솔루션, 기관이 운영 중인 웹사이트 내 게시물 개인정보 필터링 솔루션, 업무관리시스템에 등록되는 문서와 게시글의 본문 내용 또는 첨부되는 파일 안에 개인정보 포함 여부를 검사해주는 솔루션, 웹서버, DB 서버 등 서버 내 존재하는 웹소스, 파일, DB데이터를 대상으로 개인정보 포함 여부를 검색하고 결과를 보고해주는 솔루션 등 개인정보보호를 위한 다양한 솔루션이 있다. 이들 솔루션의 주요 기능을 종합하여 정리하면 <표 5>와 같다.

서울특별시는 업무관리시스템의 개인정보 보호 솔루션 구축을 통한 사전유출 예방체계를 확립하여 개인정보 보호를 강화하고, 생산문서에 대한 개인정보 보호 강화를 통해 시정의 신뢰성 제공을 위해 2015년에 ‘업무관리시스템 개인정보보호 솔루션 도입 구축’ 사업을 수행하였다. 사업의 주요내용은 업무관리시스템과 연계하여

<표 5> 시판중인 필터링 솔루션 주요기능

주요기능	기능상세
개인정보 차단기능	<ul style="list-style-type: none"> <li>- 지능적인 개인정보 차단기능 제공</li> <li>- 주민등록번호, 유사주민등록번호, 카드번호, 계좌번호, 휴대폰 번호, 이메일주소, 법인번호, 사업자번호 등 사전 차단 기능</li> <li>- 정규표현식(*, ? 등) 특수문자, 공백이 포함된 개인정보 사전 차단 기능</li> </ul>
불건전게시물 차단기능	<ul style="list-style-type: none"> <li>- 유해어 DB를 이용한 욕설, 음란 글, 상업광고 사전 차단</li> <li>- 한글 정규표현식을 적용한 특수문자, 공백이 포함된 욕설, 음란글, 상업광고 사전 차단 기능</li> <li>- 관리자 금칙어 차단 및 관리자 허용단어 차단해제 기능</li> </ul>
첨부파일 차단기능	<ul style="list-style-type: none"> <li>- 차단 문서형태 항목                             <ul style="list-style-type: none"> <li>• Microsoft Office, 한글과컴퓨터 한글, 엑셀, 워드, 파워포인트, 한글 문서의 OLE 문서, Adobe Acrobat(PDF), Rich Text Format(RTF), 알태랑(JDT8~2005), Microsoft Document Imaging(MDI), Microsoft Outlook, Open Office, 워드퍼펙트(WPD), Autodesk Drawing File(DWG), Flash Movie File(Swf), 압축파일(ZIP, TAR, GZIP, ZOO, TGZ, RAR, LHA, ALZ, BZIP2 등), 핸디소프트(HWX,HWN), 핸디소프트아리랑(HWD), XML / HTML / MHT / CHM / EML / MIME / TEXT, 프로그래밍 소스코드로 인식되는 확장자(c, cpp, h, hpp, java, rc, awk, bas, bat, aspx, php, jsp 등), 4그 외 명시되지 않고 비주기적으로 생성되는 모든 형태의 문서 파일에 대한 텍스트 추출 및 필터링 기능</li> </ul> </li> </ul>
전수검사 기능	<ul style="list-style-type: none"> <li>- 홈페이지 본문, 첨부파일, 웹서버 등에 존재한 모든 파일 검사</li> </ul>

개인정보 포함여부 확인 기능 개발, 문서 생성 시 개인정보 탐지를 위한 탐지전용시스템 구축, 개인정보 탐지 솔루션과 업무관리시스템의 연동을 위한 시스템 구축 및 지원, 개인정보 탐지 전용시스템의 최적화된 운영환경 구축 및 장애 지원이었다. 개인정보필터링시스템의 주요기능은 업무관리시스템의 개인정보 검출 기능과 개인정보 검출모듈과 업무관리시스템과의 연동이며, 상세기능은 <표 6>과 같다.

현재 공공기관 납품용 필터링 도구는 정형 방식의 민감정보를 차단할 수 있지만 비정형 대상 정보는 일정한 형식을 가지지 않아 필터링 도구를 구동하여도 민감정보를 거를 수 없다. 또한 업무담당자들이 문서작성이나 결재과정 중에 확인한 민감정보를 직접 마킹하는 기능이 부재하여 문서상의 민감정보를 사용자에게 알려주지 못하고, 정보공개 서비스용 사본 생성을 위

해 민감정보를 비식별 화하기 위한 마스킹 기능이 구현되지 않았다.

업무관리시스템 이용자 약 34만 5천명이 연간 수천 만 건 이상의 문서를 생산하고 있으며, 다양한 유형의 민감정보를 업무담당자가 직관적으로 판단하여 생산되는 모든 문서에 적용하여 처리하기가 현실적으로 어렵다. 이 때문에 업무담당자가 문서 생산과정에서 필터링 도구를 활용하여 민감정보를 1차적으로 걸러 낼 수 있어야 한다. 민감정보를 필터링하는 도구 이외에도 마킹(marking)과 매스킹(masking)을 통하여 문서 생산자가 문서를 생산하는 과정 중에 민감정보를 거르고, 기계학습 방식으로 각 기관의 특화된 민감한 정보를 관리할 수 있는 필터링 도구가 필요하다. 또한 원문정보공개 지원을 위해 업무관리시스템에서 결재되어 기관차원에서 공유하는 문서외는 별도로 대국민 서비스를 위한 발

