

데이터를 기반으로 하는 기록관리기준표 자동 점검 적용 연구

A Study on the Application of Automatic Inspection of Data-Based Record Management Criteria

홍덕용 (Deok-Yong Hong)*

초 록

행정 업무가 종이 기반 관리체계에서 디지털 매체 관리체계로 전환되었음에도 불구하고, 여전히 공공기록물 관리에 관한 법률은 “종이기록의 전자적 관리”에 기반하고 있다. 특히 정부기능분류체계(BRM)의 단위과제별 보존기간, 공개 여부 등을 포함한 기록관리기준표를 통해 생산기록이 분류되고 있으나, 이러한 체계는 복잡한 기록 생산 환경을 제대로 반영하지 못하고 있다. 그 결과 중앙정부의 노력과 선행 연구자들의 다양한 방법에도 불구하고, 기록 생산자의 주관적 판단을 충분히 반영할 수 있는 업무 운영 방식의 부재와 지속 가능한 운영을 위한 피드백의 부족으로 인해, 모든 기관에서 기록관리의 시작 단계인 기록 분류 과정에 여전히 많은 어려움을 겪고 있다. 따라서 본 연구의 목적은 공공기관의 문서등록대장 데이터를 기반으로 기록관리기준표를 자동으로 점검·적용하는 것을 제안한다. 점검·적용 방법으로 첫째, 레벤슈타인 거리 알고리즘을 적용하여 단위과제 안에 기록물들의 편철이 적절한지 확인할 수 있다. 둘째, 기록관리기준표의 단위과제와 문서등록대장 데이터 간의 상관관계 분석을 통해 사용되지 않는 단위과제 카드를 식별하고 동일 단위과제명에 대한 보존기간 점검, 공개 여부의 적정성 확인이 가능하다. 셋째, 이 방법을 실제 업무에 적용하고 성능 평가를 위해 일반사용자 집단을 모집하고 연구방법에 대한 평가를 진행한다. 연구의 결론은 첫째, 기관에서 누적된 데이터를 기반으로 기록관리기준표가 자동 점검되기 때문에 실무적으로 높은 정확도를 보여준다. 둘째, 효율성과 용이성에 대해서는 높은 결과를 얻었지만 정보성에서는 다소 부정적인 결과를 얻었다. 셋째, 실무자들의 주관적인 의견이 대부분 적용되었던 종이기반의 기록관리업무가 데이터를 기반으로 하는 기록정보관리업무로 사고의 확장전환이 가능하다.

ABSTRACT

Despite the transition of administrative work from a paper-based management system to a digital media management system, the Public Records Management Act is still based on “electronic management of paper records”. In particular, production records are classified through the record management standard including the retention period and disclosure status for each unit task of the Government Functional Classification System (BRM), but this system does not properly reflect the complex record production environment. As a result, despite the efforts of the central government and various methods of preceding researchers, due to the lack of a business operation method that can sufficiently reflect the subjective judgment of the record producer and the lack of feedback for sustainable operation, all institutions still face many difficulties in the record classification process, which is the beginning stage of record management. Therefore, the purpose of this study is to automatically check and apply the record management standard table based on the document registration ledger data of public institutions. As a method of inspection and application, first, the Levenstein distance algorithm can be applied to check whether the compilation of records is appropriate within the unit task. Second, it is possible to identify unused unit task cards, check the retention period for the same unit task name, and check the appropriateness of disclosure by analyzing the correlation between the unit task of the record management standard and the document registration ledger data. Third, this method is applied to actual work, and general user groups are recruited for performance evaluation and research methods are evaluated. The conclusion of the study is, first, because the record management reference table is automatically checked based on data accumulated by the institution, It shows high accuracy in practice. it shows high accuracy. Second, high results were obtained for efficiency and ease, but somewhat negative results were obtained for informativity. Third, it is possible to expand and convert thinking to data-based record information management, where most of the subjective opinions of practitioners were applied.

키워드: 데이터기반, 빅데이터, 기록관리기준표, 정부기능분류체계, 레벤슈타인거리알고리즘, 자동화, 분류 data base, big data, record management reference table, government function classification system, Levenstein distance algorithm, automation, classification

* 부산광역시 수영구청 기록물관리전문요원(igre@korea.kr)

■ 논문접수일자 : 2025년 11월 5일 ■ 최종심사일자 : 2025년 12월 8일 ■ 게재확정일자 : 2025년 12월 13일
■ 정보관리학회지, 42(4), 489-506, 2025. <http://dx.doi.org/10.3743/KOSIM.2025.42.4.489>

© Copyright © 2025 Korean Society for Information Management

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 (<https://creativecommons.org/licenses/by-nc-nd/4.0/>) which permits use, distribution and reproduction in any medium, provided that the article is properly cited, the use is non-commercial and no modifications or adaptations are made.

1. 서론

1.1 연구의 배경 및 목적

행정업무 프로세스가 종이매체관리체계에서 디지털매체관리체계로 전환되었음에도 불구하고 「공공기록물관리에 관한 법률」은 여전히 “종이기록의 전자적 관리”에 기반하고 있다. 20세기 이후 종이라는 물리적 매체를 제한된 공간에서 효율적으로 관리하기 위해 분류·이관·평가 및 폐기 등의 업무 방식과 규정을 디지털매체관리체계에서도 그대로 구현 시킨 것이다.

종이기록물의 경우는 ‘유일성’이 기록의 본질을 규명하는 중요한 요소였으며 도서나 다른 자료와는 구분되는 기록의 고유속성으로 자리 잡아왔다. 전자기록물 역시 조직의 업무, 목적을 위해 생산되는 점에서는 동일하지만, 유일본으로서의 속성보다는 다수의 생산자에 의해, 시간 및 공간에 구애 없이 다수의 복본 생산이 가능한 복본성의 속성이 훨씬 강하다. 전통적 의미에서 원본기록은 특징의 물리적 매체에 담겨진 기록물을 의미하며, 원본과 다른 물리적 매체에 새겨진 동일한 기록은 복제본으로 간주된다. 전자기록물은 물리적으로 형태를 가진 매체뿐만 아니라, 기술과 자료(data)가 조합되어 만들어진 결과물로 볼 수 있다. 동일 기록물에 대한 각각의 열람은 그 기록물 자체의 새로운 “원본에 대한 복사본(original copy)”일 뿐이다. 종이기록물은 종이 매체에 기록의 내용·맥락·구조가 어느 정도 반영되어 있기 때문에 물리적 보존만으로도 기록의 내용과 더불어 종합

적 맥락 파악이 가능하며, 이를 통해 증거성 및 신뢰성, 진본성이 보존된다. 전자기록물은 내용·맥락·구조가 각기 별도로 존재하는 논리적 객체의 특성을 가진다. 따라서 각기 독립적으로 존재하는 이 세요소를 동시에 보존해야하는 고도의 어려움이 발생한다(국가기록원, n.d.).

특히 같은 법 시행령에 따라 공공기관은 업무수행 과정에서 기록물을 생산하거나 접수하였을 때에는 기록물의 등록·분류·편철 등에 필요한 조치를 하여야 한다. 그 중 기록물의 분류는 정부기능분류체계(BRM)의 단위과제별로 작성하는 것을 원칙으로 하며 보존기간, 공개여부 등을 포함한 기록관리기준표를 통해 운영되고 있다. 기록관리기준표를 통한 분류는 생산된 문서를 논리적으로 구조화시켜 업무 행위나 기록물을 조직화하여 체계적으로 관리·활용하기 위한 핵심업무절차이다. 하지만 전자기록으로 생산된 방대한 기록을 출처별로 구분하는 것은 급변하는 디지털기록환경에서는 혼란 그 자체이다(이승억, 설문원, 2017, 22).

이에 국가기록원에서는 현행 기록분류체계의 문제점과 개선점을 파악하기 위하여 각급 기관 기록물관리담당자를 대상으로 설문조사를 실시하였다.¹⁾ 현행 기록분류체계의 문제점은 ‘문서 이외의 다양한 유형의 기록물에 적용할 수 없다’가 가장 높게 나타났으며 이를 개선하기 위해서는 ‘단위과제 통제’, ‘BRM정비사업 실시’ 등이 필요하다고 결과가 도출되었다. 국가기록원은 설문조사의 후속조치로 지난 3년간(2022~2024) 정부합동평가를 통해 기관 기록관리기준표 기관공통단위과제를 정비하였다. 하

1) 기록분류체계 혁신 관련 설문조사 결과보고(국가기록원 수집기획과, 2018. 8.)

