**﻿Module 4 : Data Analytics**

Syllabus:

| Lecture no | Content | Duration (Hr) | Self-Study (Hrs) |
| --- | --- | --- | --- |
| 1 | Web analytics tools and techniques | 1 | 1 |
| 2 | Click stream analysis | 1 | 1 |
| 3 | A/B testing | 1 | 1 |
| 4 | Online surveys | 1 | 1 |
| 5 | Use of Google Analytics | 1 | 1 |
| 6 | Web crawling | 1 | 1 |
| 7 | Indexing | 1 | 1 |
| 8 | Natural Language Processing Techniques for Micro-text Analysis | 1 | 1 |
| 9 | Natural Language Processing Techniques for Micro-text Analysis | 1 | 1 |

**Theoretical Background:**

Web analytics tools and techniques, including click stream analysis, A/B testing, online surveys, and Google Analytics, are grounded in statistical analysis, experimental design, and survey methodologies. Web crawling and indexing employ algorithms from information retrieval and data structures. Natural Language Processing (NLP) techniques for micro-text analysis rely on linguistics, machine learning, and deep learning concepts. These methods empower businesses to understand user behavior, optimize websites, and extract insights from data, forming the foundation for data-driven decision-making in the digital realm.

**Lecture 1**

**Learning Objective: Learners shall be able to apply knowledge in web analytics tools and techniques.**

**Web Analytics Tools and Techniques:**

Web analytics involves the collection, measurement, analysis, and reporting of web data for the purposes of understanding and optimizing web usage.

Web analytics tools and techniques draw upon statistical analysis and data mining methodologies. These tools help businesses and website owners understand user behavior, preferences, and interactions. By analyzing this data, businesses can make informed decisions to enhance user experience and improve their online presence.

**What is web analytics?**

Web analytics is the process of analyzing the behavior of visitors to a website. This involves tracking, reviewing and reporting data to measure web activity, including the use of a website and its components, such as web pages, images and videos.

Data collected through web analytics may include traffic sources, referring sites, page views, paths taken and conversion rates. The compiled data often forms a part of customer relationship management analytics ([CRM analytics](https://www.techtarget.com/searchcustomerexperience/definition/CRM-analytics)) to facilitate and streamline better business decisions.

Web analytics enables a business to retain customers, attract more visitors and increase the dollar volume each customer spends.

Analytics can help in the following ways:

Determine the likelihood that a given customer will repurchase a product after purchasing it in the past.

Personalize the site to customers who visit it repeatedly.

* Monitor the amount of money individual customers or specific groups of customers spend.
* Observe the geographic regions from which the most and the least customers visit the site and purchase specific products.

Predict which products customers are most and least likely to buy in the future.

**Web analytics tools**

* Web analytics tools report important statistics on a website, such as where visitors came from, how long they stayed, how they found the site and their online activity while on the site. In addition to web analytics, these tools are commonly used for [product analytics](https://www.techtarget.com/whatis/definition/product-analytics), [social media analytics](https://www.techtarget.com/searchbusinessanalytics/definition/social-media-analytics) and [marketing analytics](https://www.techtarget.com/searchcustomerexperience/definition/marketing-analytics).

Some examples of web analytics tools include the following:

* Google Analytics. Google Analytics is a web analytics platform that monitors website traffic, behaviors and conversions. The platform tracks page views, unique visitors, bounce rates, referral Uniform Resource Locators, average time on-site, page abandonment, new vs. returning visitors and demographic data.
* Optimizely. [Optimizely](https://www.techtarget.com/searchcustomerexperience/news/252498293/Optimizely-shore-up-DX-platform-with-Zaius-CDP-acquisition) is a customer experience and A/B testing platform that helps businesses test and optimize their online experiences and marketing efforts, including conversion rate optimization.
* Kissmetrics. Kissmetrics is a customer analytics platform that gathers website data and presents it in an easy-to-read format. The platform also serves as a [customer intelligence](https://www.techtarget.com/searchbusinessanalytics/definition/customer-intelligence-CI) tool, as it enables businesses to dive deeper into customer behavior and use this information to enhance their website and marketing campaigns.
* Crazy Egg. Crazy Egg is a tool that tracks where customers click on a page. This information can help organizations understand how visitors interact with content and why they leave the site. The tool tracks visitors, [heatmaps](https://www.techtarget.com/searchbusinessanalytics/definition/heat-map) and user session recordings.