<표 6> 서울특별시 업무관리시스템 연계 개인정보필터링시스템 기능요건

주요기능	기능상세
업무관리시스템의 개인정보 검출	<ul style="list-style-type: none"> <li>- 개인정보 연동 모듈(S/W)을 통해 전송 받은 파일에 대하여 개인정보 검출 수행</li> <li>- 전자문서 기안/결재 시 문서(본문 및 첨부파일 포함) 개인정보 검출</li> <li>- 개인정보 검출 대상은 주민등록번호, 운전면허번호, 여권번호, 외국인등록번호, 이메일주소, 전화번호, 휴대폰번호, 카드번호, 법인등록번호, 사업자번호 등이 검출</li> <li>- 관리자가 필터 항목을 설정 하여, 설정된 패턴별로 개인정보 검출 가능</li> <li>- 복수의 첨부파일이 함께 첨부된 경우에도 개인정보 검출</li> <li>- 검출된 내용에 대하여 사용자에게 상세 검출 알림 제공</li> <li>- 검출 대상 문서 종류는 HWP, MS Word, Power Point, Excel, PDF, TEXT, HWX 등에 대한 다수 포맷 지원</li> <li>- 검출한 기록을 확인할 경우, 검출 로그 등을 통해 관리자가 확인 가능</li> <li>- 로그를 보관할 최대기간 설정 기능 및 설정 지난 로그 삭제 기능</li> <li>- 검출한 결과에 대하여 통계를 제공하고 보고서 형태의 엑셀파일과 이미지형태의 그래프 출력 기능</li> <li>- 시스템 현황(트래픽, CPU, HDD, 동시접속자 등)을 그래프 형태로 모니터링 가능한 기능</li> </ul>
개인정보 검출모듈과 업무관리시스템과 연동	<ul style="list-style-type: none"> <li>- 전자문서 기안기 및 결재기 기능개선</li> <li>- 업무관리시스템과 개인정보 검출시스템간의 검출요구 연동(request) 기능</li> <li>- 업무관리시스템과 개인정보 검출시스템간의 검출결과 연동(response) 기능</li> <li>- 개인정보검출 솔루션 및 업무관리시스템과 연동을 위한 관리자 환경 지원 기능</li> </ul>

취사본 혹은 편집사본 제작기능 등이 요구된다. 즉, 필터링 도구는 정보유출의 위험을 관리하는 차원에서 유용한 도구이며 필수적이 도구이다.

### 3. 기계학습 기반의 필터링 기능의 작동 방식 분석과 시사점

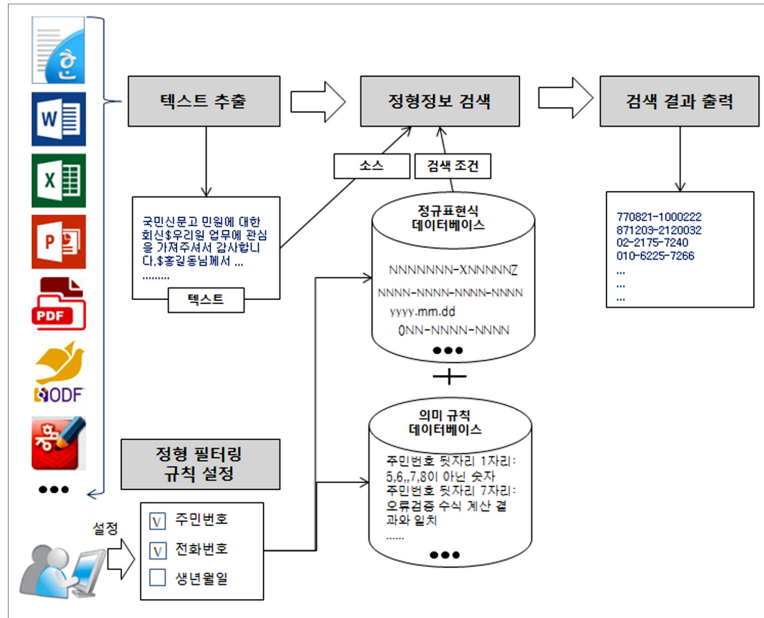
#### 3.1 정형 정보의 필터링 구현 방식

정형 정보는 규칙화할 수 있는 정보로서 외형적인 패턴을 갖는 정보이며 일정한 규칙을 갖는 문자열로 구별해낼 수 있는 정보이다. 예를 들어, 주민등록번호는 생년월일(YYMMDD) 6자리와 하이픈(-) 문자로 연결하여 뒤 7자리로 구성된다. 국내 유선전화번호는 2~3자리 지역번호, 3~4자리 국번과 4자리의 번호로 이루어진다. 이와 같이 일정한 규칙을 갖는 문자열을 표현하는 규칙을 정규표현식이라고 하며 컴퓨터 프로그래밍에서 문자열 검색과 치환에 이용되고 있다. 주민등록번호를 POXIS(Portable Operating System Interface) 문법으로 간단히 표현한다면, '숫자 6자리 + 하이픈 + 숫자 7자리' 패턴을  $[0-9]{6}-[0-9]{7}$ 로 기술한다. 그러나 이 표현식에 의하면 앞자리 생년월일의 월일 숫자에 해당되지 않는 숫자를 갖는 문자열인 999999-9999999도 허용되기 때문에 좀 더 정확한 주민번호 표현식이 필요할 것이다. 생년월일의 월 2자리 중 첫자리는 0 또는 1만이 허용되어야 하며, 날짜의 앞자리는 0~3 사이 숫자만 유효하다. 보다 더 정교하게 기술한다면  $[0-9][0-9][01][0-9][0-3][0-9]-[0-9]{7}$ 로 표현된다.

정규표현식 검색 결과의 정확도를 높이기 위하여 외적인 형식으로 표현할 수 없으나 의미적으로 유효성을 판정할 수 있는 규칙을 추가한다. 주민등록번호의 뒷자리 중 마지막 숫자는 검증식에 의해 계산된 오류 검증용 숫자로서 정규표현식으로 표현될 수 없는 규칙을 가지고 있다. 이를 본 논문에서는 '의미 규칙'이라 명명하기로 한다.

정형 정보 필터링 프로세스를 도식화하여 보면 <그림 3>과 같다. 첫 번째 단계는 필터링 대상 파일로부터 '텍스트 추출'하는 과정으로서 필터링 도구마다 구현 방법이 다양할 수 있다. 각각의 포맷별로 텍스트만을 추출하는 필터를 채용하는 방안, 공통 포맷으로서 PDF와 같은 보편화된 포맷으로 변환하여 그대로 그 포맷을 이용하는 방안, 또는 공통 포맷으로부터 다시 텍스트를 추출하는 방안이 있으며 모든 텍스트를 정확히 추출할 수 있어야 한다.

다음 단계는 추출한 텍스트를 대상으로 '정규 정보 검색'하는 과정으로 필터링 조건으로 설정한 정규표현식 데이터베이스와 의미 규칙 데이터베이스로부터 검색 조건을 설정하고 텍스트를 검색한다. 필터링할 정형 정보의 종류는 사전에 사용자에게 의해서 선택되어 설정되어 있어야 한다. 일정한 규칙을 갖는 패턴이기 때문에 사용자는 단순히 정보 종류를 선택하고 필터링 도구에 의해 해당 유형을 식별하기 위하여 정의하여 둔 규칙을 검색 조건으로 사용하도록 활성화된다. 검색 과정에서 검색 속도와 정확도는 정규표현식과 의미 규칙의 결합 및 검색 알고리즘, 정규표현식과 의미 규칙 데이터의 정확도에 의해 결정되고 이를 구현하는 필터링 도구마다 다르다.



