

Web Mining

*- Discovering interesting and useful information
from Web content and usage data*

What is Web Mining?

- Web mining is a data mining technique to extract knowledge from **web data**.
- Web data includes :
 - **web documents**
 - **hyperlinks between documents**
 - **usage logs of web sites**
- The **WWW** is huge, widely distributed, global information service centre and, therefore, It is a rich source for data mining.

World Wide Web

- ✓ There are about **1.5 billion websites**.
- ✓ But less than **400 million are active**.
- ✓ Grows at about 1 million pages a day
- ✓ By the time you finish this class, **thousands of new sites will spawn**.
- ✓ Lots of duplication (70-80%)

What is the most visited site in the world?

Go ahead — **Google** it!

Fun fact:

More than 80% of all Google searches are initiated by the Google staff, in the process of developing and refining its search algorithms.

How Many **Websites** Are There in the World

1,179,448,021

Estimations show that there are around **1.18 billion** websites in the World at the moment. **17%** of these websites are active, **83%** are inactive.

200,756,193

websites are active

252,000

new websites are
created **every day**

10,500

new websites are
created **every hour**

175

new websites are
created **every minute**

3

new websites are
created **every second**

2,000+

new websites by the time
you are done reading this
article

World Wide Web

- Diverse types of data
 - Text
 - Images
 - Audio & Video



Black Goose Bistro

The Restaurant

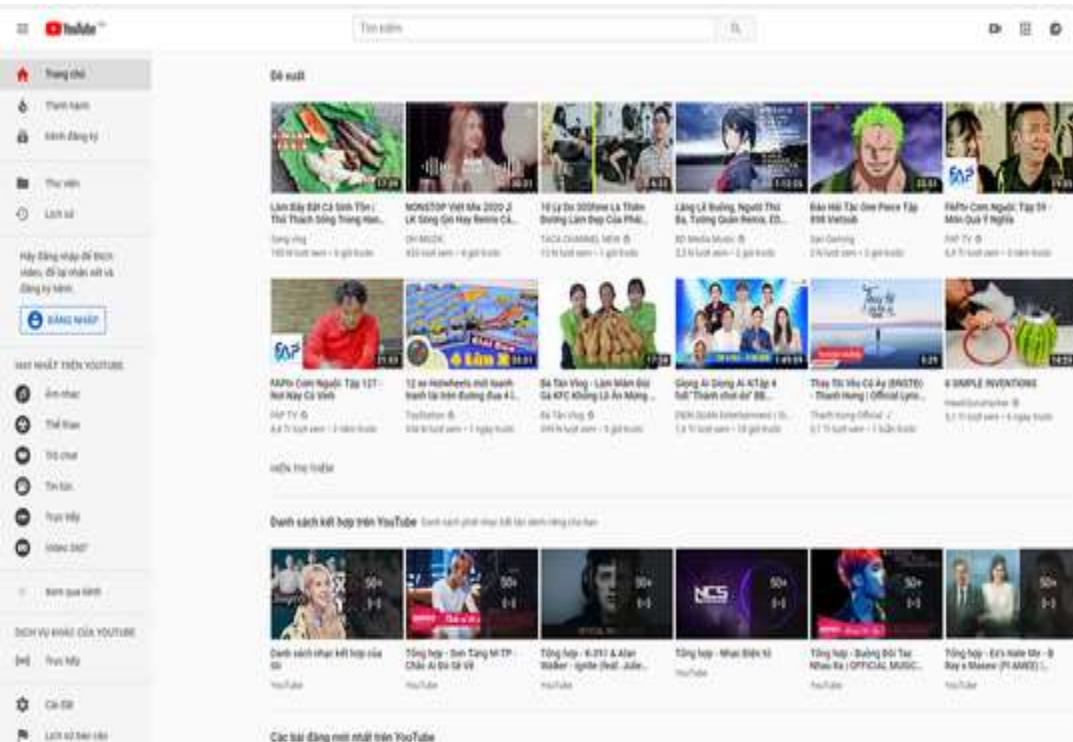
The Black Goose Bistro offers casual lunch and dinner fare in a hip atmosphere. The menu changes regularly to highlight the freshest ingredients.