*Let’s check the take away from this lecture*

Exercise:

1. What are the key metrics commonly tracked by web analytics tools like Google Analytics, and how do these metrics help in understanding user behavior and website performance?
2. Imagine you are managing an online store. Describe a scenario where you could use A/B testing to improve your website's conversion rates.

**Learning from this lecture:** Learners should be able to understand the processing and Visualizing Data.

**Lecture 2**

**Learning Objective: Learners shall be able to understand concepts behind click stream analysis and its usage.**

Click Stream Analysis:

Click stream analysis involves tracking the sequence of user interactions (clicks) as they navigate through a website or online platform. It helps in understanding user behavior patterns, preferences, and engagement levels. By analyzing the clickstream data, businesses can gain insights into:

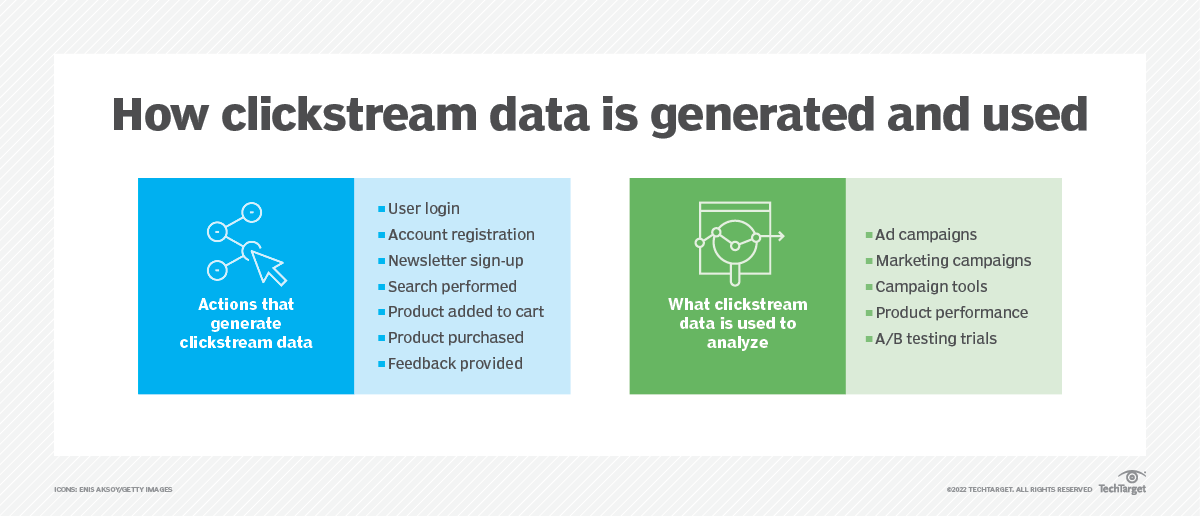
* Most visited pages: Identify popular pages that attract more user attention.
* Conversion paths: Understand the flow of users from landing pages to conversion points.
* Drop-off points: Identify where users leave the website without completing desired actions.
* User journey: Visualize the path users take to explore content and services.

**What is clickstream data?**

Clickstream data and clickstream analytics are the processes involved in collecting, analyzing and reporting [aggregate data](https://searchsqlserver.techtarget.com/definition/data-aggregation) about which pages a website visitor visits -- and in what order. The path the visitor takes through a website is called the clickstream.