〈그림 3〉 정형 정보 필터링 프로세스

정형 정보 필터링의 마지막 단계는 '검색 결과 출력'이다. 이 단계에서는 검색 조건 표시 여부, 문자열의 위치 정보 표시 여부, 필터링 사유 표시 여부 등이 필터링 도구에 따라 달리 구현될 수 있으며, 정형 정보의 출력결과는 정형 정보를 사용하는 사용자가 식별할 수 있다.

### 3.2 비정형 정보의 필터링 구현 방식

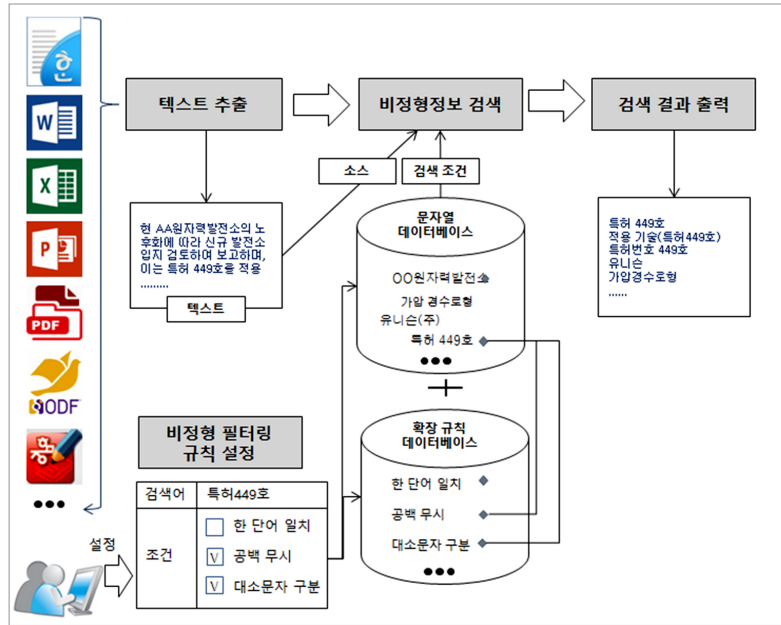
비정형 정보는 특정 조직에서 기술적, 경제적 인 의미를 갖는 정보이나 일정한 규칙과 패턴으로 검출해 낼 수 없는 문자열로 구성된 정보를 말한다. 다른 조직에서 의미 없는 낱말이 특정 조직에서는 경제적, 문화적, 기술적 영향을 갖는 의미 있는 낱말이 될 수 있다.

비정형 정보 필터링에서는 사용자가 정의한 규칙이 필터링 품질을 좌우한다. 정형 정보 설

정과 달리 비정형 정보는 검색 규칙을 사용자가 직접 정의해야 한다. 필수적으로 검색 대상 문자열이 등록되어야 하며, 필터링 도구에 따라 검색 시 확장성을 높이기 위하여 공백 문자 처리 방법, 대소문자 구분 여부, 일부 문자열만 포함하는 텍스트도 검색할지 아니면 한 단어로 정확히 일치해야 할지 등의 규칙을 추가적으로 정의한다. 사용자가 정의한 규칙을 기반으로 검색하는 비정형 정보 필터링 프로세스를 도식화하면 〈그림 4〉와 같다.

비정형 정보 필터링 프로세스의 첫 번째 단계는 정형 정보 필터링과 동일하게 대상 파일로부터 '텍스트 추출'하는 과정이다. 다양한 포맷 파일로부터 텍스트를 추출하는 방식은 정형 정보 필터링과 같은 방식을 취할 수 있다.

다음 단계에서는 사용자가 등록한 문자열과 확장 규칙을 조건으로 추출한 '정형 정보'를 검



〈그림 4〉 비정형 정보 필터링 프로세스

색한다. 문자열과 정확히 일치하는 문자열은 물론, 공백을 무시하고 일치하는 문자열, 해당 문자열을 포함한 문자열, 대소문자를 구분 또는 구분하지 않고 일치하는 문자열 등을 검색한다. 각 필터링 도구는 속도와 정확성을 고려하여 적합한 검색 알고리즘을 채택하여 구현한다.

마지막 '검색 결과 출력' 단계는 정형 정보 프로세스와 동일하게 검색 결과를 표시할 수 있다. 비정형 정보 필터링 프로세스와 정형 정보 필터링 프로세스의 차이점은 정형정보 검색 시 검색 조건을 어떻게 설정하여 필터링 도구를 구현하는 것의 차이라고 할 수 있다.

### 3.3 기계학습에 의한 필터링 구현 방식

비정형 정보 필터링 구현 방식에서 보았듯이 비정형 정보는 개발자가 정의한 규칙에 기반하

고 있기 때문에 필터링 결과의 품질은 규칙의 정확도에 의존한다. 그러나 이 규칙의 기본 전제는 특정 문자열인데 특정 문자열이라고 모든 맥락에서 민감정보인 것은 아니다. 새로운 사업 단지 설립을 위해 위치 선정을 검토하는 문서라면 이 문서 속의 주소 정보는 경제적 과급 효과를 가져 올 수 있으므로 노출 여부를 판단해야 하는 민감정보이다. 하지만 동일한 주소가 다른 문서에서 언급되었을 때는 의미 없는 단순한 위치 정보일 수도 있다. 문서 내용을 잘 아는 업무 담당자는 문서 제목과 첫 문단을 읽으면 주소를 민감정보인지 아닌지 빠르게 판단할 수 있으나, 자동 필터링을 위한 규칙으로 정의하기는 매우 어렵다.