지만 여전히 현장에서는 실제 생산되는 문서의 정보와 단위과제 미적정과 단위과제간의 공개 값·보존기간 등의 미비점이 발생하고 있다.²⁾ 이유는 여전히 종이기록관리 운영방식에 따른 중앙집중식 통제가 담당자의 자율성을 보장하지 않고 업무담당자가 바뀔 때마다 주관적인 판단으로 단위과제가 생성되어 운영되기 때문이다.

즉, 언급된 중앙정부의 노력에도 불구하고 각급 기관의 기록물관리 현장에서는 기록물의 분류를 여전히 어렵게 느끼고 있다. 이는 주관적인 종이 관리 기반의 운영과 지속가능한 업무의 피드백이 부재하기 때문이다.

본 연구의 목적은 데이터를 기반으로 정확성과 효율성을 높이고 업무의 지속가능한 운영을 위해 기록관리기준표 자동 점검·적용한 방법을 제안하고자 한다. 이를 위해서 레벤슈타인 거리(Levenshtein Distance) 알고리즘을 사용한다. 이 알고리즘은 문자열 간의 유사도를 검사하는 것으로, 한 문자열을 다른 문자열로 변환하기 위해 필요한 삽입·삭제·대체 횟수를 수치화하는 알고리즘이다. 수치화된 숫자가 높을수록 두 문자열 간의 유사도는 낮아지고, 숫자가 낮을수록 유사도가 높은 것이 특징이다.

첫째, 연구자가 근무하고 있는 기관의 업무관리시스템 문서등록대장을 데이터로 수집한다. 둘째, 문서등록대장의 데이터와 기록관리기준표의 항목들을 교차분석하여 단위과제 카드에 편철된 기록물건을 레벤슈타인알고리즘으로 거리계산을 한다. 추가적으로 사용되지 않는 단위과제 카드를 식별하고 단위과제명에 대한

보존기간 점검, 공개 여부의 적정성 확인이 가능하다. 셋째, 데이터를 기반으로 하는 기록관리기준표 자동 점검 프로토타입을 실제 업무에 적용하고 성능 평가를 위해 일반사용자 집단을 모집하고 연구방법에 대한 평가를 진행한다.

1.2 선행연구

ICT 및 정부업무환경의 변화는 현재의 전자 기록관리 체계의 전면적 전환, 즉 패러다임 전환을 요구 하고 있다(이승역, 설문원, 2017, 7). 종이시대 기록은 저장 공간 및 관리 비용의 급증에 대처하기 위함이었다면 20세기 이후 기록의 생산량이 폭발적으로 증가하는 상황과 디지털 기술의 발전은 더 이상 종이 기반의 관리체계에서 사고의 전환을 요구하는 것이다(김명훈, 2024, 80).

이에 따른 기록관리분야에서도 종이기반의 관리체계를 벗어나 대량의 데이터를 저장·분석·수집한 뒤 분석을 통해 기록관리업무의 시작단계가 처한 현황을 이해하고 분석하여 문제를 해결는 연구들이 활발히 이루어지고 있다.

임진솔 외(2023)는 공공기관의 업무환경은 시시각각 변화되는 점을 인식하고 효과적인 기록관리를 위하여 과거의 기록물분류체계가 현재의 업무가 적당한지 검토되어야한다고 제시하였다. 분류체계가 구축되고 재정비를 하려면 기관 기록물 담당자나 업무담당자의 실무 경험적 판단이 가장 중요하기 때문에 수작업으로 진행된다. 이는 조직의 변화가 적시에 반영되

2) 2023년('22년 실적) 지자체 합동평가 기록관리분야(1-3-1-다. 처리과별 기록관리기준표 관리·운영 성과)
2024년('23년 실적) 지자체 합동평가 기록관리분야(1-3-11-다. 처리과별 기록관리기준표 관리·운영)
2025년('24년 실적) 지자체 합동평가 기록관리분야(1-3-11-라. 처리과별 기록관리기준표 관리·운영 성과)

거나 전체 큰 맥락을 통합적으로 파악하기가 어렵기 때문이다. 이러한 문제를 보완하기 위해 자동화 기술을 활용한 기록물 분류체계 재정비 방안을 제안하였다.

현문수(2021)는 국가기록평가 관점에서 기록관리기준표를 분석하고, 미래 개선 방향을 제안하고자 하였다. 국가적으로 영구보존해야 하는 기록의 범주와 이를 적용하는 각 기관의 기록관리기준표를 통해 어떻게 이행하며 이행 체계의 구성할 것인지에 대한 연구를 수행하였다.

홍덕용(2021), 홍덕용, 허준석(2023)은 기록관리분야에서 빅데이터 분석의 필요성을 알아보고 활용이 가능한 업무를 통해 공공기록관리 업무의 프레임워크를 도식하고 시사점을 제시하였으며 후속 연구로 텍스트마이닝 기법을 활용하여 국내 기록관리 연구 분야 동향을 파악하였다.

김다빈 외(2021)는 지방자치단체의 단위과제 현황과 운영 및 기록관리 문제점을 조사·분석한 뒤 빅데이터 분석기술을 활용하여 개선 방안과 표준안을 제시하였다.

조지영, 설문원(2018)은 현재 공공기관의 기록 평가 제도는 기록관리기준표를 중심으로 시행되고 있으나 업무관리시스템 외 적용되지 않는 등의 복합적인 기록생산 환경을 적절하게 반영하지 못하는 등 많은 한계점을 지닌 것을 확인하여 기록관리기준표를 개선하는 방안을 제시하였다.

레벤슈타인거리 알고리즘은 1965년 Vladimir Levenshtein에 의해 고안되었다. 두 개의 문자열의 유사도를 계산하여 오타를 수정하거나 중

복되거나 오기된 키워드를 찾아내는 등 표절 여부를 판별할 때 사용되고 있다. 이 때 두 문자열이 서로 동일해지기 위해 필요한 최소 편집 연산의 정도를 통해 측정되는 정규화를 수행하는데 이를 유사성³⁾이라고 한다. 레벤슈타인거리 알고리즘은 내용·구조·맥락을 포괄하지 않고 반복적 오분류를 줄이고 분류의 정확도를 향상시킬 수 있는 것이다. 즉, 큰 문제를 작은 문제들로 나누어 풀어나감으로써 해결하는 동적 방법이다.

레벤슈타인거리 알고리즘을 활용한 연구는 Yulianto et al.(2018), 안광모 외(2013), 이명훈 외(2012) 등 다수의 논문이 있다.

Yulianto et al.(2018)은 도서관의 검색시스템의 성능을 제고하고 이용자에게 더 나은 검색 환경을 제공하기 위해 레벤슈타인거리 알고리즘을 적용하여 철자검사 및 자동완성을 도입하였다.

안광모 외(2013)는 영화평가리뷰시스템의 제작을 위해 평가리뷰텍스트를 수집 한뒤 레벤슈타인 알고리즘을 사용하였다.

이종섭, 오상엽(2003)은 각 어휘의 실제 사용빈도에 따라 가중치를 부여하여 레벤슈타인 알고리즘을 사용하여 음성인식 기능의 정확도 제고의 연구를 진행하였다.

이명훈 외(2012)는 레벤슈타인 거리 알고리즘을 휴대폰 카메라를 이용한 간판 영상에서 한글 인식에 적용하였다.

기록관리업무분야 데이터와 IT기술을 활용한 관련 연구는 다수가 있으나 데이터를 기반과 레벤슈타인 거리 알고리즘을 적용하여 자동

3) 유사성 = $1 - \frac{\text{레벤슈타인 거리}}{\text{두 문자열 중 최대 길이}}$

점검 방법을 실제 업무에 적용한 연구는 전무하다. 기존의 구축된 시스템으로 필요한 시간 내에 처리하지 못하는 대량의 데이터를 보유하고 있는 공공기록물관리업무가 현재 처해진 상황처럼 지속된다면 단순히 데이터 창고를 열어주는 문지기 역할만 수행할 뿐이다. 즉 기술을 적극적으로 활용하여 미래에 대비하여야 한다.

