



FACULTY OF INFORMATION MANAGEMENT
UNIVERSITI TEKNOLOGI MARA
ICM561 WEB ARCHIVING
SEMESTER 20254 (OCT 2025 – FEB 2026)

ASSIGNMENT 1: Case Study (30%)	
<i>Individual Assignment</i>	
Topic	Chapter 5 Archiving Website
Description	<p>Plan and design for archiving website to propose resourceful archive content</p> <p>Course Outcome:</p> <ul style="list-style-type: none"> • Discuss the requirements and problems of web archiving (A3) <p>Language: English</p>
Instructions & Content	<ol style="list-style-type: none"> 1. the Web Content related about: <ul style="list-style-type: none"> • CDIM2634A – The Building Architecture Evolution in Malaysia/other country • CDIM2634B - The Evolution of Malaysian Literature Online • CDIM2634C - The Malaysian Oil & Gas Industry • CDIM2634D - Banks and Banking Corporate Memory • CDIM2634E- Web Content about Historic Railway and Station Heritage in Malaysia • CDIM2634ST1 - The Web Content about Urban Heritage: Old Shophouses and Street Culture 2. Provide some initial or preliminary study data and evidence on the requirement of web archiving implementation 3. Criteria to proposed: <ul style="list-style-type: none"> • Selection of website, website authorization and owner (Seed list) (introduction of the WA Title) Table 1: https://docs.google.com/spreadsheets/d/16J1oxJ7dapTpIEjY03j0hMRiM5sQp0BI/edit?gid=630243912#gid=630243912 • Problem statement (why the selected WA title is crucial or important to archive) • WA objective (1. state one objective arising issue or important reason to archive the website contents (relate with problem statement) and 2. State another objective that

	<p>relate with web archiving capabilities in preserving the web content</p> <ul style="list-style-type: none"> • WA Scope of content (in term of Years or duration, website coverage within Malaysia or Asian or open in any country in the world, keywords used) • Preliminary study of website selection: Website Name & URL, Web site type, Content Management System, Operating System & Servers, SSL Certificates, No of indexed pages, No of crawled pages and Web Content Sizes (may use whatever online tools such as https://builtwith.com and https://www.xml-sitemaps.com/) (used table 2) • Discussion of Web archive preliminary findings above <ul style="list-style-type: none"> i. The success rate of crawling ii. Do security feature affect crawling process iii. Do web content size and type of media affect crawling process • Website Archive Exploration (used table 3) • Discussion on website archive finding <ul style="list-style-type: none"> i. How many websites URL achieve to archived using HTTRACK and Conifer ii. How many useful and valid content related to topic and worth for historical & Cultural web presentation iii. How many websites content authorization from local or outside Malaysia • Analyze how sustainable and preservable the web content is over time — i.e., whether the digital materials related to the theme are likely to remain accessible, authentic, and usable in the future. • Conclusion (what significant findings after the archive completed, limitations found, and proposed of future study) <p>4. Presentation of case study report should be on word processing format and complete with assignment formatting such as:</p> <ul style="list-style-type: none"> • Each reading response must be typewritten on one side only, on white A4 paper, using Microsoft Word, 1.5 line spacing, ARIAL font size 11 • Must have proper numbering, page number, justify. • Tables and diagrams (if any) should be properly captioned

1. Website infrastructure and platform identification (<https://builtwith.com/> & <https://www.xml-sitemaps.com/>)

Table 2 Website Infrastructure and Platform

Tools used:			https://builtwith.com			https://www.xml-sitemaps.com/		
Website Name	URL	Web site type	Content Management System	Operating System & Servers	SSL Certificates	Processing time	No of indexed / crawled pages	Web Content Sizes
MTA-UITM MOU EXPANDS OPPORTUNITIES	http://mrem.bernama.com/views/m.php?idm=42522	News Website		Apache 2.2 & FreeBSD	Open SSL	0:04:32s	500 (crawled: 504)	10.13 Mb

*Explain how you experienced all these attributes above affect your crawling process.

a. Website Archive Exploration

Table 3 Web archive content

Website Name	URL	Website Scope (Malaysia/ Asian/ Other country)	Topic name	HTTRACK				Conifer/ WarCreate			
				archive success (YES/NO)	Duration (start date/time -end date/time)	size of file archived	No of file archi ved	archiv e succe ss (YES/ NO)	start date/t ime	size of file archi ved	No of file archi ved
1. UTM Expert	https://expert.utm.edu.my/profile.php?id=tUZjy8caI5kSUSvpNOQbWBDHxd6xGyUGBeIMao53CXI=	Malaysia	Farrah Diana Profile	NO	NIL	NIL	NIL				
2. Harian Metro	https://www.hmetro.com.my/mutakhir/2016/03/119533/menjejak-kehilangan-sudirman	Malaysia	Menjejak 'kehilangan' Sudirman	YES	NIL	NIL	NIL				
3. Dr. Alwi Mohd Yunus Profile	https://fim.utm.edu.my/index.php/component/content/article/45-curriculum-vitae/170-alwi-mohd-yunus	Brunei	Dr. Alwi Mohd Yunus Curriculum Vitae	YES							