Clickstreams are categorized into clickstream data and clickstream analytics, which is also referred to as clickstream analysis. The clickstream data is the information collected about a user while they browse through a website or use a web browser. Clickstream analytics is the process of tracking, analyzing and reporting data on the pages a user visits and user behavior while on a webpage.

Websites use clickstream data to show how a user progressed from an initial search or landing page to buying an item or service. [Search engines](https://www.techtarget.com/whatis/definition/search-engine) use clickstream data sets to show where a user has searched for a term, when they have clicked on it and if they go back to searching after this. [Internet service providers](https://www.techtarget.com/whatis/definition/ISP-Internet-service-provider), advertising networks, and IT and telecom organizations also collect clickstream data.



**Clickstream data metrics**

Clickstream data includes the following types of [web analytics](https://www.techtarget.com/searchbusinessanalytics/definition/Web-analytics) information:

* whether the individual is a unique or repeat visitor to the website;
* the terms an individual plugs into a search engine;
* what page the individual lands on first;
* the amount of time a user spends on a page;
* the features on the page the user clicks on and engages with;
* when and where an item is added or removed from a cart;
* where the user goes next; and
* when the back button is used.

Clickstream data collected from a single session of a user interacting with a website may not be useful. However, an organization can use aggregate data gathered from many visitors to improve its website or service.

For example, if a lot of visitors leave a site after landing on a page with too little information, the organization may need to enhance the page with more valuable information. Likewise, if visitors often land on a page that isn't the website's [homepage](https://www.techtarget.com/whatis/definition/home-page), then the organization may want to redesign that page to be more inviting and informative to users.

Clickstream data does not include personal details about a user, and it is typically stored on the server that supports the website. Clickstream data is a [useful addition](https://www.aunalytics.com/what-can-clickstream-data-tell-you-about-your-business/) to data from Google Analytics.

**Clickstream analytics and how it is used**

Organizations use clickstream analytics to uncover trends and draw conclusions from different metrics about their websites. This process typically uses a web server [log file](https://www.techtarget.com/whatis/definition/log-log-file) to monitor user activity on a website.

Using the clickstream analysis, an organization can collect data on the number of page visits, views, and unique and repeat visitors. This data provides an idea of how the organization's website performs and it can help approximate the typical user experience ([UX](https://www.techtarget.com/searchcio/definition/UX-user-experience)). A website owner can then adjust the site to make it more user-friendly and increase the chance that visitors will stay longer, make a purchase or otherwise interact with the website and the organization behind it.

Because an extremely large volume of data can be gathered through clickstream analysis, many e-businesses rely on [big data analytics](https://www.techtarget.com/searchbusinessanalytics/definition/big-data-analytics) and related tools such as [Hadoop](https://www.techtarget.com/searchdatamanagement/definition/Hadoop) to interpret the data and generate reports on specific areas of interest.Clickstream analysis is effective when used in conjunction with other, more traditional market research, evaluation resources, data sources and strategies.

There are two levels of clickstream analysis: traffic analytics and [e-commerce](https://www.techtarget.com/searchcio/definition/e-commerce) analytics.

Traffic analytics

This analysis operates at the server level. It collects and analyzes the following data sets:

* how many pages are served to a user;
* how long it takes each page to load;
* how often the user hits the browser's back button; and
* how much data is transmitted before the user moves to a different webpage.

E-commerce analytics

This analysis uses clickstream data to determine the effectiveness of a website in terms of conversions and transactions. It is concerned with the following data points:

* what pages the shopper lingers on;
* what the shopper puts in or takes out of a shopping cart;
* what items the shopper purchases;
* whether the shopper belongs to a loyalty program;
* whether the shopper uses a coupon code; and
* the shopper's preferred method of payment.