2. 현황 및 필요성

2.1 기록관리기준표 현황

기록관리기준표는 2006년에 ‘공공기록물 관리에 관한 법률’이 제정되면서 기존의 기록물분류기준표의 미흡한 점을 대체하기 위해 제시되었다. ‘공공기록물 관리에 관한 법률 시행령’ 제 25조에서 공공기관은 업무과정에 기반을 둔 기록관리기준표를 작성·운영해야 하며, 안전행정부장관이 정하는 정부기능분류체계의 단위과제별로 작성하는 것을 원칙으로 하되, 공개여부 및 접근권한 등은 공공기관별로 달리 정할 수 있다고 규정하고 있다. 기록관리기준표

는 2008년부터 중앙행정기관에서 운영되기 시작하였으며, 지방자치단체에서도 점차 도입을 확대해 가고 있다(〈표 1〉 참조).

기록관리기준표는 기능분류체계의 단위과제별로 업무설명, 보존기간, 보존기간 책정사유, 비치기록물 여부, 공개여부 및 접근권한 등의 기록관리항목을 제시하고 있다. 단위과제 업무설명에는 기록물의 생산맥락을 파악할 수 있도록 업무의 성격, 업무처리절차 등을 기술해야 한다. 보존기간은 공공기록물법 시행령 [별표1] ‘기록물의 보존기간별 책정기준’, ‘처리과 공통업무 보존기간표’, ‘기간 공통업무 보존기간표’ 등을 참조하여 1년, 3년, 5년, 10년, 30년, 준영구, 영구 중 하나로 책정함을 원칙으로 한다. 보존기간 책정사유는 단위과제 보존기간 책정을 위해 실제로 참조하고 검토한 사항을 기술한다. 비치기록물 여부는 기록물 생산기관에서 카드·도면·대장 등과 같이 주로 사람·물품·관리관계 등에 관한 사항의 관리나 확인 등에 수시로 사용되는 기록물에 대하여 비치기록물로 지정하며 비치여부, 비치사유, 비치기간으로 구분하여 기술한다. 공개여부는 단위과제별로 공개 또는 비공개 기준, 비공개 대상정보 등을 기술하는데, 단위과제별로

〈표 1〉 기록관리기준표의 구조

정부기능분류체계(BRM)							기록관리항목						
조직 분류	기능분류						단위 과제 기능 유형						
처리 과명	1레벨 정책 분야	2레벨 정책 영역	3레벨 대기능	4레벨 중기능	5레벨 소기능	6레벨 단위 과제		업무 설명	보존 기간	보존 기간 책정 사유	비치 여부 및 이관 시기	공개 여부	접근 권한
							공통 / 고유	200자 이내	1,3,5, 10,30, 준영구, 영구	1,000자 이내 서술	비치/ 아님 (이관 시기)	(부분) 공개/ 비공개	열람 범위

제시하기 어려운 경우에는 기록물 유형별 또는 철별로 기술할 수 있다. 접근권한은 단위과제별로 생산되는 기록물의 접근 가능여부를 정하는 것으로 기록물 및 업무담당자를 기준으로 기록물 목록, 내용 열람 등 접근범위를 구분하여 설정해야 한다(국가기록원, 2025).

2.2 데이터를 기반하는 기록관리기준표 점검의 필요성

오늘날 공공기관의 기록관리는 대부분 전자적으로 생산되어 활용·평가·보존·폐기 등 전 생애주기에 걸쳐 효율적으로 관리되고 있다. 그러나 현행 「공공기록물 관리에 관한 법률」은 여전히 “종이기록의 전자적 관리”를 전제로 하고 있어, 실제 공공기관의 기록관리 실무환경과 제도적 기반 간의 괴리가 존재한다. 즉, 기록은

데이터 형태로 생산·관리되고 있음에도 불구하고, 이를 점검·평가하는 기준과 절차는 여전히 과거의 종이기록관리 체계를 따르고 있는 실정이다. 이러한 불일치는 기록관리의 효율성을 저해하는 주요 요인으로 작용한다.

이러한 점을 인지하고 있는 국가기록원 수집기획과에서는 기록분류체계의 문제점과 개선점을 파악하기 위하여 각급 기관 기록물관리담당자를 대상으로 기록분류체계 혁신 관련 설문조사를 <표 2>와 같이 실시하였다.

설문조사의 결과는 다음과 같다. 기록분류체계에 대한 인식과 문제점으로는 ‘문서 이외의 다양한 유형의 기록물에 적용할 수 없다’(71%), ‘단위과제 조정·관리가 용이하지 않다’(63%), ‘기록 중심의 분류체계라고 할 수 없다’(49%), ‘주요 생산 기록물 관리에 적합하지 않다’(60%) 등으로 나타났다.

<표 2> 국가기록원 기록분류체계의 문제점과 개선점 설문조사 개요

설문조사 목적		<ul style="list-style-type: none"> • 각급기관 기록물관리담당자를 대상으로 현행 기록분류체계의 문제점과 개선점 관련 의견 청취 • 새로운 기록환경에 부합하는 기록분류체계 혁신을 포함한 평가 정책 방향 수립을 위한 기초자료 수집 • 설문조사 분석결과 등을 참조하여 기록관리기준표 개선을 위한 로드맵 수립 예정
설문조사 설계	조사대상	<ul style="list-style-type: none"> • 각급기관 기록물관리담당자(총 1,607개 기관) - 기관유형: 헌법기관(국회, 대법원, 헌법재판소, 중앙선거관리위원회), 중앙행정기관(군포함), 특별지방행정기관, 대통령기록물생산기관, 지방자치단체, 교육청, 정부산하공공기관, 대학
	조사준비	<ul style="list-style-type: none"> • 설문지 작성 관련 소그룹 검토회의(2018. 6. 28. ~ 7. 12.) • 설문조사 계획보고(2018. 7. 17.) • 사전점검 실시(2018. 7. 17. ~ 7. 20.) • 각급기관 설문참여 안내 공문 발송 및 온라인 홍보(2018. 7. 27.)
	조사방법	<ul style="list-style-type: none"> • 국가기록원 홈페이지<온라인 설문조사>에 자율적 참여
	조사기간	<ul style="list-style-type: none"> • 2018. 8. 1. ~ 8. 14.(2주)
설문조사 주요내용		<ul style="list-style-type: none"> • 기록분류체계 운영 환경 • 기록분류체계에 대한 인식 • 새로운 기록분류체계 설계도입 • 현행 기록분류체계의 문제점 보완 및 개선 • 종합의견

이러한 문제점을 개선하기 위해 'BRM정비사업 실시'(46%), '단위과제 사전협의제'(50%), '핵심기록 평가도구 활용'(47%), '단위과제카드 통제(51%)', '기록관리기준표 관리항목정비'(43%)로 나타났다. 특히 '기록관리기준표 관리항목 정비'를 희망하는 응답 중 '비치기록물 해당여부'(49%), '비치사유'(53%), '비치기간'(55%), '보존장소'(51%), '보존방법'(59%)을 불필요한 항목으로 지목했고, '기록물 유형', '중요기록물', '관련 법령 보존기간', '보유기간' 등을 추가해야할 항목으로 제시하였다.

이에 따른 국가기록원의 후속조치로 지난 3년간(2022~2024) 정부합동평가를 통해 기관 기록관리기준표 중 기관공통단위과제를 정비하였다. 하지만 여전히 현장에서는 분류체계와의 미스매칭과 오분류 등 미비점이 발생하고 있다.⁴⁾ 이유는 여전히 종이기록관리 운영방식에 따른 중앙집중식 통제가 담당자의 자율성을 보장하지 않고 업무담당자의 객관적인 방법이 아니라 주관적인 판단으로 단위과제가 생성되어 편철되고 운영 중이기 때문이다.

