

재난정보 표준화를 통한 환경 재난정보 수집 및 활용

Collection and Utilization of Unstructured Environmental Disaster by Using Disaster Information Standardization

이동섭^{1*} · 김병식²¹(주)피에스글로벌 대표이사, ²강원대학교 방재전문대학원 도시환경방재학과 교수Dong Seop Lee^{1*} and Byung Sik Kim²¹CEO, PSGlobal.Co,LTD²Professor, Department of Urban Environment & Disaster Management, Kangwon National University, Samcheok 25913, Korea

Received 18 November 2019, revised 26 November 2019, accepted 5 December 2019, published online 31 December 2019

ABSTRACT: In this study, we developed the system that can collect and store environmental disaster data into the database and use it for environmental disaster management by converting structured and unstructured documents such as images into electronic documents. In the 4th Industrial Revolution, various intelligent technologies have been developed in many fields. Environmental disaster information is one of important elements of disaster cycle. Environment disaster information management refers to the act of managing and processing electronic data about disaster cycle. However, these information are mainly managed in the structured and unstructured form of reports. It is necessary to manage unstructured data for disaster information. In this paper, the intelligent generation approach is used to convert handout into electronic documents. Following that, the converted disaster data is organized into the disaster code system as disaster information. Those data are stored into the disaster database system. These converted structured data is managed in a standardized disaster information form connected with the disaster code system. The disaster code system is covered that the structured information is stored and retrieve on entire disaster cycle. The expected effect of this research will be able to apply it to smart environmental disaster management and decision making by combining artificial intelligence technologies and historical big data.

KEYWORDS: Disaster code, Environmental disaster, Environmental disaster management, Standardization

요 약: 본 연구는 재난정보 표준화를 목표로 다양한 재난 및 사고에서 생성되는 정형, 비정형 문서를 전자화된 문서로 변환하여 환경재난 정보를 생성함으로써 데이터베이스에 저장하고 환경재난 관리에 활용할 수 있는 체계를 개발하였다. 최근 4차 산업 시대를 통해 다양한 지능화 기술들이 발전하고 있다. 이러한 기술들은 환경재난, 재난관리 등의 분야에 다양한 형태로 적용되며, 환경재난 관리 업무와 융합되어 활용되고 있다. 재난정보관리는 단순히 재난업무를 지원하는 것이 아니라 과거의 환경재난 이력정보를 활용하여 인공지능 기술을 적용한 스마트 재난관리를 지원할 수 있도록 한다. 환경재난을 관리하는 중요한 요소는 재난정보이다. 재난정보는 재난의 발생에서 진행, 대응 및 계획까지의 재난 전주기에 대한 정보를 전자화된 정보로 관리하고 처리하는 행위를 의미한다. 그러나, 자연, 사회, 환경재난에 대한 상황, 대응, 대비, 복구의 정보는 주로 보고서의 형태인 핸드아웃이나 비정형 정보로 존재하고 있다. 이러한 비정형 재난안전정보는 관리 부실에 의해 사라지거나 폐기되는 경우가 많다. 이에 따라 비정형 재난안전정보는 재난정보로서의 관리를 위해 인식기술이 필요하다. 본 연구에서는 지능화된 기술을 활용하여 인쇄되거나 스캐너에 의해 이미지 또는 문서로 생성된 재난 보고서를 전자화된 문서로 변환하는 것에 중점을 두었으며, 그 후 변환된 재난정보는 재난정보관리 코드체계에 맞추어 정리하여

*Corresponding author: dslee@psglobal.co.kr, ORCID 0000-0001-7755-5598

© Korean Society of Ecology and Infrastructure Engineering. All rights reserved.

This is an open-access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

재난정보관리 데이터베이스에 저장한다. 정형, 비정형 재난정보를 생성하는 것은 스마트 재난관리의 시작으로 가장 중요한 요소이며, 이렇게 생성된 환경재난 정보는 재난정보 코드체계와 연계하여 표준화된 형식으로 관리한다. 재난코드체계는 재난 별 발생 진행 상황, 피해 규모, 대처사항 등의 정보를 저장할 수 있는 표준을 구축하였으며, 향후 이러한 많은 재난 데이터와 이력정보를 기반으로 한 인공지능 기술을 접목하여 스마트 재난관리 및 의사결정에 활용될 수 있을 것으로 판단된다.

