



The Elusiveness of AI based Records Classification

[ITRUST AI Symposium @ Lanzarote
27 Oct 2022]



MALAYSIA TEAM:

Universiti Kebangsaan Malaysia:

Umi Asma' Mokhtar, Masnizah Mohd, and Siti

Norulhuda Sheikh Abdullah

Universiti Teknologi MARA:

Azman Mat Isa

National Archives of Malaysia

Jamilah Jamal

Malaysian Administrative Modernisation and
Management Planning Unit (MAMPU)

Mohd Haizad Samian

OVERVIEW

- The idea of using AI to facilitate records and archival work is not new.
- In 1987 - Hirtle wrote about "Artificial Intelligence, Expert Systems, and Archival Automation".
- In 1991 - Stielow compared the development of archival theory with the cutting edge of artificial intelligence, characterising both as "the search for the obvious."



- In 2019,
 - 44% of records professionals agreed that their organisations use “automated tools to locate and preserve relevant information,”
 - 49% are manually deleting emails, and
 - 58% are manually deleting records from mobile devices
- Organisations are drowning in records and information. From LR survey:
 - 71% of organisations have no idea of the content in their stored data [and]
 - 79% of organisations say too much time and effort is spent manually searching and disposing of information”.



TRENDS

- Managing records is a continuous activity.
- Evolvement of technology has changed the practice.
- From industrial – digital – AI revolutions, three broad trends:
 - The (ever) increasing volume of records created
 - The (ever) increasing dominance of structure data systems over unstructured data
 - The (ever) increasing ability to re-classify and re-aggregate all records in a records system



- As Rolan et al. put it, "Today's information environments have become a 'wild frontier', decentralised and fractured, and subject to pressures that include increasing data volumes, reliance on commercial and proprietary systems, and evolving forms of records and formats".
- There have been several AI projects focusing on aspects of records and archival work that using machine learning & deep learning
 - to identify patterns that yield insights
 - to classify the content of records
 - to create internal tags for keeping track of analytics (in IMS)
 - to simplify the process of identifying records and automates the retention schedule



- The advancements of document classifications that powered by AI able to manage documents in smarter ways which to help organisations for automatic routing, sentiment analysis, genre/language identification.
- For example, The Microsoft 365 ecosystem generated a huge volume of new content scattered across four main workloads – Exchange/Outlook, SharePoint, OneDrive and Teams. A few other systems such as Yammer also added to the mix.
- Most of this information was not subject to any form of classification in the recordkeeping sense. The Microsoft 365 platform included the ability to apply retention policies to content but there was a disconnect between classification and retention.



THE COMPLEXITY

- Most of the classification works on the content.
- The complexity of records classification lie on context & structure
- What are content, context, and structure?
 - Content: text, data, symbol, image, sound, graphic, and any information forming a record.
 - Context: 3 aspects are contextual information (e.g. digital signature), relationship from record to another record, and activity that create the record.
 - E.g via doc/info attached, network, reference/code/number or metadata
 - Structure: how records are formed including: format, symbol e.g. letter, memo, official email in public office; address, date, paragraph, and signature.





KETUA PENGARAH
ARKIB NEGARA MALAYSIA
ارکيب نگارا مليسيا
KEMENTERIAN PELANCONGAN DAN KEBUDAYAAN
MALAYSIA
JALAN TUANKU ABDUL HALIM
50568 KUALA LUMPUR.



Tel : 603-6209 0500 samb / ext.602
Faks : 603-6201 3100
Portal Rasmi: www.arkib.gov.my

Nama dan
alamat
penerima

No. Rujukan

Ruj. Kami : ANM.100-2/6/1V03(1)
Tarikh : 23 Oktober 2015

Tarikh di ruang atas
sebelah kanan selepas
No. Rujukan. Ruangan
tarikh yang kosong
hendaklah diisi semasa
menandatangani surat.