특히 현재 각급기관의 기록관리기준표의 운영과 관리가 제대로 작동하는지에 대한 점검 규정이 없다. 기록관리기준표의 운영과 관리가 어디서부터 얼마나 잘못된 것인지 파악조차 할 수 없는 것이 현실이다. 만약에 규정이 있더라도 기관의 분류체계, 비공개 세부기준, 보존기간 책정, 단위과제편철의 적절성 등 다양한 항목을 사람의 손으로 일일이 확인해야 하므로 막대한 시간과 인력, 비용이 소요될 것이라고

예상된다. 더불어 수작업 점검은 점검자의 숙련도나 주관적 판단에 따라 정확성과 일관성이 떨어지는 한계를 지닌다.

이에 따라 기존에 기관에 누적된 데이터를 기반으로 하는 기록관리기준표 자동 점검 방법이 필요하다. 이 방법은 생산된 전자기록물의 메타데이터와 분류체계를 자동으로 상관분석하여, 단위업무 분류 적정성, 보존기간 설정의 일관성, 비공개 기준의 적용 여부 등을 관리되고 있는 데이터를 기반으로 정량적으로 점검할 수 있다. 예를 들어, 업무관리시스템의 문서등록대장의 기록물 제목과 단위과제의 유사도를 레벤슈타인 거리(Levenshtein Distance) 알고리즘으로 계산하거나, 메타데이터 검증 기능을 통해 대량의 전자기록을 신속하고 객관적으로 평가할 수 있다.

이와 같은 자동화 점검 방법은 기존의 수작업 방식에 비해 점검 효율성을 비약적으로 향상시키며, 동시에 정확성과 객관성을 확보할 수 있다. 나아가 이러한 방법은 공공기록물을 단순히 "보존의 대상"이 아니라, 지속적으로 관리·재활용되는 데이터 자산(Data Asset)으로 전환하는 기반이 될 것이다.

결국, 데이터를 기반으로 하는 자동 점검 방법은 디지털 행정환경에 부합하는 기록관리 패러다임의 전환이며, 공공기관 기록관리의 투명성과 신뢰성을 제고하고, 궁극적으로는 국가차원의 데이터 거버넌스를 강화하는 핵심 수단이라 할 수 있다.

4) 2023년('22년 실적) 지자체 합동평가 기록관리분야(1-3-1-다. 처리과별 기록관리기준표 관리·운영 성과)
2024년('23년 실적) 지자체 합동평가 기록관리분야(1-3-11-다. 처리과별 기록관리기준표 관리·운영)
2025년('24년 실적) 지자체 합동평가 기록관리분야(1-3-11-라. 처리과별 기록관리기준표 관리·운영 성과)

3. 연구방법

3.1 데이터 수집

연구자가 근무하고 있는 부산광역시 기초자치단체 S구청에서 데이터를 수집하였으며 업무관리시스템 문서등록대장의 데이터 항목과 그 데이터를 정제하는 방법에 대해 설명한다.

사전 연구들의 결과에 따르면 레벤슈타인거리알고리즘을 기반으로 한 연구의 정확도는 기준에 축적된 데이터의 양과 항목에 결정되기 때문에 업무관리시스템 문서등록대장의 데이터를 104만 건을 수집하여 중복된 데이터를 제거하는 등 정제 후 83만 건을 연구에 사용하였다 (<표 3> 참조).

상관관계를 적용할 분석대상은 부산광역시 기초자치단체의 수연구청 업무관리시스템 문서등록대장의 데이터이다. 2023년~2024년 2년간 생산된 기록물의 데이터로 사용한 항목은 문서번호, 보고일자, 문서구분, 제목, 수(발)신자, 보고자, 검토자, 상태, 붙임, 종류, 분리, 생

산등록번호, 공개구분, 과제카드명, 과제명, 과제보존기간, 목록공개여부, 문서비공개사유, 외부(민원인)주소, 등록구분 총 19개이다. 개발 언어로는 MS 오피스 엑셀을 기반으로 액세스 DB 즉, 기록관리기준표를 연동하여 작업하였으며 레벤슈타인 거리 알고리즘을 통해 기록관리기준표 단위과제 및 생산기록물등록대장의 기록물건 간 유사도 측정을 통한 편철의 적정여부의 확인이 가능하도록 고안하였다. 또한 기록관리기준표와 생산기록물등록대장의 기록물건의 데이터 간의 상관관계분석을 통해 미사용 단위과제카드 확인, 동일단위과제명의 보존기간 확인, 공개여부 등 확인이 가능하도록 하였다. 데이터 분석 절차는 <표 4>와 같다.

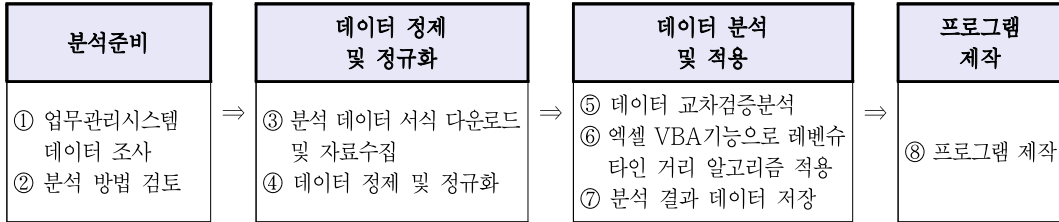
3.2 레벤슈타인 거리 알고리즘 적용

본 연구는 문자열 간 편집 거리를 계산하는 레벤슈타인 거리(Levenshtein Distance) 알고리즘을 활용하여 공공기록물과 기록관리기준표를 자동 점검하는 방법을 제안한다. 레벤슈타인

<표 3> 데이터 수집과 데이터 추출 항목

업무관리시스템 문서등록대장 화면		데이터 추출 항목	
<p>국가</p> <p>문서번호</p> <p>보고일자</p> <p>문서구분</p> <p>제목</p> <p>수(발)신자</p> <p>보고자</p> <p>검토자</p> <p>상태</p> <p>붙임</p> <p>종류</p> <p>분리</p> <p>생산등록번호</p> <p>공개구분</p> <p>과제카드명</p> <p>과제명</p> <p>과제보존기간</p> <p>목록공개여부</p> <p>문서비공개사유</p> <p>외부(민원인)주소</p> <p>등록구분</p>	<p>문서번호</p> <p>보고일자</p> <p>문서구분</p> <p>제목</p> <p>수(발)신자</p> <p>보고자</p> <p>검토자</p> <p>상태</p> <p>붙임</p> <p>종류</p> <p>분리</p> <p>생산등록번호</p> <p>공개구분</p> <p>과제카드명</p> <p>과제명</p> <p>과제보존기간</p> <p>목록공개여부</p> <p>문서비공개사유</p> <p>외부(민원인)주소</p> <p>등록구분</p>	<p>문서번호</p> <p>보고일자</p> <p>문서구분</p> <p>제목</p> <p>수(발)신자</p> <p>보고자</p> <p>검토자</p> <p>상태</p> <p>붙임</p> <p>종류</p> <p>분리</p> <p>생산등록번호</p> <p>공개구분</p> <p>과제카드명</p> <p>과제명</p> <p>과제보존기간</p> <p>목록공개여부</p> <p>문서비공개사유</p> <p>외부(민원인)주소</p> <p>등록구분</p>	<p>문서번호</p> <p>보고일자</p> <p>문서구분</p> <p>제목</p> <p>수(발)신자</p> <p>보고자</p> <p>검토자</p> <p>상태</p> <p>붙임</p> <p>종류</p> <p>분리</p> <p>생산등록번호</p> <p>공개구분</p> <p>과제카드명</p> <p>과제명</p> <p>과제보존기간</p> <p>목록공개여부</p> <p>문서비공개사유</p> <p>외부(민원인)주소</p> <p>등록구분</p>

〈표 4〉 데이터 분석 절차



거리 알고리즘은 한 단어를 다른 단어로 변경하는 데 필요한 편집의 최소 개수를 의미하며 삽입, 삭제, 대체의 과정으로 두 단어의 편집거리를 정량화한다. 공공기관에서 생산되는 기록물의 제목은 기관·부서별로 기록물의 생산자 보고서 작성 능력과 관행 차이, 띄어쓰기 오류, 철자 변형, 유사 의미 단어 사용 등으로 인해 동일 성격의 문서가 서로 다른 단위과제로 분류되거나, 문서공개값·보존기간 등이 불일치하는 문제가 지속적으로 발생하고 있다. 이러한 문제는 과거일 때는 기록물의 생산량이 적었기 때문에 생산자 또는 기록물관리전문요원의 판단만으로도 수동점검이 가능하였다. 하지만 현대에는 기록물의 생산량이 대규모로 발생되고 있고 기존의 방식으로 오분류를 점검하는 것이 어렵고, 이를 점검하였을 때 정확성 역시 생산자와 기록물관리전문요원의 업무경험과 시간에 크게 의존되고 있다. 이에 본 연구는 기록물 제목 간 유사도를 정량화하여 자동으로 오류를 점검하는 데이터 기반의 접근법을 적용하였다.