반면 정형 정보는 정규표현식으로 규칙을 기술할 수 있기 때문에 모든 것을 규칙으로 표현할 수 있을까? 정규표현식의 함정은 한 문자만

달라져도 필터링의 검색 망을 빗겨갈 수 있다는 점이다. 예를 들어 계좌번호는 금융기관에서 정한 생성 규칙을 따르기 때문에 자리수가 다르고 패턴이 다르다. 은행에 따라 NNN-NNNNNN-NN-NNN 또는 NNN-NNN-NNNNNN 등의 패턴을 갖는다. 따라서 수많은 금융기관 계좌번호 패턴을 모두 정의해줘야 되며, 중간에 숫자 그룹을 분리하는 하이픈이 없는 경우까지 고려하면 계좌번호를 찾기 위한 패턴 수는 방대해진다. 업무 담당자라면 자동 검색 대신 문서 제목만으로도 포함되어 있을 민감정보를 예측하고 계좌번호를 즉시 찾아낼 수 있을 것이다. 인간에게는 쉽지만 정규표현식으로 표현할 수 없는 맥락 정보를 고려한다면 필터링 정확도가 향상될 수 있다. 만약 근접한 텍스트에 계좌주와 같은 정형 또는 비정형의 개인정보가 존재한다면, 더욱이 이 문서가 금융거래를 내용으로 한다면 일련의 연속된 숫자는 계좌번호일 가능성이 높다.

비정형이나 정형 정보 모두의 경우에 업무 담당자의 경험과 지식이 가장 정확한 필터링 결과를 가져온다면, 필터링 도구가 업무 담당자처럼 경험하고 데이터를 누적하여 무엇이 우리 조직에서 민감정보인지 발견하고 이를 규칙으로 정의할 수 있다면 가장 정확한 필터링이 될 것이다. 풍부한 규칙 정의도 어렵지만 변화하는 문맥 정보나 의미 규칙을 필터링 규칙에 적용해야 한다는 점까지 고려한다면 스스로 규칙을 점검하고 업데이트할 수 있는 자동 규칙 관리 도구가 더욱 필요하다. 이로부터 자동 규칙 정의가 가능한 필터링 도구에 대한 요구가 발생한다.

기계학습은 컴퓨터로 하여금 경험을 통하여 자동적으로 개선해 가게 하는 컴퓨터 알고리즘

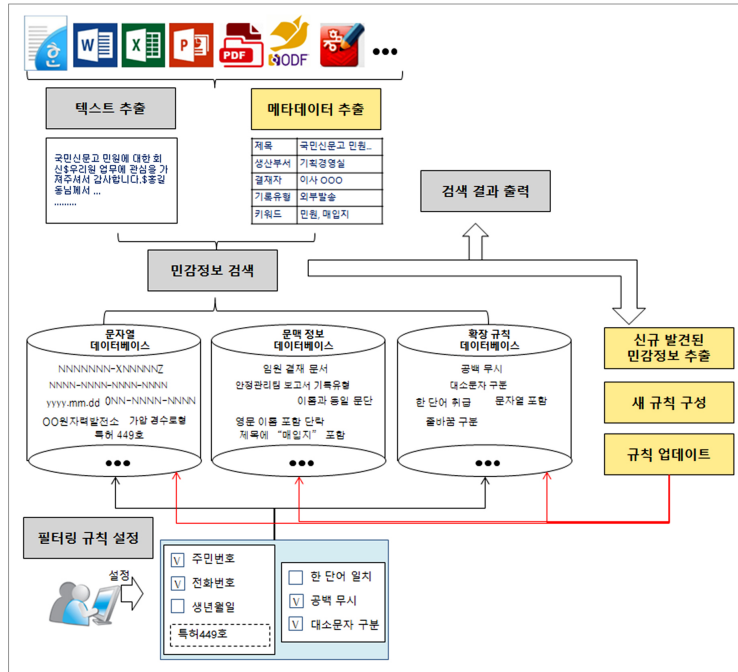
(Tom Mitchell, 1997)이다. 인공지능(AI)의 한 분야로서 명시적인 프로그램 없이 컴퓨터가 학습할 수 있도록 프로그램 하여 새로운 데이터에 대처하게 한다. 기계학습 기술을 필터링 기술에 적용함으로써 기존 규칙 데이터로부터 민감정보 검색 규칙을 찾아내서 새 규칙을 정의하고 이 규칙에 의해 다시 필터링을 실행하며 이 과정을 반복해 감으로써 규칙을 개선해가는 필터링 구현 방식을 제시하고자 한다.

기계학습 기반 민감정보 필터링 프로세스를 도식화하면 <그림 5>와 같다. 정형 정보와 비정형 정보 규칙 설정에서와 유사하게 기초 규칙 데이터베이스를 사용자가 설정하여 정규표현식 및 검색 문자열이 확장 규칙과 함께 데이터베이스를 구축한다. 이 때 문맥 정보에 관한 규칙 설정을 추가할 수도 있다.

필터링 대상이 되는 입력 파일은 다양한 포맷으로부터 텍스트를 추출하고, 문맥 정보 규칙 적용을 위해 메타데이터도 추출한다. 규칙 데이터베이스의 조건으로 텍스트와 메타데이터를 대상으로 검색을 실시한 뒤 정형 정보 및 비정형 정보와 유사하게 검색 결과를 출력한다.

기계학습 기반 필터링에서는 검색 결과로부터 필터링 도구가 자동적으로 새로운 규칙을 발견하는 프로세스가 추가된다. 최근 빠르게 발전하고 있는 기계학습, 딥 러닝(deep learning) 및 자연어 처리 기술의 많은 알고리즘을 기반으로 속도와 정확도에서 적합한 알고리즘이 연구되어야 하겠다. 새로운 규칙으로 다시 규칙 데이터베이스를 업데이트하여 규칙 개선 과정을 반복해 간다.

새 규칙 구성에서 사용자가 개입함으로써 규칙을 정규화하거나 더욱 정교화 할 수도 있을



〈그림 5〉 기계학습 기반 민감정보 필터링 프로세스

것이다. 즉, 사용자가 검색 결과를 조회하여 적용 여부를 선택하게 하고 이 결과를 새 규칙 구성의 데이터로 활용하도록 함으로써 학습 자료의 정확도를 높일 수 있다.