YBhg. Datuk Hj. Ab. Ghaffar Bin A. Tambi
Timbalan Ketua Setiausaha (Kebudayaan)
Kementerian Pelancongan dan Kebudayaan Malaysia
Pejabat Timbalan Ketua Setiausaha (Kebudayaan)
Ara 17, No. 2, Menara 1, Jalan P5/6
Presint 5
62200 PUTRAJAYA

Panggilan hormat
penerima

YBhg. Datuk,

Tajuk

JEMPUTAN KE MAJLIS PERASMIAN SAMBUTAN HARI WARISAN AUDIOVISUAL SEDUNIA 2015 ANJURAN ARKIB NEGARA MALAYSIA

Dengan segala hormatnya saya diarah merujuk kepada perkara di atas.

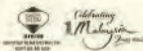
2. Sukacita dimaklumkan bahawa Jabatan Arkib Negara Malaysia bercadang untuk menganjurkan Majlis Sambutan Hari Warisan Audiovisual Sedunia 2015 pada **12 November 2015 (Khamis)** bertempat di **Bangunan Pusat Pendokumentasian dan Pemuliharaan Audiovisual Negara (PPPAV), Arkib Negara Malaysia, Jalan Tuanku Abdul Halim, Kuala Lumpur**. Majlis ini dianjurkan bersempena sambutan Hari Warisan Audiovisual Sedunia yang ditetapkan oleh UNESCO pada 27 Oktober setiap tahun. Tema sambutan pada tahun ini ialah *"Archives at Risk: Protecting the World's Identities"*.

3. Sempena majlis sambutan Hari Warisan Audiovisual Sedunia 2015 ini, Arkib Negara Malaysia akan menganjurkan Majlis Diskusi Ilmu: Pemeliharaan Bahan Audiovisual bagi memberi kesedaran kepada peserta tentang kepentingan pemeliharaan bahan audiovisual di samping mempertingkatkan perolehan, penyimpanan dan pendokumentasian bahan audiovisual yang bernilai kepada negara. Majlis turut disertai dengan acara Penyerahsipan bahan audiovisual, Pameran slaid kaca bertajuk *"Cerminan Kehidupan"*, sesi Tayangan Filem, Kem Fotografi serta Kem Animasi.

4. Untuk makluman YBhg. Datuk, program ini akan disempurnakan perasmianannya oleh **YB Datuk Mas Ermieyati binti Samsudin, Timbalan Menteri Pelancongan dan Kebudayaan Malaysia**.

...2/-

Arkib : Intipati Kepada "Masa"



Kepala surat
(Letterhead)

Penomoran
perenggan
1 kali Tab
selepas
nombor
perenggan

Isi Kandungan
Font: Arial,
Saiz: 12 - 14
Jarak: Single
hingga 1.5
Kedudukan
teks: Seimbang
rata kiri dan
kanan

Sambungan
ke muka surat
seterusnya
ditulis pada
sudut kanan
sebelah
bawah surat

ANM.100-2/6/1V03(1)

Nombor Rujukan surat atas
sebelah kiri setiap helaian surat

5. Sehubungan itu, Arkib Negara Malaysia amat berbesar hati ingin menjemput YBhg. Datuk untuk hadir ke Majlis Perasmian Sambutan Hari Warisan Audiovisual Sedunia 2015 sepertimana ketetapan berikut:

Tarikh : 12hb November 2015 (Khamis)

Masa : 9.00 pagi

Tempat : Bangunan PPPAV,
Arkib Negara Malaysia,
Jalan Tuanku Abdul Halim (Jalan Duta)
50568, Kuala Lumpur

Perasmi : **YB Datuk Mas Ermieyati binti Samsudin**
Timbalan Menteri Pelancongan dan Kebudayaan
Malaysia

Segala kerjasama dan perhatian pihak YBhg. Datuk berhubung perkara ini amatlah dihargai dan didahului dengan ucapan terima kasih.

Sekian, terima kasih.

"BERKHIDMAT UNTUK NEGARA"

Pengakhiran Surat
"Berkhidmat Untuk Negara"

Saya yang menurut perintah,

tt

(**AZEMI BIN ABDUL AZIZ**)
Ketua Pengarah
Arkib Negara Malaysia

Nama pengirim ditulis
dengan huruf besar dan
tebal

slsbaharjemputanhariAV2015MyDoc

Nama penyedia dokumen
Font: Arial, Saiz: 8

1 kali jarak

3 kali jarak
Single spacing

LAILA/khidmatnasihat/tdrive1

Pewujud surat
tajuk/perkara
lokasi simpanan pewujud

-2-
Muka surat di
bahagian tengah
bawah surat

InterPARES
TrustAI



- Records management will become a data science, overseeing algorithms that apply record classifications and/or record retention and access rules.
- To an algorithm everything is data. Algorithms, like humans, tend to understand data best when they view it in the context of its originating application. E.g.
 - Email is best understood within email systems, or within repositories that can replicate the structure and functioning of email systems.
 - No longer any necessity to move content out of one structured database (such as an email system) into another system.