레벤슈타인 거리는 두 문자열을 동일하게 만드는 데 필요한 삽입, 삭제, 교체 횟수를 계산함으로써 유사도를 평가하는 것이다. 예를 들어 “회의록”과 “회의 록”은 띄어쓰기 하나만 다르므로 거리가 1이며, “업무보고”와 “업무보고서”는 접미사 “서”가 추가된 것뿐이므로 거리 1로

계산된다. 반면 “회의록” 제목의 기록물건과 “대책보고” 제목의 단위과제는 동일 업무이지만 의미가 전혀 다른 문서는 거리 4 이상으로 나타나며 유사성이 낮다. 이러한 거리값으로 기록물건 제목과 단위과제 간 관계를 정량적으로 판단할 수 있게 하여, 미사용 단위과제, 중복된 단위과제 보존기간, 잘못된 문서공개 재분류 등을 자동으로 식별하는 데 유용한 것이다. 문서공개재분류에서도 기록관리기준표 단위과제를 생성할 때 초기에 공개값을 지정하는 것이 아니라 생산된 기록물의 공개 데이터로 기록관리기준표 단위과제의 공개값을 지정하는 동일 분류 그룹으로 묶어 정확성을 향상시킬 수 있었다.

특히 공공기관에서는 동일 문서가 “업무보고”, “업무 보고”, “업무보고서”와 같이 소폭의 표기 차이 때문에 서로 다른 단위과제로 편철되는 경우가 많은데, 본 연구에서 적용한 유사도 임계값(거리 1~3)을 기준으로 자동 그룹화한 결과, 전체 데이터의 약 18~26%가 실제로 중복된 문서로 확인되었다. 이는 단위과제 통합, 표준 제목 정립, 기록관리 기준의 정합성 확보에 큰 기여를 할 수 있는 수치이다.

즉, 레벤슈타인 거리를 활용한 본 연구의 기록관리기준표 자동 점검 방식은 기존의 종이기반의 수작업 중심 점검 방식보다 빠르고 정확하며, 기록물의 제목 표준화, 단위과제 정합성

확보, 문서공개값 분류 정확도 향상, 보존기간 설정 오류 방지 등 다양한 영역에서 높은 효과를 보여줄 수 있다. 본 연구는 데이터를 기반으로 하는 기록관리의 혁신을 위하여 알고리즘의 활용 가능성을 실증하였으며, 향후 Word2VEC (워드임베딩) 기반 의미 유사도 분석과 결합할 경우 기록의 내용, 구조, 맥락 등을 고려한 기록관리기준표 자동 점검 분류 체계를 구축할 수 있을 것으로 기대된다.

3.3 단위과제카드 편철 유사도 및 보존기간, 공개여부 정확도 측정 결과

업무관리시스템의 문서등록대장에 등록된 기록물의 제목과 기록관리기준표 단위과제 안에 편철된 기록물 간의 문자 간의 사이를 레벤슈타인 거리 알고리즘으로 계산한다. 키(Key)가 기록물제목, 값(value)이 레벤슈타인 거리로 이

루어진 엑셀 VBA 자료형을 생성하고 문자열 간의 유사도를 분석하였다. 수치화된 숫자가 높을수록 두 문서등록대장에 등록된 기록물 제목과 기록관리기준표 단위과제 안의 문자열간의 유사도는 낮아지고, 숫자가 낮을수록 유사도가 높은 것으로 개발하였다. 공공기관은 별도의 서버나 고성능 연산 환경을 갖추지 않은 경우가 많기 때문에 엑셀 VBA 기반 환경에서도 경량으로 구현할 수 있음을 입증하였다. 연구자가 구현한 방식은 단일 PC 환경에서도 신속하게 수천~수만 건의 기록 데이터를 비교·분석할 수 있다는 장점을 갖는다. 이는 기록관리 업무의 자동화·지능화를 추진하는 데 있어 즉시 적용 가능한 실용적 접근이다.

〈표 5〉는 본 연구에서 제안하는 ‘데이터를 기반으로 하는 기록관리기준표 점검 자동 방법’을 통해 기록관리기준표 단위과제와 기록물건 제목 간의 유사도 측정을 통한 편철 적정 여부

〈표 5〉 데이터를 기반으로 문서 간의 유사도 측정을 통한 편철 적정 여부 분석 결과

연번	단위과제명	평균유사도	점검결과
1	해양오염일반	42.99	(일자리경제과) 지역공동체일자리사업 현장 작업 교육일지, 전기요금 납부건의 등 해당 문서철과 관련 없는 문서 다수 확인
2	운송주선업 일반관계철	42.73	(교통행정과) 문서철 다수가 화물 결격사유로 되어있으나 ‘화물결격사유조항’(전국)단위과제가 별도로 있는 것으로 확인
3	유통조례제정관련	42.00	(일자리경제과) 유통조례와 제정관련과 관계 없는 문서철 발견(1건)
4	건축 (산지전용)협의	40.23	(일자리경제과) 실제로 산지전용 관련 공문이 아니나, 별도 처리할 카드가 없어 해당 카드로 문서철 함
5	책임보험 관계철	38.41	(교통행정과) 고시송달 단위과제카드가 별도로 있으나 책임보험 관계철 내 고시 송달 관련 공문이 있는 것으로 확인
6	음식물 쓰레기수거	38.37	(자원순환과) 문서철은 잘되어 있으나, 과제카드명을 ‘음식물쓰레기수거’로 수정하는 것이 적절함
7	부동산 실명제업무	38.27	(토지정보과) 고시송달 단위과제카드가 별도로 있으나 부동산실명제업무에 송달 관련 공문이 있는 것으로 확인
8	조경협의	37.50	(일자리경제과) 단순 건축허가, 건축물 표시변경 등 조경협의와 관련없는 공문이 문서철로 되어있는 것으로 보임, 건축물 관련 단위과제카드 별도로 생성필요

결과를 보여주고 있다.

일자리경제과 '해양오염일반'의 단위과제는 평균유사도가 42.99로 실제로 편철된 내용을 살펴보면 지역공동체일자리사업 현장 작업 교육일지, 전기요금 납부건의 등 해당 문서철과 관련 없는 문서 다수 확인되었다. 교통행정과의 '운송주선업일반 관계철'의 단위과제는 평균 유사도가 42.73으로 편철된 내용은 화물 결격 사유가 편철되어있으며 또한 부서에서는 '화물 결격사유조회(전국)' 단위과제가 존재하고 있다. 토지정보과의 '부동산실명제업무' 단위과제는 평균유사도가 38.27로 실제로 편철된 내용에는 부동산실명제업무와 관련이 없는 고시승달 관련 공문이 포함되어있었다.

다음으로 생산된 기록물의 공개값과 기록관리기준표의 단위과제 공개값 간 사이의 정확도를 비교분석한 결과를 <표 6>으로 나타냈다. 19,167건의 불일치가 나온 "화물결격사유조회(전국)"의 단위과제의 공개값은 대국민 공개이지만 실제로 문서편철은 비공개로 진행되고 있다. 다음

으로 12,130건의 불일치가 나온 "멸실인정"의 단위과제의 공개값은 대국민 공개이지만 실제로 문서편철은 부분공개로 진행되고 있다. 알 수 있는 점은 기록관리기준표의 단위과제의 공개 기준값을 "대국민공개"로 생성한 뒤 실제 문서를 생산할 시 비공개 또는 부분공개로 결정한다는 것이다. 즉, 기준공개값과 실제공개값의 괴리가 발생하고 있는 것을 알 수 있다.

