



Current Status and Development Trend of Applying Artificial Intelligence to Records Management in the Digital Age

Yi-Ling Wu, National Dong Hwa University
Thomas Kaluvi, National Dong Hwa University

Outline

- Introduction
- Current Status of Applying Artificial Intelligence to Records Management
 - Australia
 - United Kingdom
 - United States of America
- Contributions and Conclusions



Introduction

- **Background**

- Rapid development of artificial intelligence (AI)
- High volumes of records escalate

- **Motivation**

- Lack of persuasive cases of research in AI and Records Management (RM)
- Reluctant to adopt the AI technology in RM
 - Lack of training in artificial intelligence techniques
 - Risk aversion

Introduction (cont.)

- **Objective**

- To shed some light on progress made by developed countries in helping the records manager with artificial intelligence

Current Status of Applying Artificial Intelligence to Records Management

- Australia
- United Kingdom
- United States of America



Australia (1/3)

- **Whole-of-Government Digital Records Platform (National Archives of Australia, 2017)**
 - The Platform will automate the records management function taking into consideration user needs and technological advancements
 - Information is categorized, indexed, managed and disposed of in accordance with Australian Government records management obligations
 - Using technologies: cognitive computing, keyword extraction, machine learning and auto-indexing

Australia (2/3)

- **Australasian Digital Recordkeeping Initiative (Council of Australasian Archives and Records Authorities, 2014)**
 - The public records institutions of Australia and New Zealand face similar, if not identical, challenges in the digital era
 - ADRI was formed to articulate and promote a common approach to digital recordkeeping
 - The tools were installed and tested at the Public Record Office Victoria (PROV), State Records New South Wales
 - Production and management of digital files; preservation of archives
 - Online retention and disposal application, Australasian digital recordkeeping initiative glossary of disposal triggers

Australia (3/3)

- **Machine learning and Records Management (The New South Wales State Archives, 2018)**
 - There has been very little adoption of the machine learning technology
 - To explore the application of machine learning in records management
 - Scikit-learn: a free and open source machine-learning software
 - Used a corpus of unstructured data for testing of the retention or disposal mechanism
 - Assisting with the classification and disposal of unclassified unstructured data

United Kingdom (1/3)

- **The application of Technology- assisted review to born-digital records transfer (The National Archives, 2016)**
 - Born-digital records pose many challenges for government departments : records volume, sensitivity, disclosure, lack of structure
 - The National Archives conducted trials of eDiscovery software on 100,000 born-digital Records
 - Technology-assisted review using eDiscovery software can support government departments during appraisal, selection and sensitivity review
 - Topic modelling algorithm automatically detects groups of topics from the content of documents

United Kingdom (2/3)

- **Machine learning in the Archives (The National Archives, 2018)**
 - As we reach the late 90s and the new millennium the records will change dramatically
 - Training to uncover the mystery of machine learning
 - 35 members to develop system for automation of records
 - Non-technical team explored the ethical implications of the use of machine learning in archival practice
 - Identified two areas for further investigation: automated recognition of coding languages, and topic modelling catalogue descriptions

United Kingdom (3/3)

- **Machines reading the archive: Handwritten Text Recognition (HTR) software (The National Archives, 2018)**
 - This project focuses on our collection of handwritten court wills between 1384 and 1858 known as PROB 11 wills
 - They are not the easiest things to read
 - The National Archives has been running a pilot project to test out the feasibility of HTR software
 - Transkribus platform offers the potential to use computers to ‘read’ handwritten documents
 - The Transkribus software works by training a model on accurate transcriptions of documents
 - HTR software produces an automatic transcription
 - Had some good results from a model trained

United States of America (1/3)

- **Automated Electronic Records Management Report (National Archives and Records Administration, 2014)**
 - Managing Government Records Directive by the end of 2019
 - The Directive encourages NARA, agencies, and stakeholders to automate records management.
 - The processes and tools that agencies currently use to manage electronic records are not adequate to support consistent compliance
 - Automated tools for managing electronic records could reduce the recordkeeping burden
 - Machine learning or predictive coding as used in eDiscovery

United States of America (2/3)

- **Open Source Tools for Records Management (National Archives and Records Administration, 2015)**
 - Agencies must manage all permanent electronic records in an electronic format by 2019
 - Open source tools are generally free and available in a time of shrinking agency budgets
 - By December 31, 2014, obtain external involvement for the development of open source records management solutions
 - NARA identified open source tools that could be used for records management Tasks
 - NARA introduced efforts and tools to the Federal records management community

United States of America (3/3)

- **The Access Restriction Checker (Georgia Tech Research Institute, 2005)**
 - Review of Presidential electronic records for access restrictions is an intellectually demanding task
 - The increasing volume of electronic records from Presidential administrations
 - To develop an automated tool that could use this knowledge to support archivist's decisions in reviewing Presidential Records
 - We have begun prototyping such a tool, which we call the Access Restriction Checker.
 - The prototype illustrates the overall process emphasis has been on application of rule-based reasoning
 - Case-based reasoning and rule-based reasoning in decision making algorithm, and machine learning

Contribution and Conclusion

- **Contributions**

- Understanding Current status
 - Strategies, Initiatives Projects and research
- Research direction
 - Offering suggestions on Application of AI

- **Conclusions**

- Persuasive research in AI and records management needed
- Adoption the AI technology in Electronic records management
- Training in artificial intelligence techniques

Thank you