### 3.4 필터링 도구의 고도화 방향

원문정보공개에 대비한 필터링 도구의 고도화 방향은 필터링 도구의 활용과 관리 단계로 구분하여 볼 수 있다. 먼저, 필터링 도구의 활용 단계에서 고도화 되어야 할 부분은 정형정보와 비정형정보를 모두 거를 수 있어야 하고, 맥락 정보와 연계하여 거를 대상을 추천하여 업무담당자가 비공개 여부를 결정할 수 있어야 한다. <표 1>에서 제시된 개인정보에는 정형정보와 비정형정보가 포함되어 있다. 주민등록번호, 운

전면허번호, 전화번호, 생년월일 등은 일정한 형식을 갖춘 정보로 복잡한 알고리즘 없이 필터링 도구를 만들어 정보를 거를 수 있으며 현재 이런 기술은 상용화 되어 있다. 하지만 개인정보 중 비정형 대상정보인 범죄이력, 교육정보 등은 문맥에 따라 다를 수도 있어 이들 정보를 문맥에 따라 판단하여 비공개 여부를 설정해야 한다. 업무담당자에게는 본문 파일뿐만 아니라 첨부 파일의 내용을 일일이 확인하여 비공개 대상정보를 판단하는 것은 업무의 부담이 될 수 있다. 필터링 도구는 업무담당자의 업무효율성을 높일 수 있도록 지원해야 하는데 이를 위해 맥락정보와 연계하여 거를 대상을 추천하고, 걸러야 할 정보 목록을 가중치에 따라 제시 할 수 있는 기능이 요구된다. 필터링 도구를 통해 문서 생산 중 확인된 비공개 대상 정보는 업무담당자가

최종판단하여 공개여부를 결정해야 한다. 만약, 업무담당자의 판단이 기존의 학습된 규칙에서 어긋나는 경우 해당 사례들을 제시하여 정보공개를 통한 비공개 정보 유출을 줄여야 한다.

다음으로, 필터링 도구의 관리 단계에서 고도화 되어야 할 기능은 새로운 규칙을 학습하여 적용할 수 있는 있어야 한다. 원문정보공개 대상 기관의 민감정보는 기관의 특성에 따라 그 유형이 다양하며, 특정하여 관리하기 어렵다. 공공기관의 경우 「공공기록물 관리에 관한 법률」 제35조(기록물의 공개여부 분류)에 따라 비공개로 재분류된 기록물에 대해 재분류된 연도부터 매 5년마다 공개 여부를 재분류하여 한다. 공개재분류 과정상에는 기준서를 작성해야 하는데 기준서에는 비공개 기록물의 내용정보를 정리해야 한다. 비공개 정보는 기관의 업무 성격에 따라 차이가 있으며 기준서에 제시되는 비공개 정보 목록은 바로 각 기관의 민감정보로 활용할 수 있다. 필터링 도구는 각 기관의 기준

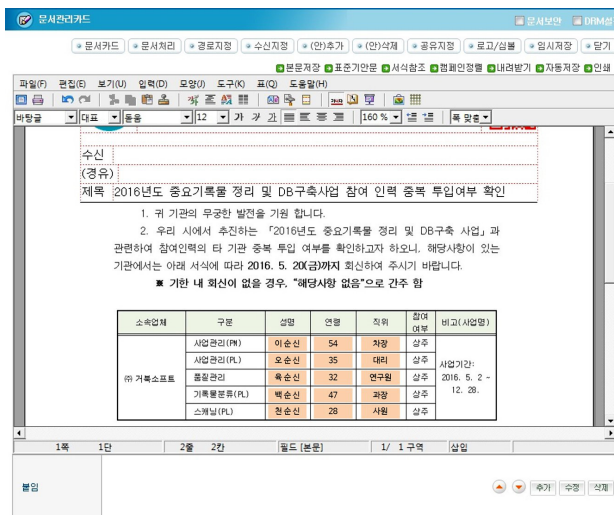
서에 정리된 비공개 정보 목록을 필터링 도구에 입력하여 새로운 규칙을 학습할 수 있어야 하며, 이를 통해 각 기관의 맞춤형 필터링 도구 활용이 가능해 질 수 있다.

#### 4. 원문정보공개 프로세스의 재설계

##### 4.1 원문정보공개에 필요한 도구들

원문정보공개를 위해 민감정보를 필터링 하는 도구 외에도 민감정보를 표시하는 마킹 도구와 정보공개 서비스용 사본 생성을 위한 마스킹 도구가 필요하다.

마킹 도구는 업무담당자가 문서 편집기를 활용한 문서 생산이나 결재 과정 중에 민감정보를 표시하는 도구를 말한다. 업무담당자는 문서편집기에서 문서 작성 중에 민감정보를 입력한 경우 <그림 6>과 같이 민감정보를 마킹 표기를 한



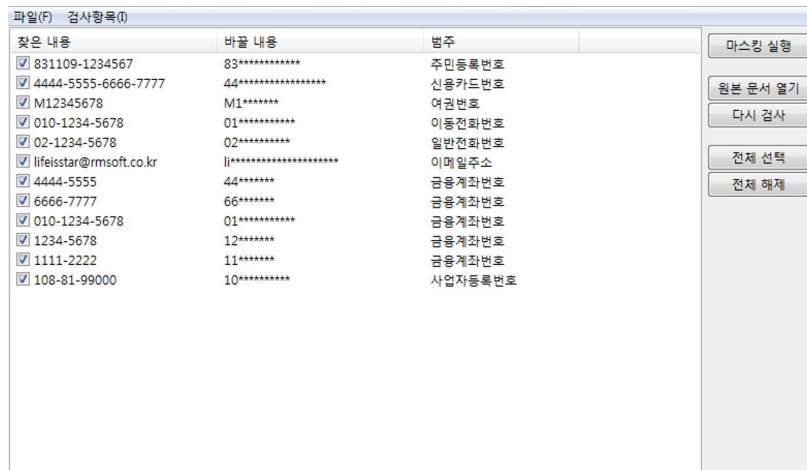
<그림 6> 업무관리시스템의 문서편집기 상에 마킹 기능을 구현한 화면 예시

다. 마킹 도구는 민감정보를 문서상에 표기하여 사용자에게 알려주는 알람 도구일 뿐만 아니라 양질의 추가 학습자료를 생산한다. 또한, 마킹한 정보들이 컴포넌트의 메타데이터로 등록하고, 관리하여 기관의 민감정보 관리 도구로 활용이 가능하다.

마스킹 도구는 정보공개 서비스용 사본 생성을 위해 민감정보를 비식별 화하기 위한 도구이다. 기관에서 관리하는 민감정보를 마스킹 도구에 검사항목으로 등록하여 마스킹할 문서를 불러오기 하여 마스킹 도구를 구동한다. <그림 7>과 같이 문서에 작성된 민감정보를 마스킹 도구를 활용하여 비식별 정보로 변경이 필요한 내용을 선택하고 '마스킹 실행'을 수행한다. 마스킹 수행이 완료되면 정보공개 서비스용 사본이 생성되고, 업무담당자는 생성된 사본을 최종 검토한다.