다음으로 생산된 기록물의 보존기간과 기록관리기준표의 단위과제 보존기간의 정확도를 분석한 결과를 <표 7>로 나타냈다.

부서간에도 같은 단위과제명을 사용하고 있지만 이들 간의 보존기간이 상이하다. 가장 많이 보존기간이 불일치한 단위과제는 일자리경제과의 "타시군구통보철"이다. 기록관리기준표 내 소기능에 따라 6개로 분류되어 있으며 업무 성격에 따라 1년과 3년으로 나누어져 있다. 다음으로 불일치한 단위과제는 남천1동·남천2동·수영동 등 행정복지센터와 안전관리과의 "안전관리일반"이다. 행정복지센터에서는 보존기간

<표 6> 생산된 기록물의 공개값과 기록관리기준표의 단위과제 공개값 간 사이의 정확도 분석

구분	내용	불일치 상위 15건
생산된 기록물과 기록관리기준표 단위과제 간의 공개여부 일치 문서 건수	317,435	
생산된 기록물과 기록관리기준표 단위과제 간의 공개여부 불일치 건수	515,169	
생산된 기록물과 기록관리기준표 단위과제 간의 공개여부 정확도	38.1%	

〈표 7〉 생산된 기록물의 공개값과 기록관리기준표의 단위과제 공개값 간 사이의 정확도 분석

구분	내용	불일치 상위 7건																
생산된 기록물과 기록관리기준표 단위과제 간의 보존기간 일치 문서 건수	824,746	<table border="1"> <caption>불일치 상위 7건</caption> <thead> <tr> <th>단위과제</th> <th>정확도</th> </tr> </thead> <tbody> <tr> <td>부산광역시청</td> <td>10,072</td> </tr> <tr> <td>부산광역시청</td> <td>3,775</td> </tr> <tr> <td>부산광역시청</td> <td>1,526</td> </tr> <tr> <td>부산광역시청</td> <td>697</td> </tr> <tr> <td>부산광역시청</td> <td>631</td> </tr> <tr> <td>부산광역시청</td> <td>59</td> </tr> <tr> <td>부산광역시청</td> <td>35</td> </tr> </tbody> </table>	단위과제	정확도	부산광역시청	10,072	부산광역시청	3,775	부산광역시청	1,526	부산광역시청	697	부산광역시청	631	부산광역시청	59	부산광역시청	35
단위과제	정확도																	
부산광역시청	10,072																	
부산광역시청	3,775																	
부산광역시청	1,526																	
부산광역시청	697																	
부산광역시청	631																	
부산광역시청	59																	
부산광역시청	35																	
생산된 기록물과 기록관리기준표 단위과제 간의 보존기간 불일치 건수	7,858																	
생산된 기록물과 기록관리기준표 단위과제 간의 보존기간 정확도	99%																	

이 5년, 안전관리과에서는 3년으로 지정되어있다. 그 다음으로 일자리경제과·자원순환과·환경위생과의 단위과제 “차량운행일지”이다. 일자리경제과·자원순환과에서는 3년, 환경위생과에서는 5년으로 보존기간이 불일치한다. 안에 들어있는 기록물건을 조사해본 결과 업무성격은 동일하였다.

4. 기록관리기준표 자동 점검 프로토타입 업무적용과 시사점

4.1 업무 적용 및 사용자 평가


앞에서 수행된 전자문서생산시스템의 문서 등록대장의 기록물 제목과 기록관리기준표 단위과제 안에 편철된 기록물들의 평균유사도를 레벤슈타인 거리 알고리즘을 사용하여 측정하고 보존기간과 공개여부 정확도를 측정하고 결과를 부산광역시 S구청 부서에 〈그림 1〉과 같이

적용하였다.

객관적인 데이터만으로 부산광역시 S구청 기록관리기준표를 자동으로 점검할 수 있었는데 단위과제카드 내 유사도가 낮은 하위 8건 중 2건을 재편철, 동일단위과제명이나 부서간 보존기간이 상이한 20건 중 8건을 재편철할 수 있었다. 특히 사용하지 않는 단위과제카드 520건 중 159건을 삭제할 수 있었다.


이에 대한 데이터를 기반으로 기록관리기준표를 자동으로 점검하는 방법에 대한 사용자평가를 진행하였다. 부산광역시 S구청 부서 기록물관리책임자 34명을 대상으로 정확성, 용이성, 정보성, 효율성에 대한 문항으로 구성된 리커트 5점 척도를 사용한 설문지를 통해 진행하였다. 기록관리업무의 전문지식이 없는 부서 기록물관리책임자를 평가 집단에 포함시켜 실무 관점에서 사용성이 높은지 평가하여 실제 도입 가능성 검증, 구조화된 평가 설계를 통한 성능 검증하였다.

설문에 참여한 표본 통계와 평가문항은 〈표 8〉



수영구

정당한 인식이 없기 때문에 수정주의 주입입니다



수신 수신자 참조
(경유)
제목 데이터 기반의 기록관리기준표 운영을 위한 단위과제 편일적정 및 미사용여부 조사 협조

- 행정지원과-22587(2024. 6. 14.)호 및 민원어권과-11710(2024. 9. 2.)호 관련입니다.
- 지난 3년간 정부합동평가를 통해 기록관리기준표 기관공통 단위과제를 정비하여 고도화가 되었으나 문서편일성 여전히 오분류와 미사용 단위과제가 발생하고 있습니다.
- 이에 따라 우리 구에서 생산된 기록물과 기록관리기준표에 대한 빅데이터 분석 결과를 바탕으로 단위과제 편일의 적정 및 미사용 단위과제 삭제여부를 조사하고자 하니 (서식1)을 2024. 9. 13. 까지 제출하여 주시기 바랍니다.

가. 조사기간 : 2024. 9. 5.(목) ~ 9. 13.(금)
나. 조사대상 : 해당부서 단위과제
다. 조사내용 :
- 동일단위과제명이나 부서간 보존기간이 상이한 20건의 정리 여부
- 단위과제카드 내 문서유사도 하위 8건의 재편일 여부
- 미사용 단위과제 카드 520건의 삭제 여부

라. 향후계획
- 해당부서 단위과제 재편일 및 미사용단위과제 삭제 여부를 조사 후 시스템 반영
- 2025년(2024년 실적) 기록관리분야 정부합동평가 우수사례 제출 예정

◆ 빅데이터 분석 현황 ◆

- 분석기간 : 2024. 6. 7.(금) ~ 6. 14.(금)
- 분석대상 : 2023년 수영구 온나라시스템에서 생산된 기록물 약 83만건(1년간)
- 분석방법 : 레벤슈타인 거리 알고리즘 및 범정부 데이터분석시스템(Tabieau)
- 분석결과 : 기록관리기준표 단위과제 및 기록물건 유사도 측정을 통한 편일적정 및 미사용 단위과제 카드 확인, 동일단위과제명의 보존기간 확인 등

업무보고

등록번호	민원어권과-12919	주무관	여유기록물과장	합제어권과장
등록일자	2024. 9. 24	홍익훈	조병준	김재 2024. 9. 24
결재일자	2024. 9. 24	합조장	기록관리과장	민대기
과제구분	양문민 과제			

제 목 : 데이터 기반의 기록관리기준표 운영을 위한 단위과제 편일적정 및 미사용여부 조사 결과보고

우리구에서 수행한 빅데이터 분석 결과를 바탕으로 단위과제 편일적정 및 미사용 단위과제 삭제여부를 조사한 결과임(데이터기반 행정 활성화)

○ 추진근거
○ 행정지원과-22587(2024. 6. 14.)호 및 민원어권과-11710(2024. 9. 2.)호 관련

◆ 빅데이터 분석 현황 ◆

- 분석기간 : 2024. 6. 7.(금) ~ 6. 14.(금)
- 분석대상 : 2023년 수영구 온나라시스템에서 생산된 기록물 약 83만건(1년간)
- 분석방법 : 레벤슈타인 거리 알고리즘 및 범정부 데이터분석시스템(Tabieau)
- 분석결과 : 기록관리기준표 단위과제 및 기록물건 유사도 측정을 통한 편일적정 및 미사용 단위과제 카드 확인, 동일단위과제명의 보존기간 확인 등