핵심어: 표준화, 환경재난, 재난코드, 환경재난 관리

1. 서론

최근에 지진, 화재, 미세먼지, 적수사고 등 다양한 환경재난관련 사고가 많이 발생하고 있다. 이러한 환경사고의 위험요소로부터 국민이 안심할 수 있는 환경재난에 대해 다양한 분야에서 연구가 활발히 진행되고 있다. 특히 4차 산업혁명 시대의 사물인터넷, 클라우드, 빅데이터, 인공지능 등의 분야가 환경 및 재난 분야와의 융복합을 통해 정보를 생산하고 관리하여 생명과 직접적으로 관련된 의미 있는 정보를 생산할 수 있는 체계적인 관리에 대한 관심이 커지고 있다. 해외 선진국에서는 환경재난 정보가 관리측면에서 운영 및 활용성의 중요성이 증가되고 있으며, 세계적으로도 공공 데이터(재난, 사고)의 개방, 공유 활용을 위해 노력하고 있다. 국내의 정보관리 측면에서는 환경재난 업무와 정보의 불일치성이 존재한다. 이에 따라 재난업무 수행 시 각 담당자의 경험과 노하우, 정보의 이력과 재난 및 사고관리에 필요한 정보를 취합하고 과거 재난 및 사고 상황에 대한 정보를 축적하고 활용하는 체계와 시스템 구축이 필요성이 강조되고 있다(Lee et al. 2018). 특히 환경재난 정보는 규격화된 형태 또는 통일된 형태의 구조로 관리되지 않았기 때문에 정보의 추적 및 관리가 미흡하여 효율적인 업무를 수행할 수 없었다. 특히 본 연구에서는 이러한 재난 및 사고정보에 대한 전주기적 관점에서 가장 중요한 정보의 생성, 수집 부분 중 비정형 데이터를 활용한 환경재난 정보의 생산에 대한 효율적인 방안을 제시하고자 하고 있다. 최근 재난 및 환경사고의 보고서나 서류 등에 포함되어 있는 정보가 중요함에 따라 기존에 종이나 이미지의 형태로 관리하던 문서를 관리하기 위해 문서에 대한 전자화 시스템이 도입되었다. 그러나 시스템으로 관리되지 않는 비정형 데이터는 담당자 자리에 존재하거나 관리 부실로 인해 사라지는 경우가 종종 발생되고 있다. 이러한 비정형 데이터는 과거에는 담당자가 일일이 수작업으로 데이터화시켜야 하는 번거로움이 있으므로, 이러한 것을 인공지능 기술

을 활용하여 자동화된 방법으로 비정형 데이터를 정형 데이터로 변환하는 작업을 통해 의미 있는 정보로 변환하여 정보로서의 가치를 확보해야 한다. 본 연구에서는 환경재난에 대한 정보로 인식하기 위해 비정형화된 정보를 전자화된 문서로 변환하고, 변환된 광학문자를 Lee et al. (2018)에서 제시하는 재난안전 코드체계로 변환하여 데이터베이스에 저장하는 방안을 제시하였다. 향후 비정형 정보 자동화 생산기술과 데이터베이스화된 환경재난 정보를 활용할 수 있는 방법을 제시하여 환경사고 및 재난 관리의 예방, 대비, 대응 및 복구 업무 지원을 위한 정보를 우선적으로 제공하는 기술로 활용하며, 향후 인공지능을 활용한 환경재난 의사결정지원 체계의 기반이 될 것으로 판단한다.

2. 연구 방법

본 연구는 환경재난 정보에 대한 관리를 위해 정형, 비정형 환경재난 정보를 생산하고 환경재난 표준화에 맞춰 데이터베이스에 저장한다. 또한, 재난관련 업무 담당자의 업무를 지원하기 위해 환경재난 정보의 활용에 대한 기법과 방법을 제시하는 것이다. 재난 및 사고와 관련된 비정형 정보는 문자인식을 통해 정형화시키고, 정형화된 정보는 환경재난 데이터베이스로 저장하기 위해 환경재난 관리 코드체계를 적용하였다. 우선적으로 환경재난정보를 수집하였고, 환경 재난안전 데이터를 전자화할 수 있는 기법을 적용하여 재난안전 코드로 변환할 수 있는 방법을 제시하였다.