**Benefits of clickstream data analysis**

There are a number of benefits organizations can get from clickstream data and clickstream analytics. Among them are the following:

* User information. The data collected can include search terms used, pages landed on, webpage features used and the addition or removal of items from a cart, all of which can lead to more actionable insights.
* User routes. Organizations can use data analysis to view the different routes their online visitors or customers take to reach a page or to make a purchase.
* Customer trends and insights. Collecting and analyzing the clickstreams of a large number of visitors lets an organization identify trends in the following areas:
  + how visitors get to the website;
  + what they do once there;
  + how long they stay on a page;
  + the number of page visits visitors make; and
  + the number of unique and repeat visitors.
* UX. If a majority of users quickly leave a page or website, it could be a sign that the page is poorly optimized or doesn't contain enough information of value. Clickstream data enables an organization to [recognize UX shortcomings](https://www.techtarget.com/searchcustomerexperience/tip/5-digital-customer-experience-trends-for-businesses-to-consider), enabling them to make necessary changes.
* Digital marketing. Clickstream data can be used to determine the amount of traffic coming from ad banners and campaigns. Such data provides insight as to which advertisements are most effective and lead to customer [conversion rate optimization](https://www.techtarget.com/whatis/definition/conversion-rate-optimization). Clickstream analysis can also derive what times of day, month or year a marketing strategy is most effective.

*Let’s check the take away from this lecture*

Exercise:

Question 1: What is clickstream analysis used for?

A) Tracking weather patterns  
**B) Analyzing user interactions on a website**C) Studying ocean currents  
D) Monitoring stock market fluctuations

Question 2: What does clickstream data include?

A) Personal details of users  
B) IP addresses of visitors  
**C) Pages visited, time spent, and actions taken by users**  
D) Social media profiles of users

Question 3: What is the primary purpose of traffic analytics in clickstream analysis?

A) Calculating website revenue  
**B) Analyzing user behavior and page loading times**  
C) Evaluating customer satisfaction  
D) Predicting future market trends

Question 4: How can clickstream analysis benefit digital marketing strategies?

A) By predicting the weather patterns  
**B) By optimizing customer conversion rates and analyzing advertisement effectiveness**  
C) By analyzing geological formations  
D) By monitoring space exploration data

**Learning from this lecture:** Learners learned concepts behind click stream analysis and its usage.

**Lecture 3**

**Learning Objective:**  Learners shall be able to understand A/B Testing with handson.

**A/B Testing**

A/B testing, also known as split testing, is a method used in marketing and experimentation to compare two versions of a webpage, advertisement, or other digital content to determine which one performs better in achieving a specific goal or outcome. It involves randomly dividing a sample of users into two groups: Group A and Group B.

In A/B testing, Group A is exposed to the original or control version (referred to as variant A) while Group B is exposed to a modified version (referred to as variant B).The two variants differ in a specific aspect, such as a different headline, call-to-action button, layout, or color scheme. The performance of each variant is then measured, and statistical analysis is applied to determine if there is a significant difference in their effectiveness.

The primary objective of A/B testing is to gather data-driven insights about user behavior and preferences to optimize the performance of a particular element or feature. It helps answer questions like:

1. Which version leads to higher click-through rates, conversions, or sales?

2. Which variant improves user engagement or reduces bounce rates?

3. Which design or content element generates more user interactions or time spent on the page?

By systematically testing different variants and measuring the outcomes, A/B testing allows marketers and designers to make informed decisions based on empirical evidence rather than assumptions or subjective opinions. It helps identify areas of improvement, validate hypotheses, and optimize digital experiences to achieve desired business objectives.

To conduct an effective A/B test, it is important to define clear goals and metrics, ensure an adequate sample size for statistical significance, and carefully track and analyze the results. A/B testing platforms or tools can assist in managing the test setup, randomization, data collection, and statistical analysis.

It's worth noting that A/B testing is an iterative process, and multiple rounds of testing may be required to refine and optimize digital assets continually. By utilizing A/B testing, businesses can make data-driven decisions, improve user experiences, and ultimately achieve better results in their marketing and optimization efforts.

