

개인정보 유출 걱정 없이 안전하게 전자문서 개방공유한다

- KISTI NTIS센터, PDF에 포함된 개인정보만 제거하는 전자문서 개인정보보호 기술 개발 -
- 현재까지 총 25개 공공기관에 무상으로 기술 이전 -

- 개인정보 유출 없이 안전하게 전자문서를 개방·공유할 수 있는 길이 열렸다.
- 국가R&D 보고서 원문 성과물 전달관리기관인 한국과학기술정보연구원(원장 한선화, 이하 KISTI)이 전자문서(PDF)에 포함된 개인정보만을 자동 식별하여 제거하는 기술을 개발하였다.
 - KISTI NTIS센터 R&D시스템개발실에서는 전자문서 개인정보보호 기술을 개발하고 2014년 9월 “개인정보를 보호하는 장치, 개인정보를 보호하는 방법 및 개인정보를 보호하는 프로그램”을 저장하는 저장매체”로 특허 출원을 했다.
- 기존에는 전자문서 내 개인정보 유출을 방지하기 위해 대부분 전자문서를 암호화하는 기술을 사용하고 있다. 그리고 전자문서가 유통되는 과정에서 개인정보가 탐지될 경우 해당 전자문서 자체를 삭제하여 개인정보 유출을 방지하고 있다.
 - 그러나 전자문서 전체가 배포되거나 열람되지 못할 경우 ‘정부 3.0 공공데이터 개방 정책’에 따른 정보 공유와 개방이 어렵다는 문제점이 발생했다.
- KISTI가 이번에 개발한 기술은 기존의 문서를 개인정보 검출 프로그램으로 처리해 주민등록번호, 이메일, 전화번호, 주소, 금융정보 등의 개인정보가 선별적으로 제거된 결과물을 생성한다.

- 1 -

- 전자문서 형태의 정보 공개가 필요한 기관의 전자문서 개인정보보호는 물론 관련 업무의 효율성 제고에 크게 기여할 수 있을 것으로 기대된다.
- 지난해 한국연구재단 등 19개 기관에 이어 올해 4월까지 한국기초과학지원연구원 등 6개 기관에 무상으로 기술이전하였다.
- KISTI는 오는 9월 아래아 한글 문서에 대한 개인정보보호 기술 개발을 완료하고 기술이전을 시작할 계획이다.
 - 또한 앞으로 마이크로소프트 워드(MS-Word), ODF, XML 등의 포맷에 대한 기술 개발도 이어나갈 예정이다.(끝)



그림 1 개인정보검출 프로그램을 실행한 결과

연구 결과 개요

전자문서 개인정보보호 기술

(출원명 : 개인정보를 보호하는 장치, 개인정보를 보호하는 방법 및 개인정보를 보호하는 프로그램을 저장하는 저장매체, 출원 번호 : 10-2014-0117439, 출원 일자 : 2014.09.04.)

□ 연구의 차별성

최근 개인정보 유출 사고로 인해 개인적 또는 산업적 피해가 증가함에 따라 개인정보 유출 방지를 위해 데이터 유출 방지 기술(DLP: Data Loss Prevention)을 많이 사용하고 있다. 이러한 데이터 유출 방지 기술의 경우에는 네트워크를 통해 개인정보가 유출될 수 있는 모든 서비스를 감시하고, 개인정보 유출을 방지하기 위해 전자문서를 암호화하여 보호하는 기술을 사용하고 있다. 또한 기존의 전자문서 개인정보보호 기술의 경우에는 전자문서에서 개인정보가 탐지될 경우 해당 전자문서 자체를 삭제한다. 그러나 전자문서 전체가 배포되거나 열람되지 못하도록 하는 경우, 정보화 시대에 정보를 공유하고 전달하기가 어려운 문제점이 발생한다. 예를 들어 전자문서 암호화를 이용하여 전자문서를 암호화하는 방식이 사용될 수 있지만, 이러한 경우 개인정보 보호를 위해 다른 정보의 공유와 전파를 막게 되는 문제점이 있다.

본 기술에서는 전자문서에서 개인정보인 주민등록번호, 이메일, 전화번호, 주소, 금융정보 등을 (1) 정규표현식(Regular Express)을 이용하여 검출하고 각 개인정보 항목에 대한 유효성을 검증하여 (2) 개인정보만을 제거하거나 마스킹 하여 전자문서를 개인정보 유출 없이 안전하게 개방·공유할 수 있다는 점에서 큰 의미가 있다.

□ 연구의 활용성

본 연구를 통해 전자문서에 포함된 개인정보만을 선별적으로 제거함으로써 (1) 전자문서 형태의 국가연구개발 보고서를 연구자 개인정보 유출 없이 공유할 수 있으며 (2) 정부 3.0 공공데이터 개방 정책에 따른 전자문서 개방에 따른 개인정보 유출을 방지 할 수 있다. 그러므로 개인정보 검증 시 기존의 수작업에 의존하였던

- 3 -

보고서 등의 전자문서에 기재된 개인정보를 자동으로 검출하고 삭제할 수 있다. 그리고 전자문서의 관리기관의 관리자도 쉽게 사용할 수 있고 향후 공개되는 전자문서의 개인정보 강화에 기여할 수 있다.

전자문서 개인정보보호 기술은 개인정보만을 오류 없이 식별하고 제거할 수 있는 기반 기술로서 현재 국가R&D보고서원문 성과물 전달관리기관인 한국과학기술정보연구원에서 국가연구개발사업 보고서원문 DB 구축 시 개인정보 제거에 활용되고 있으며, 2014년 11월 과제관리(전문)기관인 한국연구재단 등 19개 기관, 2015년 3월 한국기초과학지원연구원 등 6개 기관에 무상 기술 이전하여 과제관리(전문)기관의 전자문서 개인정보보호체계 구축에 활용되고 있다. 그러므로 본 연구결과는 전자문서 형태의 정보 공개가 필요한 기관의 전자문서 개인정보보호 기반 기술로 적합하다.

□ 실용화를 위한 과제

상용화를 위해서는 모든 전자문서 포맷에 대한 보호가 보장되어야 한다. 이를 위해 현재 PDF, HWP 포맷을 확장하여 MS-WORD, ODF, XML 포맷에 대한 개발이 필요하다.

□ 기 타

1. 한국과학기술정보연구원 주요사업의 지원으로 수행중인 “국가R&D 성과물 공유·확산”으로 진행된 연구결과이다.

- 4 -