2.1 정형 및 비정형 환경재난 정보의 수집

이미지로 표현되는 문자와 같은 비정형 환경재난 정보를 수집하기 위하여 광학문자인식은 인공지능을 활용한 응용 분야 중 패턴 인식 분야에서 활발하게 연구되고 있는 분야 중 하나이다. 기본적으로 광학문자인식은 이미지에서 텍스트 내용을 추출할 수 있도록 개발된 기

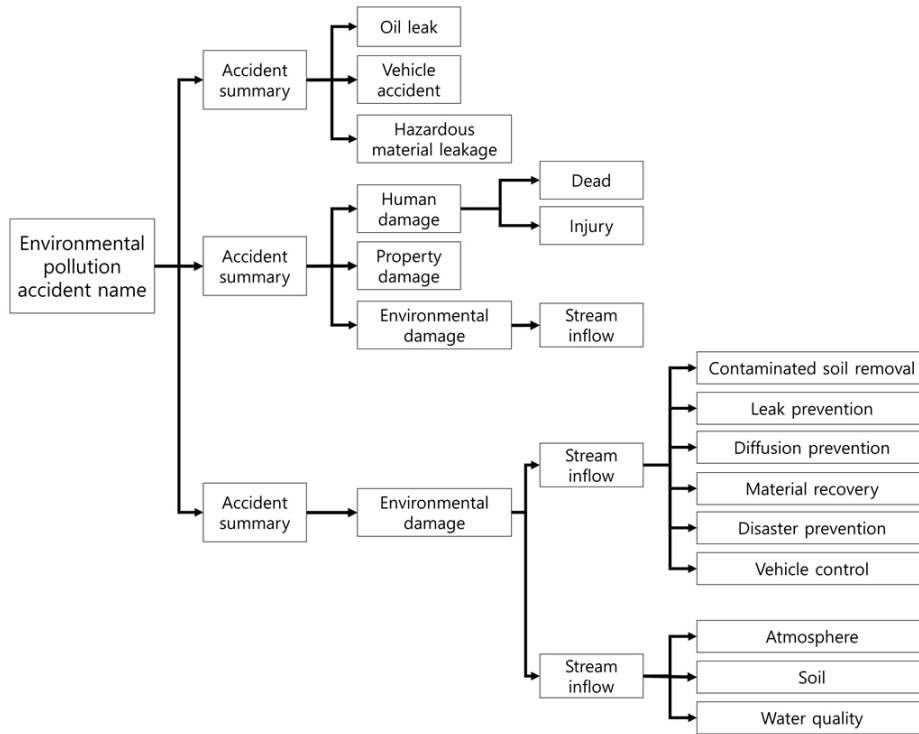


Fig. 1. Classification of Environmental Disaster Management (Lee et al. 2018).

법이다. 기계적으로 인식해야 할 텍스트는 손으로 쓰거나 컴퓨터에서 다양한 형태의 글꼴로 입력된 문서를 스캔이나 보고서 형태로 출력된 인쇄물이며, 이러한 이미지를 컴퓨터가 인식할 수 있는 인코딩된 텍스트로 변환하는 것이다. 특히 인공신경망 모델을 활용한 문자인식 기법은 최근 인공지능 및 머신러닝 기법을 활용하여 최대의 인간이 인식할 수 있는 능력을 과학적인 방법으로 시도되고 있다. 문자인식을 위해 이미지 문자는 선택적 다각형 텍스트 영역이 정의된 이진(Binary) 이미지라고 가정하여 3단계를 통해 문자를 인식한다. 첫 번째 단계에서는 이진 이미지의 구성요소와 저장된 문자의 구성요소의 분석이다. 이를 통해 문자의 윤곽선과 윤곽선의 수를 검사하여 문자를 쉽게 감지할 수 있다. 두 번째 단계에서는 각 단어를 차례로 인식한다. 세 번째 단계에서는 인식되지 않은 문자에 대해 다시 인식하는 작업을 수행한다. 마지막 단계에서는 대체 문자들을 찾기 위한 작업을 수행한다 (Smith 2007). 이미지화된 문자로 구성된 비정형 데이터와는 달리 정형 데이터는 컴퓨터가 인식할 수 있는 자료 구조로 구성되어 있기 때문에 바로 환경재난 표준화 체계에 적용할 수 있다.