#### 4.2 결재문서 본문 작성 프로세스

결재문서 생산 과정에 필터링 도구 활용은 크게 본문 작성 과정, 붙임파일 등록 과정, 결재 과정으로 구분된다. 업무담당자는 문서 작성 과정 중 업무관리시스템의 문서편집기에서 <그림 3>과 같이 본문 작성 중에 민감정보에 한하여 기본적인 마킹(붉은색 음영으로 표기)을 한다. 본문 작성이 완료(저장)되면 필터링 도구가 구동<sup>8)</sup> 된다. 업무담당자가 마킹한 내용과 필터링 도구가 거른 내용이 팝업 창에 목록으로 구분되어 <그림 8>과 같은 형태로 제시된다. 이 때, 마킹한 내용이나 필터링된 내용이 왜 공개유예 대상인지 '민감정보 유형', '정보공개법 제9조 제1항의 '비공개 근거', '비공개 대상정보'가 추천 값으로 제시된다. 업무담당자는 마킹 항목별로 공개유예 사유와 호수를 수정할 수 있다. 임



<그림 7> 마스킹 도구의 기능이 구현 화면

8) 필터링 도구 구동방식은 저장 시 자동으로 구동하게 할 것인지, 업무담당자가 수동으로 구동시킬 것인지를 업무담당자별 옵션으로 선택할 수 있어야 한다. 또한, 필터링 도구가 이미 마킹된 정보는 스킵할 것인지도 선택사항이며, 필터링 도구 작동 초기에 어떤 유형의 정보들을 대상으로 거를 것인지도 선택사항이어야 한다.

구분	파일명	민감정보 유형	페이지	확인	비공개 근거	비공개 대상정보	필터링 유무
본문	2016년도 중요기록물 정리 및 DB구축사업...hwp	(정형) 개인정보	1p	보기	제6호	이순신, 오순신, 육순신...	Y
첨부	01_중요기록물 정리 및 DB구축사업 업체 정보.hwp	(비정형) 영업상 내부관리자료	2p	보기	제7호	(주)거북소프트 영업이익	Y
	02_참여 인력 개인정보.xlsx	(정형) 개인정보	1p	보기	제6호	640702-1234567, ...	Y

〈그림 8〉 민감정보 필터링 이력 조회

력을 마치면 필터링 도구를 통해 확인한 민감정보를 본문에 적용시킬 것인지 ‘필터링 유무’를 체크하고 팝업창을 닫는다. 팝업 창을 닫으면 이 내용이 해당 컴포넌트의 메타데이터로 저장되어 관리된다. 다시 본문 파일 편집기에서 필터링 도구를 통해 확인한 정보들이 마킹된 상태로 보인다. 업무담당자는 육안으로 민감정보를 재확인 후 본문 파일을 저장한다. 만약, 마킹된 내용 중 오류가 있다면 마킹을 해제할 수 있으며, 저장하는 순간 컴포넌트의 메타데이터들도 수정되어 저장된다.

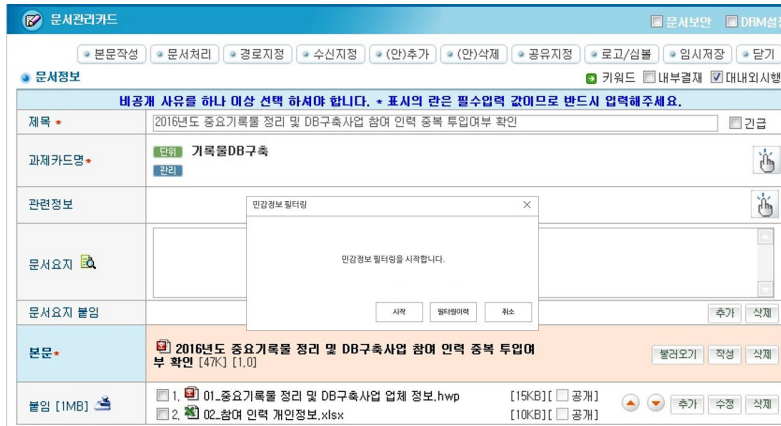
붙임파일 등록 과정 중에 필터링 도구는 붙임 파일을 업로드할 때마다 〈그림 9〉와 같은 형태로 구동된다. 또한, 필터링 도구가 거른 내용은 〈그림 8〉과 같은 팝업 창 목록으로 뜬다. 본문 작성 과정에 필터링 도구 구동 결과를 수정한 것과 마찬가지로 공개유예 대상과 정보를 수정할 수 있고, 입력 후 이 내용들이 해당 컴포넌트의 메타데이터로 저장된다.

결재 경로에서 중간 결재자와 최종 결재자는 컴포넌트별 공개 구분 값과 사유 등을 살펴보면 문서 내용뿐만 아니라 민감정보를 함께 검토한다. 만약, 검토과정 중에 문서를 생산한 업무담당자와 의견이 다르다면 필터링 도구를 구동하여 민감정보를 수정하여 문서를 제출할 수 있다. 컴포넌트의 메타데이터는 결재 경로의 마지막

결재자까지 추가 기입에 의해 완성된다.

결재가 완료되면 공개여부 값 등도 함께 확정된다. 해당 결재 건에서 처음으로 마킹된 유형의 정보는 기록관에 통보된다. 정보공개담당자가 마킹정보를 검토하고 승인여부를 결정한다. 마킹된 정보가 공개유예될 수 있는 대상으로 판단되면 기록관리기준표와 비공개세부기준에 반영하여 관리한다. 만약, 마킹한 정보가 공개유예 대상이 아니라고 판단되면 결재문서 생산 처리과와 협의를 거쳐 공개유예 여부를 최종확정한다.

필터링 도구를 통해 관리되는 공개유예 대상은 〈그림 10〉~〈그림 11〉과 같이 정형, 비정형 민감정보 구분되어 관리된다. 일반적인 민감정보는 필터링 도구 도입 시 기본 값이 입력되어 있으며, 업무담당자들이 필터링 도구를 활용하여 축적한 민감정보들이 기관의 특수한 민감정보로 저장된다. 정형 민감정보는 특정 형식을 가지는 민감정보로 일정한 패턴을 민감정보 형식에 입력하여 필터링이 가능하다. 이에 반해 비정형 민감정보는 각 기관에 따라 등록되는 정보가 상이하며 특정 패턴을 가지지 않는 특성이 있어 지속적인 관리가 필요하다. 각 기관의 공개유예 대상 목록의 관리는 기록물의 공개재분류 작업 시 활용할 수 있으며, 이들 정보를 활용한 공개재분류 솔루션 개발도 가능할 것으로 판단된다.