- Organisations will have the technical possibility of having one overall structure/schema for organising records.
 - But this dream is likely to remain elusive due to the fact that data created within a structured dataset is usually much more meaningful and manageable within the structure of that dataset than it would be outside of it.
- In Malaysia, manual template for classification scheme is prepared by the NAM, for general administrative records (100 - 500).
- NAM worked with MAMPU since 2013 to develop digital document management system and include the file classification plan in it.



FILE CLASSIFICATION PLAN- NAM

- Classification File (Template)
 - 100 to 500: General
 - 600 onwards : Function-based

Code	Type
100	Administrative (including ICT, legal, IP, risk management, office management, ISO compliant, etc)
200	Land, Building, and Infrastructure
300	Asset and Storage
400	Finance
500	Human Resource
600	Function according to core business (Eg. NAM – Records Management, Mampu – Project Management for public sector)

CONTOH KLASIFIKASI FUNGSIAN SEKOLAH-SEKOLAH NEGERI PAHANG

Lampiran 6

KOD	FUNGSI/ AKTIVITI/ SUB- AKTIVITI	KOD	TRANSAKSI / FAIL	DESKRIPSI
600	TADBIR URUS SEKOLAH			Perkara-perkara berkaitan tadbir urus sekolah, dasar, peraturan, takwim, kurikulum, peperiksaan, pemantauan Pengajaran Dan Pembelajaran (P &P) dan naziran.
600-1	Dasar dan Peraturan			Perkara-perkara berkaitan penyediaan dasar, peraturan dan garis panduan.
600-1/1	Penyediaan Dasar dan Peraturan	600-1/1/1	Garis Panduan Disiplin	Perkara-perkara berkaitan penyediaan garis panduan disiplin, hukuman dan denda bagi kesalahan yang dilakukan mengikut kategori.
		600-1/1/2	Buku Panduan Peraturan Sekolah	Perkara-perkara berkaitan penyediaan buku panduan dan draf peraturan sekolah.
		600-1/1/3	Peraturan Asrama	Perkara-perkara berkaitan penyediaan buku panduan dan draf peraturan asrama.
600-2	Takwim Sekolah			Perkara-perkara berkaitan takwim dan jadual waktu kelas.
600-2/1	Jadual Waktu Persekolahan	600-2/1/1	Jadual Waktu Induk/Kelas	Perkara-perkara berkaitan penyediaan, penyemakan dan pelaksanaan Jadual Waktu induk/ kelas sekolah.
		600-2/1/2	Jadual Waktu Peribadi	Perkara-perkara berkaitan penyediaan Jadual Waktu peribadi setiap guru dan kakitangan.
		600-2/1/3	Jadual Waktu Kelas Bimbingan	Perkara-perkara berkaitan penyediaan Jadual Waktu Kelas Bimbingan/Tambahan sekolah.
		600-2/1/4	Jadual Waktu Anjal	Perkara-perkara berkaitan penyediaan dan pelaksanaan Jadual Waktu Anjal sekolah.
		600-2/1/5	Jadual Guru Gantian	Perkara-perkara berkaitan penyediaan Jadual Waktu Guru Gantian.