2.2 환경재난 정보 표준화

환경재난에 대한 상황관리는 재난관련 담당자의 업무를 지원하기 위한 기술 개발로서 각종 재난 발생 시 인명 및 재산피해를 최소화하거나 방지하기 위하여 우선적으로 재난상황실에서 환경재난에 대한 정보 수집 및 신속한 상황전파와 초동 조치 및 지휘 업무를 수행하기 위한 모든 활동의 의미한다 (Lee et al. 2018). 재난정보 표준화 기술에서 표준화(Standardization)는 다양한 분야에서 다양한 형태로 존재하며, 사전적 의미로는 사물, 개념, 방법 및 절차 등에 대해서 합리적인 표준(Standard)을 설정하고 이를 활용하기 위한 규칙, 지침 또는 가이드 등을 만드는 조직적인 행위를 말한다. 본 연구에서 환경재난 정보 표준화의 의미를 방재적인 측면과 환경재난 정보적인 측면에서 정의하였다. 방재적 측면에서의 표준화 의미는 환경재난 정보의 생산, 운영, 관리, 유통(공유)에 대한 정보의 통일성과 일관성을 확보할 수 있도록 하여 효과적으로 정보를 활용할 수 있도록 한다. 환경재난 정보적인 측면으로 표준화는 정보를 표준화된 코드로 변환 및 분류하고 환경재난 정보

Table 1. The Principle of Environmental Disaster Data Management

Tag	Type	Example
Principle 1	Elimination of duplication data	The environmental disaster data is managed by the system to manage unnecessary duplication data related disaster
Principle 2	Compliance with all data standard	The environmental disaster data code, data management and data standard are established. All data comply with these standards
Principle 3	Unification of disaster data input	The environmental disaster must be entered only once. Those data are mainly generated and entered in the code system in one place. It has to be minimized duplication of data and it should be only be done for performance and availability reasons to ensure data accuracy and consistency
Principle 4	Improvement of data quality	Data quality management is carried out to improve data reliability and search results by adhering to the environmental disaster code and standards. Performing quality control is continuously carried out including post-event improvement of environmental disaster data

의 종류 및 규칙에 대한 표준을 정하는 행위로 정의하였다 (Jung 2018). Lee et al. 2018에서 제시하는 환경재난 관련 코드체계는 다음과 같다.

2.3 환경재난 데이터베이스 표준화

환경재난 정보를 체계적으로 활용하기 위하여 정보에 대한 표준을 정의하였다. 환경재난에 대한 업무, 절차의 기준, 정보 간 상호운용성을 보장하기 위한 데이터와 형식 표준화를 대상으로 하였다. 데이터베이스 표준으로는 Korea Database Promotion Center에서 제시하고 있는 데이터 표준은 정보관리 업무를 수행함에 있어서 필요한 데이터 및 데이터 셋을 말하여 연계 표준은 정보 시스템 사이 또는 수요자 사이의 정보를 연계하기 위한 표준을 말한다 (Korea Database Promotion Center 2006). 본 연구에서는 환경재난 정보를 관리하기 위한 데이터의 속성, 구조, 항목에 대한 명명 규칙과 데이터 표준관리체계의 지속적인 유지를 위한 품질관리가 이루어지도록 한다. 환경재난 정보관리의 원칙 (Table 1)을 수립하여 정보의 신뢰성을 확보한다.

환경재난 정보의 표준화를 위해 용어, 코드, 데이터 도메인 등의 속성을 정의하여 환경재난 데이터베이스에 일관되게 적용한다. 환경재난 정보의 데이터의 속성과 용어를 정의하여 표준용어으로써 관리하고 표준화된 용어사전을 통해 정보를 표준화한다. 이러한 표준용어는 표준화된 대상 용어, 영문 용어 명, 약어 명, 정의 및 데이터 형태로 정의되며, 향후 데이터베이스의 속성을 정의하기 위해 활용된다.