○ 추진배경
○ 지난 3년간 정부합동평가를 통해 기록관리기준표 기관공통 단위과제를 정비하여 고도화가 되었으나 오분류 등 미비점이 발생
○ 빅데이터분석을 통해 기록관리데이터 기반 행정업무의 활성화를 위한

○ 추진결과
○ 추진기간 : 2024. 9. 5.(목) ~ 9. 20.(금)
○ 추진대상 : 해당부서 기록관리기준표 미비점 발생 단위과제
○ 추진결과
- 동일단위과제명이나 부서간 보존기간이 상이한 20건 중 8건 재편일
- 단위과제카드 내 문서유사도 하위 8건의 2건 재편일
- 미사용 단위과제 카드 520건 중 159건 삭제

○ 향후계획
○ 해당부서 단위과제 재편일 및 미사용단위과제 삭제 여부를 조사 후 시스템 반영
○ 2025년(2024년 실적) 기록관리분야 정부합동평가 우수사례 제출 예정

붙임 1. (서식1)수영구 기록물관리 빅데이터분석에 따른 조사 1부.
2. 수영구 기록관리기준표 데이터분석 결과 보고(행정지원과) 1부.
3. 기관공통업무 단위과제 및 과제카드 정비 결과 보고(민원어권과) 1부. 끝.

2024. 9. 24. 민원어권과 기록물전문위원 홍 덕 용

〈그림 1〉 데이터기반의 기록관리기준표 자동 점검 프로토타입 업무적용 관련 공문 및 결과보고

〈표 8〉 사용자 평가 인구 통계와 평가문항

구분		빈도	비율(%)	항목	문항
성별	남성	9	26.5	정확성	자동 점검의 정확도는?
	여성	25	73.5		자동 점검에 오류여부는?
연령	20대	20	58.8	용이성	자동점검으로 편일 이해정도는?
	30대	10	29.4		자동점검으로 재편일이 편리한지?
	40대	4	11.8	정보성	자동점검이 정확한 내용을 제공하는지?
	50대	0	0		결과가 세부적으로 도출되었는지?
학력	고졸	0	0	효율성	이 방법이 향후에도 지속적으로 사용할 의도가 있는지?
	대졸	36	100		시간의 단축정도는?
	석사	0	0		
	박사	0	0		

과 같다. 실험에 참여한 총원은 34명이었으며 성별은 남성이 9명(26.5%), 여성이 25명(73.5%)로 나타났다. 연령대는 20대가 20명(58.8%)로 가장 많았고 이어서 30대가 10명(29.4%), 40대 4명(11.8%)순으로 나타났다. 학력은 대학교 졸업이 100%로 나타났다.

응답자의 전체 평균은 4.60점(표준편차=0.48)으로 나타나, 데이터를 기반으로 하는 기록관리 기준표의 자동 점검 방법이 전반적으로 만족도가 매우 높은 것으로 분석되었다. 특히 효율성(M = 4.76)이 가장 높게 나타났으며, 이는 수작업 대비 점검 소요 시간이 현저히 단축되고 업무 생산성이 향상되었음을 의미한다. 또한 정확성(M = 4.62)과 사용용이성(M = 4.55) 역시 높은 수준을 보여, 자동점검 결과에 대한 신뢰성과 시스템의 직관적 활용성을 입증하였다. 정보성(M = 4.47)은 다른 항목에 비해 약간 낮지만 여전히 높은 수준으로 평가되어, 결과의 세부 표현 및 시각화 기능 개선의 가능성을 시사한다.

설문에 참여한 수영구청 부서 기록물관리책임자들은 데이터를 기반으로 하는 기록관리 기준표 자동점검 기능이 문서등록대장과 기록관리 기준표 간의 불일치 항목을 정확히 식별하고, 유사 문서명 탐지와 분류체계 정비에 유용하였으며, 인적 오류를 최소화하고 업무 시간을 절감하는 데 크게 기여했다고 응답하였다. 특히 레벤슈타인 거리(Levenshtein distance) 기반의 문자열 유사도 계산 방법은 기존의 수작업 검토 방식보다 정확성과 일관성이 높다는 평가를 받았다. 응답자 다수는 자동점검 방법을 공공기록관리의 디지털 전환을 촉진하는 실무형 도구로 인식하였다.

4.2 시사점 및 기대효과

본 연구를 통해 다음과 같은 시사점을 도출하고 기대할 수 있다.

첫째, 생산된 전자기록물의 단위과제 편철 부 적합 여부를 자동으로 점검할 수 있었다. 기록물을 생산할 시 담당자별 표현 차이, 오타자, 띄어쓰기 차이 등을 넘어서는 객관적 유사도 기반으로 판단할 수 있는 것이다. 기록물제목과 단위과제명이 불일치하는 경우를 신속히 파악하여 오분류 가능성 최소화할 수 있다. 결과적으로, 생산기록물의 배열 정확성을 높이고 기록 생산자의 주관적 판단의 의존도를 낮춰 분류의 표준성을 확보할 수 있는 것이다.

둘째, 기록관리기준표 단위과제와 실제 문서 데이터 간 상관분석을 통한 기준표 점검 체계를 구축할 수 있었다. 기록관리기준표와 실제 업무에서 생산된 문서등록대장 데이터의 관계를 상관 분석을 통해 실제 업무에서 전혀 활용되지 않는 미사용 단위과제를 식별하여 자동으로 추출할 수 있다. 또한 기록관리기준표의 불필요한 항목을 정비하거나 구조에 대해서 개선이 가능하다. 유사하거나 동일한 단위과제명에서 서로 다른 보존기간이 부여된 경우 자동 확인할 수 있으며 기관 간·부서 간 기록관리의 정합성과 합리성 제고할 수 있다. 그리고 기록관리 기준표의 공개구분과 실제 기록 등록 시의 공개구분을 비교할 수 있다. 이는 곧 반복적 불일치 발생 시 기준표의 공개기준 재검토가 가능하고 법적·행정적 위험을 줄이고 예측 가능한 기록관리 체계 구축할 수 있는 것이다. 이로써 기록관리기준표를 정적·비순환형 관리체계가 아니라, 실제 업무 데이터 흐름을 반영하

여 지속적으로 개선되는 동적·순환형 관리체
계로 전환할 수 있는 것이다.

5. 결 론

본 연구는 행정업무 환경이 종이 기반에서
디지털 기반으로 전환되었음에도 불구하고, 여
전히 「공공기록물 관리에 관한 법률」이 ‘종이기
록의 전자적 관리’ 체계를 중심으로 운영되고 있
는 현실적 한계를 극복하고자 하였다. 이를 위해
본 연구는 공공기관의 문서등록대장 데이터를
활용하여 기록관리기준표를 자동으로 점검·적
용할 수 있는 방법론을 제안하였다.

첫째, 레벤슈타인 거리 알고리즘을 적용하여
문서등록대장에 등록된 기록물 제목과 기록관리
기준표의 단위과제에 포함된 기록물들 간의 문
자(키워드) 사이의 유사도를 정량적으로 비교
하였다. 문자열 유사도 기반 자동 분류 검증, 데
이터 상관 분석을 통한 기준표 점검 등의 접근으
로, 기록관리의 시작 단계인 분류의 어려움을 해
결하고, 기관별 기준표 운영의 지속가능성을 확
보하며, 나아가 ‘종이기록의 전자적 관리’라는
기존 패러다임을 실질적 디지털 기록관리 체계
로 전환하는 데 기여하였다.

둘째, 실제 업무 데이터를 반영하여 기준표를
유지·관리하는 새로운 방식은 기존 연구나 정

부 차원의 노력에서 부족했던 피드백 기반의 순
환형 기록관리 체계를 제공한다는 점에서 높은
실무적 가치를 가진다.

본 연구의 결과는 크게 세 가지로 요약할 수
있다.

첫째, 기록물의 자동분류는 내용, 구조, 맥락
을 반영해야하지만 기관에서는 기록물건 제목
이 단위과제명을 기능적으로 반영하는 관행이
존재하기 때문에 기관에서 누적된 데이터를
기반으로 실무적관점에서 높은 정확도를 보여
준다.