*Let’s check the take away from this lecture*

Exercise:

Question 1: What is the main purpose of A/B testing in digital marketing?

A) To compare two different products  
**B) To compare two versions of digital content**C) To analyze social media engagement  
D) To calculate website traffic

Question 2: What does Group A represent in an A/B test?

A) The control group exposed to variant B  
**B) The experimental group exposed to variant A**C) The group not participating in the test  
D) The group exposed to both variants simultaneously

Question 3: Why is statistical analysis important in A/B testing?

A) To create visually appealing content  
**B) To determine if differences in outcomes are statistically significant**  
C) To identify the most popular color schemes  
D) To track user engagement on social media

Question 4: What kind of decisions can businesses make based on A/B testing results?

A) Decisions based on personal opinions  
**B) Data-driven decisions backed by empirical evidence**C) Random decisions to experiment with different content  
D) Decisions solely based on competitors' strategies

Question 5: Which factor is crucial in setting up an effective A/B test?

A) Using a small sample size for quicker results  
B) Implementing changes to both variants simultaneously  
**C) Randomizing user assignment to groups**  
D) Relying solely on subjective opinions for content changes

**Learning from this lecture:** Learners learned A/B testing.

**Lecture 4**

**Learning Objective:**  Learners shall be able to understand how online surveys are conducted with tools.

**Online Surveys**

Online surveys are a popular method of collecting data and gathering insights from a targeted audience. They involve the distribution of a set of questions through digital platforms, such as websites, email, social media, or survey platforms, to collect respondents' opinions, feedback, or demographic information.

Here are some key aspects of online surveys:

1. Design: Online surveys can be designed using various question formats, including multiple-choice, open-ended, Likert scale, or rating scales. The design should be user-friendly, visually appealing, and easy to navigate to encourage higher response rates.

2. Target Audience: Online surveys allow researchers to reach a wide range of participants, including specific demographics or individuals with particular interests. Targeting the right audience ensures that the collected data is relevant to the research objective.

﻿

3. Convenience and Accessibility: Respondents can complete online surveys at their convenience, from any location with internet access. This convenience factor leads to higher response rates compared to other data collection methods, such as phone or in-person interviews.

4. Data Collection: Online surveys automatically collect and store responses in a centralized database, eliminating the need for manual data entry. This streamlines the analysis process and reduces the chances of data entry

errors.

5. Data Analysis: Online survey platforms often provide tools for data analysis, including generating summary statistics, creating visualizations, and exporting data for further analysis. Researchers can gain insights by analyzing response patterns, identifying trends, and drawing conclusions from the collected data.

6. Cost and Time Efficiency: Online surveys are generally more cost-effective than traditional methods, such as paper-based surveys or phone interviews. They eliminate printing and mailing costs and can be conducted within a shorter time frame, allowing for quicker data collection and analysis.

7. Flexibility: Online surveys offer flexibility in terms of survey length, question branching, and customization options. Researchers can tailor surveys to specific research objectives and easily modify them as needed.

However, it's important to note that online surveys have limitations. Response bias may occur, as respondents self- select to participate. It's essential to consider the representativeness of the sample and account for potential biases in the analysis. Additionally, not all populations have equal access to the internet, which can limit the reach of online surveys.

*Let’s check the take away from this lecture*

Exercise:

Question 1: What is the primary advantage of online surveys in terms of data collection?

A) Higher response rates due to personal interaction  
**B) Convenient completion from any location with internet access**C) Immediate analysis of responses during the survey  
D) Greater accuracy in capturing nuanced responses

Question 2: How do online surveys streamline the data analysis process?

A) By requiring manual data entry for each response  
B) By eliminating response patterns and trends  
**C) By automatically collecting and storing responses in a centralized database**  
D) By limiting customization options for researchers

Question 3: What is a potential limitation of online surveys related to respondent participation?