3. 결과 및 고찰

3.1 비정형 환경재난 정보의 자동 생성

환경재난의 비정형 데이터에 대해 광학문자인식 기술을 활용하여 환경재난 정보를 생산하였다. 한글이 포함된 비정형 환경재난 정보를 95% 이상으로 인식률을 높여 정보의 활용적인 측면을 높였다. Fig. 2와 같은 비정형 자료의 문자인식의 결과로 환경재난 정보 생성에 적합하였으며, 문자인식을 기반으로 하는 환경재난 정보 생산 자동화 시스템을 구축하였다

3.2 환경재난 정보의 계층화

재난정보 자동화 프로그램의 가장 중요한 기술은 문자인식기능을 포함한 모듈을 통해 비정형 재난안전정보의 전자화된 문서를 생성 및 관리의 기능과 전자화된 데이터로 분류하고 구조화하여 의미 있는 재난안전정보로 변환하는 기능이다. 특히 비정형 자료를 체계적으로 관리하기 위해 원본 데이터와 연계할 수 있는 기능을 포함하고 있다. 재난안전정보 자동화 기술인 비정형 자료가 텍스트로 변환된 자료 (Fig. 3)를 1차로 환경재난 코드와 매핑 및 가공하여 재난안전 데이터베이스에 저장, 관리한다. 이렇게 변환된 자료를 계층화할 수 있도록 전자화된 자료를 분석하고 계층화 (Table 2)를 위한 분석을 수행한다. 비정형 이미지를 계층화하기 위해 변환된 자료를 분석한 결과는 다음과 같다.

3.3 환경재난 정보 생산 자동화 시스템

본 연구에서는 환경재난의 비정형 자료를 정형화된

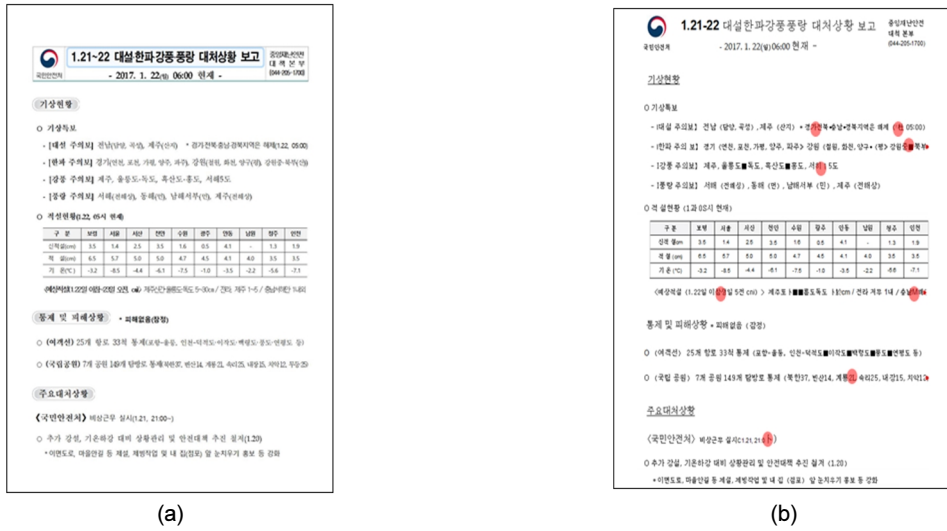


Fig. 2. Character Recognition (b) by Optical Character Recognition engine using unstructured image document (a).