Catering Services

fun... we'll do the cooking. Black Goose Catering can handle events from snacks for to elegant corporate fundraisers.

Hours and Hours

Massachusetts;
rough Thursday 11am to 9pm,
Saturday, 11am to midnight



YouTube

Tìm kiếm

Trang chủ

Thư viện

Lịch sử

Hãy đăng nhập để theo dõi kênh, để lại nhận xét và đăng ký kênh.

Đăng nhập

Đề xuất

Lâm Tây Bật Cá Linh Tinh
1:24

MUSIC Việt Nam 2020
1:24

10 lý do để bạn là một người yêu
1:24

Lâm Tây Bật Cá Linh Tinh
1:24

Bản Hát Tài Năng Tập 1
1:24

MUSIC Việt Nam 2020
1:24

Danh sách kết hợp trên YouTube

Các tài khoản mới trên YouTube

Web Mining

- Web mining is **the application of data mining techniques to discover useful information from the World Wide Web.**
- It uses automated methods to extract both **structured** and **unstructured** data from web pages, server logs and link structures.

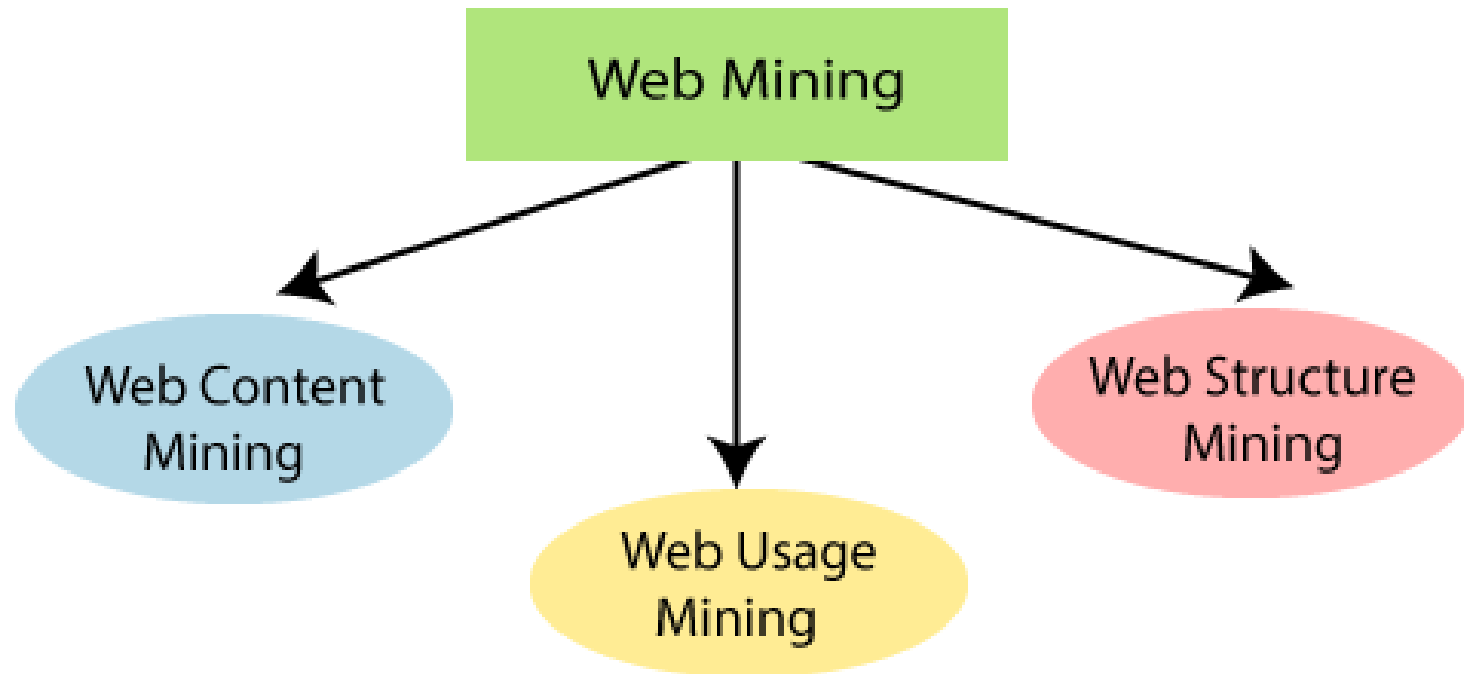
Web Mining

Examples:

- **Web search**, e.g. Google, Yahoo, Bing, Dogpile ,Duck Duck Go Ecosia ,Gigablast ,...
- **Specialized search**: e.g. *Squool Tube* - Search for factual, educational videos., *Elephind* - search the world's historical newspaper archives.
- **eCommerce** : e.g. Amazon,Flipkart,eBay,Fiverr,Upwork.
- **Advertising**, e.g. Google Adsense
- Improving Web site design and performance

Web Mining

Types of Web Mining



Web Content Mining

- **Web content** mining can be used to extract useful data, information, knowledge from the web page content.
- In web content mining, **each web page** is considered as an individual document.
- The primary task of **content mining** is data extraction, where **structured** data is extracted from **unstructured** websites.

Web Content Mining

- Web content mining can be utilized to distinguish topics on the web.

For Example, if any user searches for a **specific book** on the search engine, then the user will get a list of suggestions.

- The technologies that are normally used in web content mining are **NLP** (Natural language processing) and **IR** (Information Retrieval).
- Techniques in Web Content Mining :
 - Classification
 - Clustering

Web Content Mining

data mining Book - Google Search

https://www.google.co.in/search?q=data+mining++Book&ei=hrS5YafiAZmm2roPiaS3uAQ&ved=0ahUKEwinvcIWvuX0AhUZk1YBHQnSDUcQ4dUDCA4...

Google

data mining Book

X

Settings

Sign in

https://data-flair.training › ... › Data Mining Tutorials

15 Best Data Mining Books To Learn Data Mining - DataFlair

Top **Data Mining Books** - 1. Introduction to Data Mining - 2. An Introduction to Statistical Learning: with Applications in R · 3. Data Science for Business: What ...

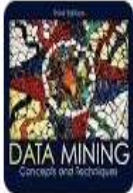


http://myweb.sabanciuniv.edu › files › 2016/02 PDF

Data Mining. Concepts and Techniques, 3rd Edition (The ...

The **book** gives quick introductions to database and **data mining** concepts with particular emphasis on data analysis. It then covers in a chapter-by-chapter ...

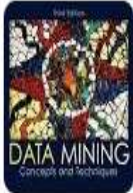
740 pages



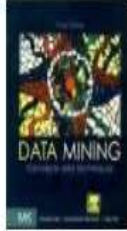
https://www.amazon.in › Data-Mining-Concepts-Techni...

Buy Data Mining: Concepts and Techniques (The Morgan ...

Data Mining: Concepts And Techniques (The Morgan Kaufmann Series In Data Management Systems) explains all the fundamental tools and techniques involved in the ...



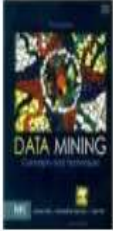
Ads · Shop data mining Book



Data Mining - Concepts A...

₹594

Flipkart



Data Mining: Concepts A...

₹625

MeriPustak...
Free delivery



Data Mining: Concepts a...

₹295

Amazon.in

→

More on Google

Web Usage Mining

- Web usage mining is the application of identifying or discovering interesting **usage patterns** from large data sets.
- And these patterns enable you to understand the user behaviors or something like that.
- Web usage mining is used for mining the web log records (access information of web pages) and helps to discover the user access patterns of web pages.
- **Web server** registers a web **log entry** for every web page.

Web Usage Mining

- The main source of data here is **Web Server** and **Application Server**.
- Log files are created when a user/customer interacts with a web page.
- Techniques in Web Usage Mining :
 - Association Rules
 - Classification
 - Clustering

Web Usage Mining

Advantage:

- This technology has enabled **e-commerce** to do **personalized marketing**, which eventually results in higher trade volumes.

Disadvantage:

- This technology when used on data of **personal nature** might cause concerns. The most criticized **ethical issue** involving web usage mining is the invasion of **privacy**.