A) Limited survey length  
**B) Response bias due to self-selection**C) Inability to customize survey questions  
D) Limited availability of data analysis tools

Question 4: What advantage do online surveys offer in terms of cost efficiency?

A) Higher printing and mailing costs compared to traditional surveys  
B) Immediate data analysis leading to reduced costs  
**C) Elimination of printing and mailing costs**  
D) Longer time frames for data collection

Question 5: What is a key consideration for researchers when using online surveys to ensure reliable data analysis?

A) Increasing survey length for comprehensive data  
B) Ignoring response patterns to focus on individual responses  
**C) Tailoring surveys to specific research objectives**D) Avoiding customization options to maintain consistency

**Learning from this lecture:** Learners learned how online surveys are conducted with tools.

**Lecture 5**

**Learning Objective:**  Learners shall be able to understand Google Analytics.

﻿

**What is Google Analytics?**

Google Analytics is a web analytics service that provides statistics and basic analytical tools for search engine optimization ([SEO](https://www.techtarget.com/whatis/definition/search-engine-optimization-SEO)) and marketing purposes. The service is part of the Google Marketing Platform and is available for free to anyone with a Google account.

Google Analytics is used to track website performance and collect visitor insights. It can help organizations determine top sources of user traffic, gauge the success of their marketing activities and campaigns, track goal completions (such as purchases, adding products to carts), discover patterns and trends in user engagement and obtain other visitor information such as demographics. Small and medium-sized retail websites often use Google Analytics to obtain and analyze various [customer behavior analytics](https://www.techtarget.com/searchcustomerexperience/news/252441433/Customer-behavior-analytics-brings-marketing-analytics-together), which can be used to improve marketing campaigns, drive website traffic and better retain visitors.

**How does Google Analytics work?**

Google Analytics acquires user data from each website visitor through the use of page tags. A [JavaScript](https://www.theserverside.com/definition/JavaScript) page [tag](https://www.techtarget.com/whatis/definition/tag) is inserted into the code of each page. This tag runs in the web browser of each visitor, collecting data and sending it to one of Google's data collection servers. Google Analytics can then generate customizable reports to track and [visualize](https://www.techtarget.com/searchbusinessanalytics/definition/data-visualization) data such as the number of users, bounce rates, average session durations, sessions by channel, page views, goal completions and more.

The page tag functions as a [web bug](https://www.techtarget.com/whatis/definition/Web-bug-Web-beacon) or web beacon, to gather visitor information. However, because it relies on [cookies](https://www.techtarget.com/searchsoftwarequality/definition/cookie), the system can't collect data for users who have disabled them

Google Analytics has distinct benefits and limitations. Pros generally relate to the platform being powerful, free and user-friendly.

Google Analytics also provides the following benefits:

* The service is free, easy to use and beginner friendly.
* Google Analytics offers a variety of metrics and customizable dimensions. Many different types of useful insights can be captured using this platform.
* Google Analytics also contains many other tools, such as data visualization, monitoring, reporting, predictive analysis, etc.

*Let’s check the take away from this lecture*

Exercise:

Question 1: What is the primary purpose of Google Analytics?

**A) Tracking website performance and collecting visitor insights**  
B) Managing social media accounts  
C) Creating digital advertisements  
D) Designing website layouts

Question 2: How does Google Analytics acquire user data from websites?

A) Through email notifications  
**B) Via JavaScript page tags inserted into webpage code**C) Through phone calls made to website visitors  
D) By analyzing website images

Question 3: What types of insights can be obtained using Google Analytics?

A) Weather forecasts and traffic updates  
B) Cooking recipes and fashion trends  
**C) Number of users, bounce rates, page views, and goal completions**D) Historical events and political news

Question 4: What are some benefits of Google Analytics?

A) Provides weather forecasts and travel recommendations  
B) Requires a subscription fee for basic usage  
**C) Free, easy to use, and offers various metrics and customizable dimensions**D) Only applicable to large enterprises

Question 5: What is a limitation of Google Analytics regarding data collection?