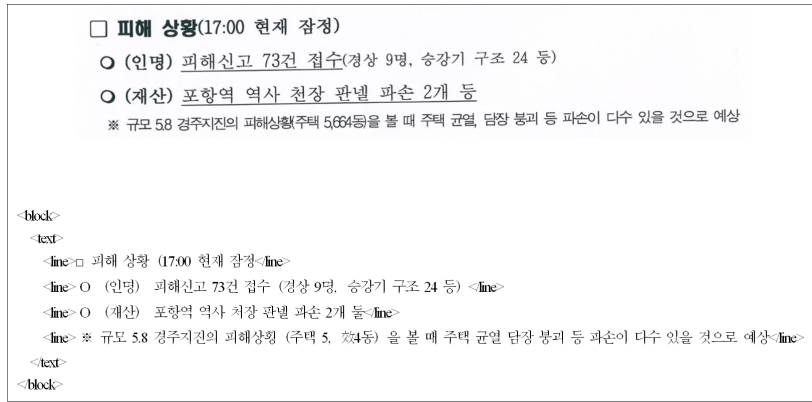


Fig. 3. Hierarchical Structure for unstructured data.

Table 2. Hierarchical data from automation program

Tag	Type	Example
<code><block> ... </block></code>	Subject and contents for the highest title	One structure
<code><text> ... </text></code>	Contain context of title	Rainfall Control and damage status Action Plan
<code><line> ... </line></code>	Contain items of title	□, ○ Details

자료로 변환하여 환경재난 코드에 연계하고 환경재난 정보를 관리하는 데이터베이스에 저장하기 위해 문자 인식모듈, 비정형 자료 등록, 관리, 변환된 텍스트 관리, 분류, 코드와의 연계 체계를 구성하기 위해 다음과 같은 프로세스를 정립하였다(Fig. 4).

문자인식 모듈은 비정형 자료의 변환을 통해 문서의

계층화를 통해 1차로 환경재난 코드와 매핑 및 가공하여 환경재난 데이터베이스에 저장한다. 데이터 관점에서 비정형 데이터가 데이터베이스에 저장되는 절차(Fig. 5)를 통해 환경재난 정보를 활용할 수 있는 기반이 된다.

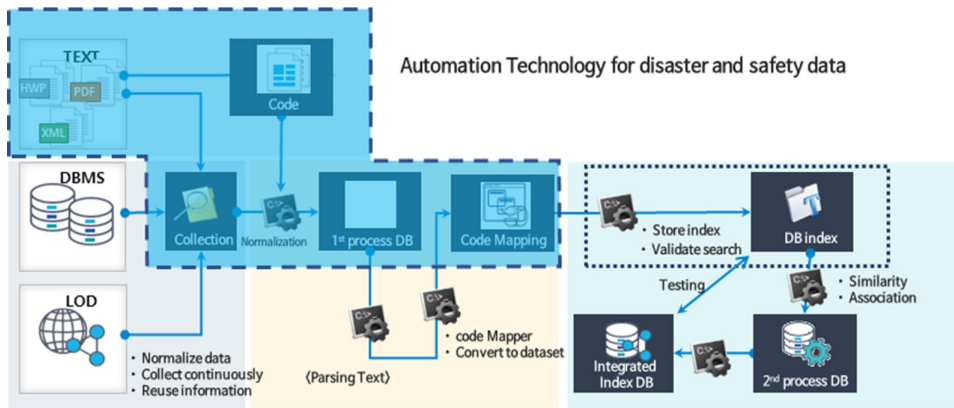


Fig. 4. The overview of Automated Collecting data for Environment Disaster Management.