둘째, 사용자 평가를 진행한 결과 효율성과
용이성 항목에 대해서는 사용자들로부터 높은
결과를 얻었지만 정보성에서는 다소 부정적인
결과를 얻었다. 이는 생산된 전자기록물을 분류
하는 것이 과연 유의미한지에 대한 의문을 가지
고 있는 것이다. 향후 생산된 전자기록물의 심
층적연구를 통해 전자기록물의 분류업무에 대
한 실효성 연구를 진행하겠다.

셋째, 실무자들의 주관적인 의견이 대부분 적
용되었던 종이기반의 기록관리업무가 데이터를
기반으로 하는 기록정보관리업무로 사고의 확
장전환이 가능하다는 것을 의미한다. 법과 제도
를 개선하고 업무환경을 종이기반의 관리체계
에서 데이터를 기반으로 하는 관리체계로 전환
하는 것이 필요하며 기록관리를 수행하는 업무
담당자들도 역량을 강화할 필요가 있다.

참 고 문 헌

공공기록물관리에 관한 법률, 법률, 제21066호,
국가기록원 (2005). 기록관리혁신 실무매뉴얼. 대전: 국가기록원.

- 국가기록원 (2018). 기록분류체계 및 기록관리기준표 재설계. 대전: 국가기록원.
- 국가기록원 (2025). 기록물관리지침(공통메뉴얼). 대전: 국가기록원.
- 국가기록원 (n.d.). 기록관리업무소개.
출처: <https://www.archives.go.kr/next/newmanager/electronicConcept.do>
- 김다빈, 이은정, 류한조 (2021). 빅데이터 분석 방법론을 활용한 지방자치단체 단위과제 운영 지원도구 개발 연구. 기록학연구, 70, 85-117. <https://doi.org/10.20923/kjas.2021.70.085>
- 김명훈 (2024). 디지털 시대 기록의 폐기에 대한 재조명: 정보관리 기능의 발전 및 확대와 관련하여. 한국기록관리학회지, 24(1), 73-88. <https://doi.org/10.14404/jksarm.2024.24.1.073>
- 안광모, 김윤석, 김영훈, 서영훈 (2013). Levenshtein 거리를 이용한 영화평 감성 분류. 디지털콘텐츠학회논문지, 14(4), 581-587. <https://doi.org/10.9728/dcs.2013.14.4.581>
- 이명훈, 양형정, 김수형, 이귀상, 김선희 (2012). 간편영상에서 한글 인식 성능향상을 위한 가중치 기반 음소 단위 분할 교정. 한국콘텐츠학회논문지, 12(2), 105-115.
<https://doi.org/10.5392/JKCA.2012.12.02.105>
- 이승익, 설문원 (2017). 전자기록관리정책의 재설계에 관한 연구. 기록학연구, 52, 5-37.
<https://doi.org/10.20923/kjas.2017.52.005>
- 이중섭, 오상엽 (2013). 개선된 Levenshtein Distance 알고리즘을 사용한 어휘 탐색 시스템. 디지털융복합연구, 11(11), 367-372. <https://doi.org/10.14400/JDPM.2013.11.11.367>
- 임진술, 한희정, 오효정 (2023). 공공기관 기록물 분류체계 재정비를 위한 지능화 방안: L 기관 사례를 중심으로. 정보관리학회지, 40, 137-156. <https://doi.org/10.3743/KOSIM.2023.40.2.137>
- 조지영, 설문원 (2018). 복합적 기록 생산 환경에서의 평가를 위한 기록관리기준표 개선방안: A구 '공장설립 및 등록' 업무를 사례로. 기록학연구, 58, 147-190.
<https://doi.org/10.20923/kjas.2018.58.147>
- 현문수 (2021). 국가기록평가의 관점에서 본 환경 분야 기록관리기준표 분석. 기록학연구, 67, 139-203.
<https://doi.org/10.20923/kjas.2021.67.139>
- 홍덕용 (2021). 공공기록관리분야의 빅데이터 활용 방법과 시사점 제안. 한국기록관리학회지, 21(4), 1-18. <https://doi.org/10.14404/JKSARM.2021.21.4.001>
- 홍덕용, 허준석 (2023). 비정형 텍스트 데이터분석을 활용한 기록관리 분야 연구동향. 한국기록관리학회지, 23(4), 73-89. <https://doi.org/10.14404/JKSARM.2023.23.4.073>
- Yulianto, M., M., Arifudin, R., & Alamsyah, A. (2018). Autocomplete and spell checking levenshtein distance algorithm to getting text suggest error data searching in library. Scientific Journal of Informatics, 5(1), 75. <https://doi.org/10.15294/sji.v5i1.14148>

• 국문 참고문헌에 대한 영문 표기
(English translation of references written in Korean)

- Act on the Management of Public Records, Act No. 21066.
- Ahn, Kwang-Mo, Kim, Yun-Suk, Kim, Young-Hoon, & Seo, Young-Hoon (2013). Sentiment classification of movie reviews using Levenshtein Distance. *Digital Contents Society*, 14(4), 581-587. <https://doi.org/10.9728/dcs.2013.14.4.581>
- Cho, Ji-Young & Seol, Moon-Won (2018). Redesigning disposition authority template for the records produced in complex system environment: A case study of factory founding and registration activity in local government. *The Korean Journal of Archival Studies*, 58, 147-190. <https://doi.org/10.20923/kjas.2018.58.147>
- Hong, Deokyong & Heo, Junseok (2023). Research trends in record management using unstructured text data analysis. *Journal of Korean Society of Archives and Records Management*, 23(4), 73-89. <https://doi.org/10.14404/JKSARM.2023.23.4.073>
- Hong, Deokyong (2021). Big data utilization and policy suggestions in public records management. *Journal of Korean Society of Archives and Records Management*, 21(4), 1-18. <https://doi.org/10.14404/JKSARM.2021.21.4.001>
- Hyun, Moon Soo (2021). A study on the records management reference table of environmental authorities from the perspective of appraisal and selection of national archives. *The Korean Journal of Archival Studies*, 67, 139-203. <https://doi.org/10.20923/kjas.2021.67.139>
- Kim, Dabeen, Leem, Eunjung, & Ryu, Hanjo (2021). Research on development of support tools for local government business transaction operation using big data analysis methodology. *The Korean Journal of Archival Studies*, 70, 85-117. <https://doi.org/10.20923/kjas.2021.70.085>
- Kim, Myoung-hun (2024). Reconsideration on the disposal of records in digital era: Regarding the development and expansion of information management functions. *Journal of Korean Society of Archives and Records Management*, 24(1), 73-88. <https://doi.org/10.14404/jksarm.2024.24.1.073>
- Lee, Jong-Sub & Oh, Sang-Yeob (2013). Vocabulary retrieve system using improve Levenshtein Distance algorithm. *Journal of Digital Convergence*, 11(11), 367-372. <https://doi.org/10.14400/JDPM.2013.11.11.367>
- Lee, Myung-Hun, Yang, Hyung-Jeong, Kim, Soo-Hyung, Lee, Guee-Sang, & Kim, Sun-Hee (2012). Weighted disassemble-based correction method to improve recognition rates of

- Korean text in signboard images. *The Journal of the Korea Contents Association*, 12(2), 105-115. <https://doi.org/10.5392/JKCA.2012.12.02.105>
- Lee, Seung-eok & Seol, Moon-won (2017). A study of redesigning electronic records management policies. *The Korean Journal of Archival Studies*, 52, 5-37. <https://doi.org/10.20923/kjas.2017.52.005>
- Lim, Jinsol, Han, Hui-Jeong, & Oh, Hyo-Jung (2023). An intelligent approach for reorganization record classification schemes in public institutions: Case study on L institution. *Journal of the Korean Society for Information Management*, 40, 137-156. <https://doi.org/10.3743/KOSIM.2023.40.2.137>
- National Archives (2005). *Practice Manual for Record Management Innovation*. Daejeon: National Archives.
- National Archives (2018). *Redesignated the Record Classification System and Record Management Standards*. Daejeon: National Archives.
- National Archives (2025). *Archives Management Guidelines (Common Manual)*. Daejeon: National Archives.
- The National Archives of Korea (n.d.). *Introduction to Record Management*. Available: <https://www.archives.go.kr/next/newmanager/electronicConcept.do>