A) It collects data even if users disable cookies  
**B) It relies on cookies, so it can't gather data for users who have disabled them**C) It collects data only from specific geographical regions  
D) It requires users to input data manually

**Learning from this lecture:** Learners learned Google Analytics.

**Lecture 6**

**Learning Objective:**  Learners shall be able to understand web crawling.

**Web Crawling and Indexing**

Web crawling and indexing are fundamental processes in the field of web search engines, allowing for efficient and comprehensive search results. Here's an explanation of both processes:

**1. Web Crawling:**

Web crawling, also known as web scraping or spidering, refers to the automated process of systematically browsing the World Wide Web to discover and gather information from web pages. It involves the use of web crawlers or bots, which are computer programs that navigate through web pages by following hyperlinks.

The web crawler starts with a list of seed URLS and proceeds to visit each URL, extracting the content and following any outgoing links on the page. This process continues recursively, enabling the crawler to access and collect data from a large number of web pages. The collected data can include text, images, metadata, and other relevant information.

﻿Web crawling is vital for search engines as it allows them to continuously discover and update their index of web pages. It enables search engines to provide up-to-date and comprehensive search results by ensuring that new content is indexed and existing content reindexed.

*Let’s check the take away from this lecture*

Exercise:

Question 1: What is the primary purpose of web crawling?

A) Sending emails  
B) Automating online purchases  
**C) Discovering and gathering information from web pages**  
D) Social media management

Question 2: What are web crawlers or bots used for in web crawling?

A) Ordering products online  
B) Automating phone calls  
**C) Navigating through web pages and following hyperlinks**  
D) Managing calendar events

Question 3: Why is web crawling important for search engines?

A) It helps in organizing email accounts  
B) It ensures that new web pages are created  
**C) It allows search engines to discover and update their index of web pages**  
D) It is used for weather forecasting

**Learning from this lecture:** Learners learned the concept behind web crawling..

**Lecture 7**

**Learning Objective:**  Learners shall be able to understand Indexing.

**Indexing:**

Once web pages are crawled, the next step is indexing. Indexing is the process of organizing and storing the collected web page data in a structured manner to facilitate quick and efficient search retrieval. The data is typically stored in an index database that allows for fast searching and retrieval of relevant information. During indexing, various techniques are employed to analyze and extract key information from web pages. This can include extracting text content, identifying important keywords, generating metadata, and establishing relationships between different web pages. The indexed data is organized in a way that makes it searchable based on user queries.

Indexing allows search engines to quickly retrieve relevant search results based on user queries. By matching the search query against the indexed data, search engines can rank and display the most relevant web pages to the

user.

Web crawling and indexing are continuous processes as new web pages are constantly added to the web and existing pages are updated. Search engines employ sophisticated algorithms to ensure efficient crawling, effective indexing, and accurate retrieval of search results.

It's worth noting that web crawling and indexing raise important ethical considerations, such as respecting website owners' terms of service, adhering to copyright laws, and being mindful of privacy concerns. Responsible web crawling and indexing practices involve respecting website guidelines and ensuring the appropriate use of collected data.

*Let’s check the take away from this lecture*

Exercise:

Question 1: What is the primary purpose of indexing in web crawling?

A) Sending emails  
**B) Organizing and storing web page data for efficient search retrieval**C) Designing website layouts  
D) Managing social media accounts

Question 2: How do search engines use indexed data to provide search results?

A) By displaying web pages randomly  
**B) By matching the search query against the indexed data to rank and display relevant web pages**  
C) By displaying only the latest web pages  
D) By organizing web pages alphabetically

Question 3: What are some ethical considerations related to web crawling and indexing?