PANDUAN PENGURUSAN REKOD SEKTOR AWAM



100 PENGURUSAN PENTADBIRAN

100-1 Perundangan

- 100-1/1 Akta / Enakmen/ Ordinan / Rang Undang-undang / Undang-undang Kecil Agensi**
- 100-1/1/1 Penggubalan Akta/ Enakmen/ Ordinan /Rang Undang-undang/ Undang-undang Kecil
 - 100-1/1/2 Pindaan Akta/ Enakmen/ Ordinan /Rang Undang-undang /Undang-undang Kecil
- 100-1/2 Akta / Enakmen / Ordinan/ Rang Undang-undang / Undang-undang Kecil Luar Agensi**
- 100-1/2/1 Ulasan Atas Penggubalan Akta / Enakmen / Ordinan / Rang Undang-undang / Undang-undang Kecil Luar Agensi
- 100-1/3 Pekeliling Dan Peraturan Dalam Agensi**
- 100-1/3/1 Pekeliling Agensi

100-1/1/1

100: Function - Administration

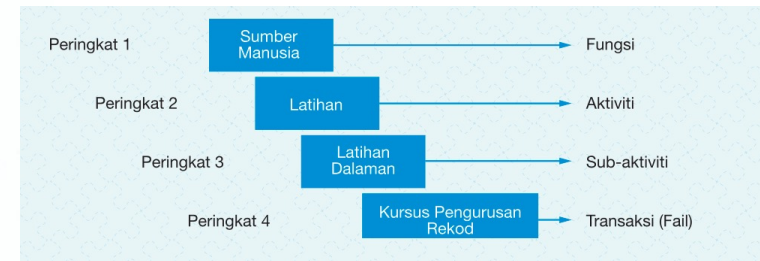
1/: Activity – Public Relations

1/: Sub Activity - Visit

1: Transaction (File) – Domestic Visit
- International Visit

TEMLAT KLASIFIKASI FAIL URUSAN AM (SIRI 100 - 500) SEKTOR AWAM

ARKIB NEGARA MALAYSIA 2018



MANUAL PROCESS

1. Preliminary discussion
2. Workshop execution
3. Proposed draft by department
4. Refinement of the draft (NAM)
5. Review or update (dept)
6. Approval

BIL	PERKARA	BULAN PERTAMA				BULAN KEDUA				BULAN KETIGA				BULAN KEEMPAT				BULAN KELIMA				BULAN KEENAM			
		1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
1.	Perbincangan Awal																								
2.	Pelaksanaan Bengkel																								
	2.1 Penyediaan Draf KF Urusan Am dan Fungsian																								
	2.2 Penyediaan Draf Deskripsi KF Urusan Am dan Fungsian																								
	2.3 Pengharmonian Draf KF Urusan Am dan Fungsian																								
3.	Jabatan kemukakan draf yang telah dipersetujui oleh Pengurusan Atasan Jabatan: 3.1 KF Urusan Am dan Fungsian 3.2 Deskripsi KF Urusan Am dan Fungsian																								
4.	Pengharmonian Peringkat Arkib Negara Malaysia																								
5.	Pindaan dan Pengemaskinian oleh Jabatan																								
6.	Kelulusan dan Perakuan kepada Jabatan																								

6 months preparation, at least



DIGITAL DOCUMENT MANAGEMENT SYSTEM- MAMPU



DIGITAL DOCUMENT MANAGEMENT SYSTEM

User Name

Password

Rekod Rasmi

Request OTP (Email)

