Vidyavardhini's College of Engineering and Technology

Department of Artificial Intelligence & Data Science

AY: 2024-25

Class:	TE	**Semester:**	V
Course Code:		**Course Name:**	DWM

Name of Student:	Sainath Khot
Roll No. :	20
Assignment No.:	6
Title of Assignment:	
Date of Submission:	
Date of Correction:	

Evaluation

Performance Indicator	Max. Marks	Marks Obtained
Completeness	5	
Demonstrated Knowledge	3	
Legibility	2	
Total	10	

Performance Indicator	Exceed Expectations (EE)	Meet Expectations (ME)	Below Expectations (BE)
Completeness	5	3-4	1-2
Demonstrated Knowledge Legibility	3	2	1
Legibility	2	1	0

Checked by

Name of Faculty :

Signature :

Date :

DWM (6)

Q1. Apply your knowledge of a web crawler in context mining. Describe the basic component of a web crawler & explain how it decide which page to visit first what are 2 common technique it use to avoid visiting the same pages multiple times?

→ A web crawler, also known as a spider or bot, is an automated program that systematically browses the web to collect & index information from web pages for various purposes, such as search engine & data analysis.

Basic component:

- URL frontier: A queue that manages URL to be visited, prioritizing based on criteria like relevance or freshness.
- Downloader: Fetches web page content from the URLs in the URL frontier
- Parser: Analyzes page to extract text links and metadata while identifying new URL to add to the frontier
- Storage: Saves crawled data in a structured format, usually in a database.

Page Prioritization: Crawlers use strategies like Breadth-first search (BFS) or Priority base to decide which page to visit first.

Avoiding Page Revisit: - URL De-duplication: Keeps a list of visit URL's to avoid revisiting

- Canonicalization : Ensure different versions of the same page are identified and only the canonical version is crawled.

Q2) Explain what web usage mining and how it can be beneficial for a website. Provide an example of how web pages usage mining can help improve user experience on an e-commerce site. What type of data would you analyze to gather insights from user behaviour?

→ Web usage mining analyzes user behaviour from web logs to identify patterns that improve user experience & site performance

Benefits: 1) Personalization, adapts content & recommendations based on user preference

2) Performance Improvement: Helps resolve issue like slow-loading pages that affect engagement.

E-Commerce Example: Analyzing user navigation and click pattern can identify bottlenecks in the purchase process, such as users abandoning their carts at the shipping stage. This insight can lead to optimizing the checkout experience.

Key Data Types:- ① Access Logs :- show visits of pages ② Clickstream Data : Detailed records of user clicks analyzing these helps e-commerce sites enhance user experience and boost conversion rates.