Web Structure Mining

- Web structure mining is the application of discovering **structure information** from the web.
- The structure of the web **graph** consists of web pages as **nodes**, and hyperlinks as **edges** connecting related pages.
- Structure mining basically shows the **structured summary** of a particular website. It identifies **relationship** between web pages linked by information or direct link connection.
- Techniques in Web Structure Mining :
 - Association Rules
 - Classification

Web Structure Mining

Example:

- The most important application in this regard is the **Google search** engine, which estimates the ranking of its outcomes primarily with the PageRank algorithm.
- The rank of a page is decided by the number and quality of links pointing to the target node.
- In general, Web structure mining can be very useful to companies to determine the connection between two commercial websites.

Web Structure Mining

books on data mining - Google

https://www.google.co.in/search?q=books+on+data+mining&ei=o7W5YfPrO8LI2roPm6C_uAQ&oq=Book&gs_lcp=Cgdn d3Mtd2I6EAM YADIFCAAQQL...

books on data mining

Sign in

15 Best Data Mining Books To Learn Data Mining - DataFlair

Top **Data Mining Books** · 1. Introduction to **Data Mining** · 2. An Introduction to Statistical Learning: with Applications in R · 3. Data Science for Business: What ...

https://www.quora.com › What-is-the-best-book-for-lea...

What is the best book for learning data mining? - Quora

18-Mar-2016 · 4 answers

1- Introduction to **Data Mining** [Introduction to **Data Mining**]. This **book** is very helpful for beginners in **Data Mining**. · 2- **Data Mining** Concepts and Techniques [...

Which is the best **book** to learn **data mining**? - Quora 5 answers 15 Jun 2015

Which are the best websites, and the best **books** to learn ... 2 answers 10 Sept 2014

Is there any **book** about data preparation for **data mining**? 1 answer 30 Oct 2018

Where can I find **books**/documents on Orange **Data** ... 2 answers 15 May 2017

More results from www.quora.com

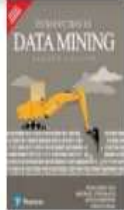
https://www.jigsawacademy.com › blogs › data-science

15 Best Data Mining Books in 2021 - Jigsaw Academy

3. Data Science for Business: What you need to know about **data mining** and data-analytic thinking ... This **book** focuses on visualising the business problems data- ...

https://www.amazon.com › bestsellers › books

TRUSTWORTHY



Introduction to Data Mining...

₹778

Amazon.in

Free delivery



Data Mining: Concepts...

₹597.55

MeriPustak...

Free delivery



Data Mining: Introductory...

₹200

Amazon.in

Free delivery



Data Mining & Warehousin...

₹175

Amazon.in

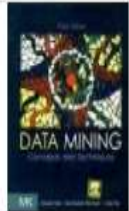


Data Mining Techniques

₹199

Was ₹380

Amazon.in

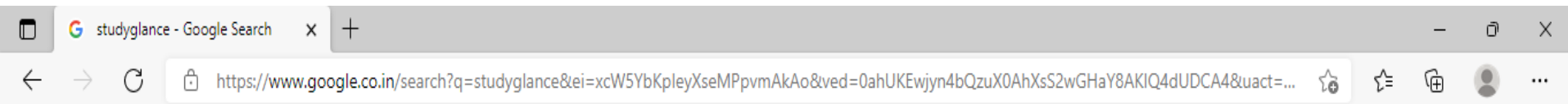


Data Mining - Concepts A...

₹594

Flipkart

Web Structure Mining



studyglance



Sign in

All

Maps

Videos

Images

News

More

Tools

About 58,300 results (0.33 seconds)

https://studyglance.in

Study Glance | Tutorials, Power Point Presentations, Lecture ...

Study Glance provides Tutorials , Power point Presentations(ppts), Lecture Notes, Important & previously asked questions, Objective Type questions, ...

Python Programming Lab ...

Study Glance provides Tutorials , Power point Presentations ...

Lab Programs

Study Glance provides Tutorials , Power point Presentations ...

Lecture Notes

Study Glance provides Tutorials , Power point Presentations ...

HTML Menu

Hyper Text Markup Language (HTML) is a simple and basic ...

[More results from studyglance.in »](#)

https://in.pinterest.com › studyglance

study glance (studyglance) – Profile | Pinterest

Activate Windows
Go to Settings to activate Windows.