<그림 9> 붙임 파일 첨부 후 필터링 도구 구동

선택	민감정보 유형	민감정보 형식	비공개 근거	등록(수정)일
<input type="checkbox"/>	주민등록번호	#####-#####	제6호	2017.01.01.
<input type="checkbox"/>	운전면허번호	##-##-#####-##	제6호	2017.01.01.
<input type="checkbox"/>	휴대전화	01#-#####	제6호	2017.01.01.
<input checked="" type="checkbox"/>	신용카드번호	#####-#####-#####	제6호	2017.01.30.
⋮	⋮	⋮	⋮	⋮

<그림 10> 정형 민감정보 관리

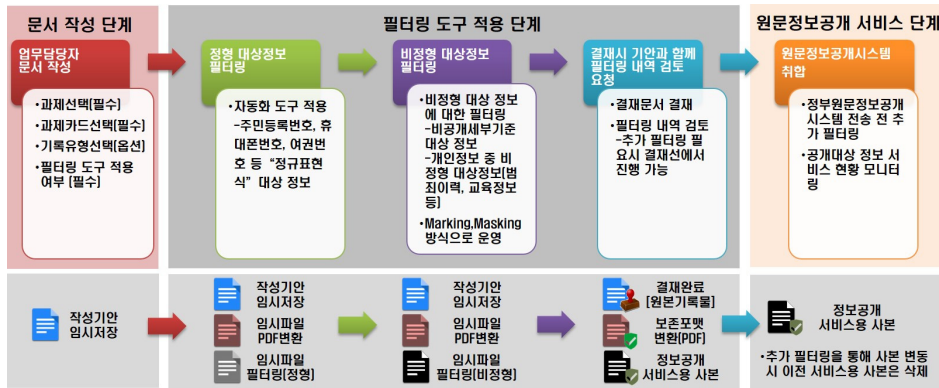
선택	민감정보 유형	민감정보 용어	비공개 근거	등록(수정)일
<input type="checkbox"/>	안보 관련 정보	총무계획, 물자계획, 전사자랑...	제2호	2017.01.01.
<input checked="" type="checkbox"/>	재판 및 수사 관련 정보	심판청구서, 답변서 제출, 변호기일...	제4호	2017.01.03.
<input type="checkbox"/>	계약관리	유자계약명단, 설계서, 원가계산서...	제5호	2017.01.05.
<input type="checkbox"/>	중소기업	#원, 신용등급, 지원금액...	제7호	2017.01.25.
⋮	⋮	⋮	⋮	⋮

<그림 11> 비정형 민감정보 관리

### 4.3 원문정보공개 프로세스

원문정보공개 서비스를 고려한 문서 기안의 전체적인 프로세스를 재설계하면 <그림 12>와 같다. 업무담당자는 업무관리시스템을 활용하여 기안문을 작성하기 위해 과제와 과제카드를 선택한 후 기안문을 작성한다. 기안문 작성이 마치면 작성한 기안문이 임시저장되고, 기안문에 대한 필터링 도구가 적용된다. 필터링 도구 적용 단계는 정형 대상정보 필터링, 비정형 대상정보 필터링 후 업무담당자에게 필터링 내역 검토 요청이 이루어진다. 정형 대상정보 필터링 과정에서는 자동화된 도구가 적용되어 주민

등록번호, 휴대폰번호, 여권번호 등 정규표현식 대상 정보가 필터링 되고, 각 기관의 비공개세 부기준 대상 정보와 개인정보 중 비정형 대상 정보가 필터링 된다. 필터링된 내역은 업무담당자에게 제시되고, 업무담당자는 필터링 내역을 검토한 후 비공개 대상정보를 선택해 준다. 업무담당자의 필터링 내역 검토가 완료되면 결재완료 원문 기록물이 생산되고, 보존 포맷 변환, 정보공개서비스용 사본 생성이 이루어진다. 원문정보공개시스템으로 문서를 전송하기 전에 추가적인 필터링 도구 구동으로 불필요한 정보가 공개되는 것을 막는다.



〈그림 12〉 원문정보공개 프로세스 재설계

### 5. 맺음말

2017년 1월 현재까지 12,179개 공공기관에서 원문정보공개 서비스를 진행하고 있으며, 하루에도 수만건의 원문정보가 공개되고 있다. 현재 일부 공공기관에서는 개인정보 필터링을 위해 솔루션을 도입하여 활용하고 있으나 하루에도 수만건 이상 생산되는 전자결재문서를 개인정보 필터링만으로 공공기관에 따라 비공개정보를 선별하여 공개 할 수 없는 실정이다. 이것은 공공기관에서 사용중인 생산시스템이 생산부분만을 고려하여 설계되어 공개를 대비한 시스템 구축이 되지 않았다는 것에 그 원인이 있다. 본 연구에서는 현재까지 개발된 필터링 도구의 한계점을 살펴보고 고도화된 필터링 도구가 갖추어야 할 기능요건을 분석해 보았다. 또한, 필터링 도구를 효과적으로 활용하기 위한 프로세스를 재설계하여 각 공공기관에서 원문정보공개 서비스를 위한 업무를 보다 효율적으로 할 수 있는 민감정보 필터링 요건에 대해 연구하였다. 연구를 통해 확인한 민감정보 필터링 요건은 다음과 같다.

기계학습 방식의 필터링 도구를 개발하여 공개유예 할 정보를 각 기관 맞춤형으로 관리해야 한다. 현재 상용화되어 판매되고 있는 개인정보 필터링 도구는 정형화된 패턴의 개인정보만을 걸러낼 수 있어 기관의 정보공개 업무담당자는 원문정보공개 서비스를 위해 민감정보를 직접 확인해야 하는 업무 부담을 가지고 있다. 기계 학습 방식의 필터링 도구는 업무담당자가 문서 생산이나 결재 과정상에 민감정보를 마킹하여 필터링 내역으로 관리할 수 있어 기관 맞춤형 필터링 도구로 할 수 있다. 기계학 방식의 필터링 도구의 개발은 반복적인 민감정보 확인 작업을 최소화 할 수 있고, 각 기관에서 공개 유예되어야 하는 정보를 사전에 알려주어 원문정보공개 과정에서 비공개 대상정보를 사전에 차단하는데 효과성이 높다.