Login

Forgot Password

© 2018 Hak Cipta Kerajaan Malaysia



MEJA BANTUAN DDMS 2.0



<https://helpdesk.ddms.malaysia.gov.my>



helpdeskdms@mampu.gov.my



03-61512966 / 03-88725117

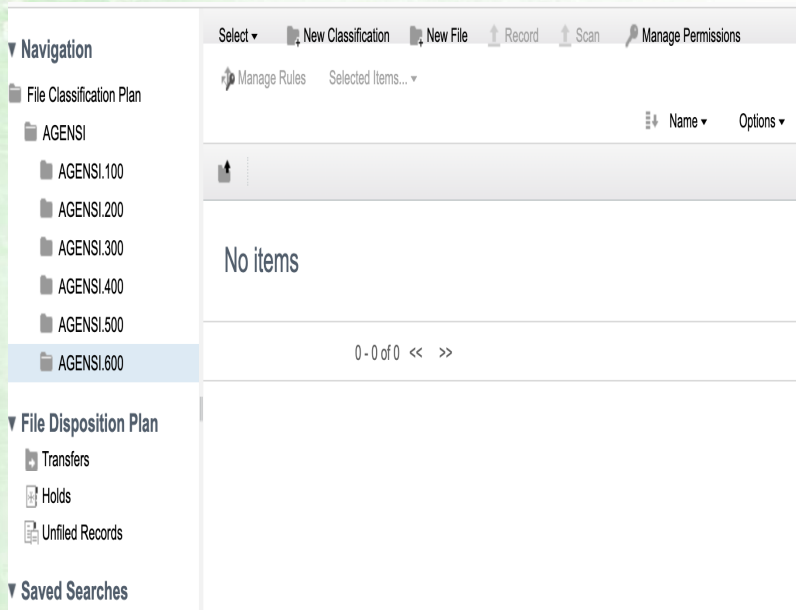
InterPARES
TrustAI



- The basic objectives of the DDMS 2.0 System are to:
 - replace the use of physical records with digital records;
 - minimise the use of paper;
 - enable digital records and physical records to be stored and maintained electronically; and
 - enable all government affairs such as presentation materials, internal circulars and letters between agencies/departments are managed or made electronically.



About the System



- Classification is according to FAT, and records are created in transaction folder.
 - Note: no creation of records (yet) in version 2.0. Records are created outside of the system.
- Using Alfresco as vendor, but fully maintained by the MAMPU
- Security – using Information Right Management (IRM).
 - So far, no incidents and attempt yet. Any classified documents must use IRM to open the document. Restricted-confidential-secret-top secret
- No AI-based technology used in the system.

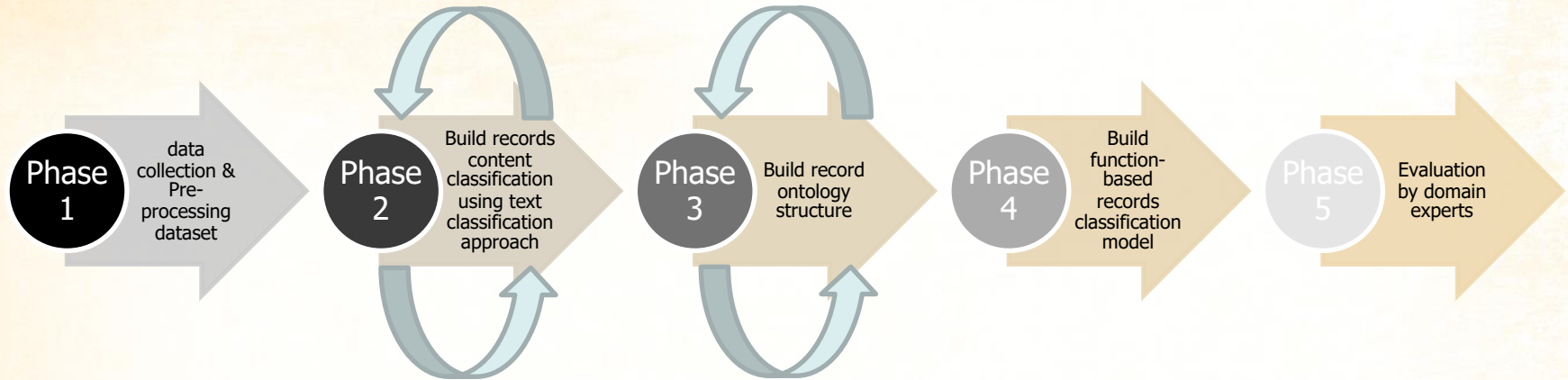


METHODOLOGY

- This project proposes a model that can automatically classify a record based on function.
- The function-based record classification model requires content, structure, and context.
- Thus, to perform an automatic function-based record classification model, tools like automatic text classification (for the content) and record ontology structure (structure and context) are needed to extract the content, structure, and context from a record.



5 PHASES



TAKE AWAY

- Classification is the "sleeping beauty of information science," and classification is indeed an indispensable technique for record-keeping (Bowker & Star, 1991).
- No matter how advanced the technology is, we still believe that we can protect the trust of people through records. Therefore, it is crucial to maintain records trustworthiness.
- To maintain the trustworthiness, it must be implemented according to the right theory, in this sense we refer to Archival Science Theory and Diplomatic.



Acknowledgement

This study draws on research conducted
as part of InterPARES Trust AI and
supported by the Social Sciences and
Humanities Research Council of Canada.

THANK YOU

InterPARES
TrustAI