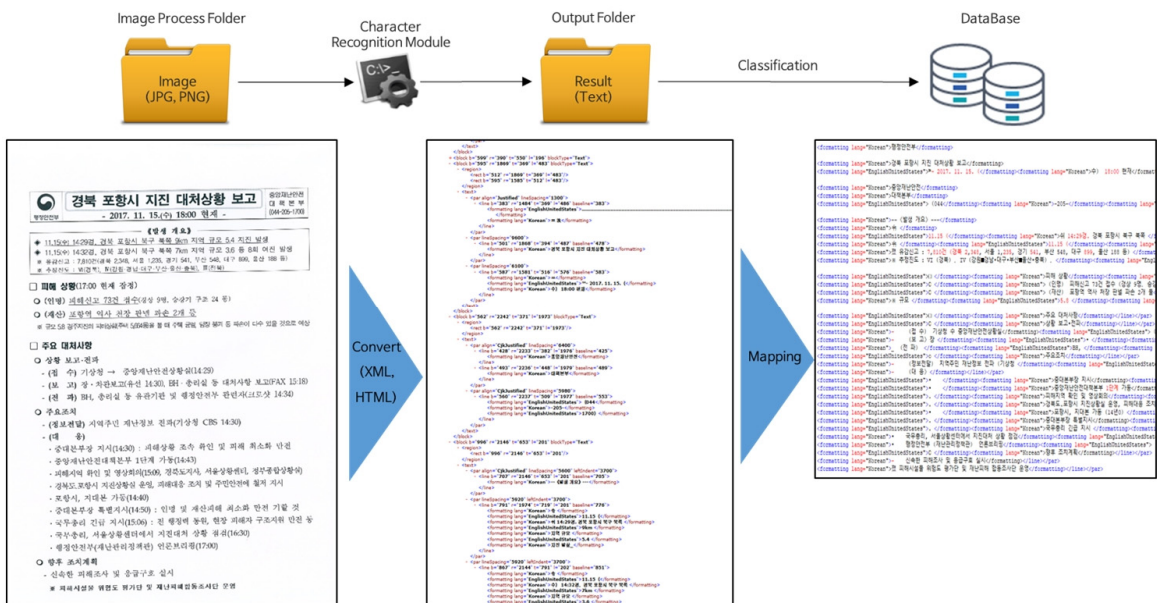


Fig. 5. The Mapping and Classification of Environmental Disaster Information.

4. 결론

본 연구를 통해 환경재난의 비정형 데이터를 활용을 위해 환경재난 정보의 표준화 및 자동화 생산 기술이 필요하다는 것을 확인하였다. 향후 4차 산업혁명에서의 환경재난 정보와 공간정보와 같은 빅데이터와의 융합, 활용에 대한 재난관리 패러다임의 변화에서도 볼 수 있듯이 환경재난 정보뿐만 아니라 모든 재난정보는 재난 담당자 중심의 활용과 재난대응에 이용할 수 있는 새로운 정보로서의 가치를 얻을 수 있음을 알 수 있다. 환경재난 정보를 자동으로 생산하기 위해 문자인식 기술을

활용한 재난정보 자동화 기술은 비정형 데이터의 인식률이 약95% 이상의 결과가 나왔기 때문에 이를 활용한 환경재난관련 비정형 문서들의 변환을 통해 환경재난 데이터베이스를 구축할 수 있을 것으로 생각되며, 이를 통해 재난 데이터를 활용할 수 있을 것으로 판단된다. 이러한 환경재난 정보의 효율적이고 체계적인 활용을 위해 비정형 자료의 데이터 수집, 환경재난 정보의 표준체계 등의 환경재난 정보 관점에서 정보 생산 시스템의 구축 방안을 제시하였다. 이를 기반으로 환경재난 관리 업무를 지원하기 위하여 재난업무 담당자의 업무별 맞춤형으로 환경재난 정보를 활용할 수 있는 시스템

의 방향을 제시하고자 과거의 환경재난 이력 정보 제공으로 인한 재난대응 담당자의 효과적 업무수행의 효과를 가지고 올 수 있을 것으로 기대한다. 본 연구를 통해 환경재난 정보의 활용 체계를 구축하여 정보공유 및 재난정보의 추적, 활용이 가능하게 하여 신속하게 대국민 재난정보를 제공할 수 있는 체계를 구축하여 경험기반의 재난관리 및 지능형 재난관리의 의사결정을 지원할 수 있는 시스템으로 확장할 수 있을 것으로 기대한다.

감사의 글

본 연구는 행정안전부 극한재난대응기반기술의 연구비지원 (2017-MPSS31-004)에 의해 수행되었습니다.

References

- Lee, Y. et al. 2018. Development of Environmental Disaster Information Standard Codes. *Journal of Korea Society of Hazard Mitigation* Vol. 18:649-654. (in Korean)
- Jung, I.-S. 2018, Development of Code System for Systematic Accumulation and Utilization of Disaster Safety Data. *Journal of Korea Academia-Industrial cooperation Society* Vol 19: 167-175. (in Korean)
- Smith, R. 2007. An Overview of Tesseract OCR Engine. *International Conference on Document Analysis and Recognition*, IEEE.
- Korea Database Promotion Center 2006. *The Guideline of Data Quality Management Maturity Model*.