원문정보공개 지원을 위해 문서 생산단계에서부터 원문정보공개 서비스 단계까지 원문정보공개를 위한 프로세스 상에 민감정보 필터링 도구의 활용이 요구된다. 업무담당자가 문서 생산시 민감정보를 마킹 한 후 필터링 도구를 활용하여 1차적으로 민감정보를 걸러내고, 정형·비

정형 정보를 내역으로 관리되어야 한다. 문서 결재 과정상에는 문서 내용만의 검토뿐만 아니라 필터링 내역 검토와 추가 또는 수정해야 하는 민감정보를 2차적으로 걸러내어 민감정보 유출을 최소화해야 한다. 원문정보공개 서비스 단계에서는 결재완료 원문 기록물이 생산되고, 보존 포맷 변환, 정보공개서비스용 사본 생성이 민감정보 필터링 도구를 활용하여 이루어져야 한다. 원문정보공개 프로세스 상에 민감정보 필터링 도구를 활용하여 문서 생산과 결재 과정상에 업무담당자가 문서 내용뿐만 아니라 공개유예 대상 정보를 함께 검토함으로써 불필요한 정보가 공개되는 것을 사전에 효과적으로 막을 수 있다.

민감정보 필터링 내역 관리를 통해 공공기관 별 민감정보를 유형을 파악할 수 있고, 이를 활용하여 공개재분류 업무상에 적용할 수 있다. 민감정보 유형에 따라 공통 민감정보, 고유 민감정보로 구분할 수 있으며, 공통 민감정보는 타 공공기관에서도 관리되는 민감정보를 말하고, 고유 민감정보는 개별 기관에 특별히 관리되는 민감정

보를 말한다. 민감정보 필터링 내역에는 업무담당자가 문서 생산이나 결재과정 중에 등록한 민감정보 내역이 포함되어 있으며 이것은 각 기관의 공개유예 대상 정보 목록으로 볼 수 있다. 이 민감정보 내역은 민감정보 필터링 도구 활용시 사용할 수 있고, 5년 주기의 공개재분류 과정에서도 활용 할 수 있다. 추후에는 민감정보 내역을 활용한 공개재분류 솔루션을 개발하여 업무를 보다 효율적으로 수행할 수 있기를 기대한다.

원문정보공개 지원을 위한 민감정보 필터링 도구의 개발은 비공개 대상 정보의 유출을 사전에 차단하여 보다 효과적인 정보공개 업무를 수행할 수 있다는데 의의가 있다. 또한 각 기관의 민감정보 유형의 관리를 통해 공개재분류 업무와 같은 기록관리 업무에도 활용할 수 있다. 공공기관의 설명책임성 실현을 위해 민감정보를 보다 체계적으로 관리 할 필요가 있는 것이 이런 이유들 때문이며, 기록관리분야 산업상의 변화를 일으켜 제4차 산업혁명시대를 대비하는 것이 앞으로의 과제이다.

## 참 고 문 헌

- 국가기록원 (2006). 검색일자: 2017. 1. 22. [www.archives.go.kr](http://www.archives.go.kr)
- 국가기록원 (2016). 2016년 기록물 관리지침. 대전: 국가기록원.
- 나라장터 (2016). 검색일자: 2017. 1. 22. [shopping.g2b.go.kr](http://shopping.g2b.go.kr)
- 정보공개포털 (2015). 검색일자: 2017. 1. 22. [www.open.go.kr](http://www.open.go.kr)
- 정보소통광장 (2015). 검색일자: 2017. 1. 22. [opengov.seoul.go.kr](http://opengov.seoul.go.kr)
- 정연수, 김미현, 최윤정 (2007). 2007년 개인정보분쟁조정 사례집. 서울: 한국정보보호진흥원, 개인정보분쟁조정위원회.
- 행정자치부 (2015). 개인정보보호 지침. 서울: 행정자치부.
- 행정자치부 (2016). 2015년 정보공개 연차보고서. 서울: 행정자치부

- 허준석, 홍덕용 (2016). 원문정보공개서비스를 위한 전자기록구성요소에 관한 연구. 기록학연구, 50, 351-388.
- 황진현, 임진희 (2015). 체계적인 원문정보공개를 위한 정부산하공공기관 기록관리시스템 기능요건 연구. 기록학연구, 46, 325-363.
- Tom Mitchell (1997). Machine Learning. McGraw hill.

[ 관련법규 ]

- 「공공기록물 관리에 관한 법률」 (법률 제12844호).
- 「개인정보 보호법」 (법률 제14107호).
- 「전자서명법」 (법률 제12762호).
- 「정보통신망 이용촉진 및 정보보호 등에 관한 법률」 (법률 제14080호).

• 국문 참고자료의 영어 표기

(English translation / romanization of references originally written in Korean)

- Freedom of Information Portal (2015). Retrieved January 22, 2017, from [www.open.go.kr](http://www.open.go.kr)
- Heo, J. S. & Hong, D. Y. (2016). A Study on the Electronic Records Element for Full Text Opening Information Service. Korean Society Of Archival Studies, 50, 351-388.
- Hwang, J. H. & Yim, J. H. (2015). A Study on Functional Requirements of Records Management Systems for Government-Affiliated Organizations to respond to Immediate Disclosure of Original Official Documents. Korean Society Of Archival Studies, 46, 325-363.
- Jung, Y. S., Kim, M. H., & Choi, Y. J. (2007). Personal Information Dispute Settlement Casebook 2007. Seoul: Korea Information Security Agency, Personal Information Dispute Resolution Committee.
- Korea Public Procurement (2016). Retrieved January 22, 2017, from [shopping.g2b.go.kr](http://shopping.g2b.go.kr)
- Ministry of the Interior (2015). Privacy Guidelines. Seoul: Ministry of the Interior.
- Ministry of the Interior (2016). Annual Report on Public Disclosure in 2015. Seoul: Ministry of the Interior.
- National Archives of Korea (2006). Retrieved January 22, 2017, from [www.archives.go.kr](http://www.archives.go.kr)
- National Archives of Korea (2016). Guidelines for Archives Management 2016. Daejeon: National Archives of Korea.
- Seoul Information Communication Plaza (2015). Retrieved January 22, 2017, from [opengov.seoul.go.kr](http://opengov.seoul.go.kr)