**A) Respecting website owners' terms of service, adhering to copyright laws, and being mindful of privacy concerns**  
B) Collecting data without permission for marketing purposes  
C) Ignoring website guidelines to gather more data  
D) Storing collected data without any restrictions

**Learning from this lecture:** Learners learned the concept behind Indexing..

﻿**Lecture 8**

**Learning Objective:**  Learners shall be able to understand Natural Language Processing Techniques for Micro-text Analysis.

**Natural Language Processing Techniques for Micro-text Analysis**

Micro-text analysis refers to the process of analyzing short and concise text data, such as social media posts, tweets, product reviews, and chat messages. Natural Language Processing (NLP) techniques play a crucial role in extracting meaningful insights from micro-text data.

**NLP Techniques in micro-text analysis:**

• Tokenization: Tokenization is the process of breaking down a text into individual tokens or words. In micro-text analysis, tokenization helps in segmenting short text data into meaningful units, allowing for further analysis. Part-of-speech Tagging: Part-of-speech tagging assigns grammatical labels (e.g., noun, verb, adjective) to each word in a sentence. It helps in understanding the syntactic structure of micro-texts and extracting relevant information.

.

• Named Entity Recognition (NER): NER identifies and classifies named entities, such as people, organizations, locations, and dates, within micro-texts. It helps in extracting specific entities mentioned in the text and understanding their relationships.

• Sentiment Analysis: Sentiment analysis determines the sentiment or opinion expressed in a micro-text, whether it is positive, negative, or neutral. It is commonly used in analyzing product reviews, social media posts, and customer feedback.

• Topic Modeling: Topic modeling is a technique that discovers latent topics within a collection of micro-texts. It helps in identifying the main themes or subjects discussed in the text data.

• Emotion Detection: Emotion detection aims to identify the emotional content or sentiment expressed in micro- texts. It goes beyond simple sentiment analysis by detecting specific emotions such as joy, anger, sadness, or surprise.

﻿

Text Classification: Text classification involves categorizing micro-texts into predefined categories or classes. It is useful for tasks like spam detection, topic classification, and sentiment-based categorization.

• Word Embeddings: Word embeddings represent words or phrases as numerical vectors in a high-dimensional space. They capture semantic relationships between words, enabling algorithms to understand the context and meaning of micro-texts.

• Named Entity Disambiguation: Named Entity Disambiguation resolves ambiguities in named entities by associating them with their specific meanings or entities. It helps in disambiguating references to people, locations, or organizations mentioned in micro-texts.

Text Summarization: Text summarization techniques condense lengthy micro-texts into shorter summaries while preserving key information. They are useful for extracting the most important points from a large amount of micro-text data.

• Named Entity Linking (NEL): Named Entity Linking connects named entities mentioned in micro-texts to their corresponding entries in a knowledge base or database. It helps in enriching the understanding of entities and enables further exploration of related information.

• Entity Sentiment Analysis: Entity sentiment analysis focuses on determining the sentiment or opinion expressed towards specific named entities within micro-texts. It provides a more granular understanding of sentiment by associating it with particular entities.

·

Aspect-Based Sentiment Analysis: Aspect-based sentiment analysis goes beyond overall sentiment and analyzes the sentiment associated with specific aspects or features mentioned in micro-texts. It is particularly useful for product reviews, where different aspects of a product are discussed.

• Opinion Mining: Opinion mining, also known as sentiment mining or sentiment analysis, involves extracting

subjective information, opinions, and attitudes from micro-texts. It helps in understanding public opinion and sentiment trends.

*Let’s check the take away from this lecture*

Exercise:

Question 1: Which natural language processing (NLP) technique is used to determine the sentiment expressed in a micro-text, such as a social media post or product review?

A) Tokenization  
B) Named Entity Recognition (NER)  
**C) Sentiment Analysis**  
D) Topic Modeling

Question 2: What does Named Entity Recognition (NER) aim to identify and classify within micro-texts?

A) Semantic relationships between words  
B) Emotions expressed in the text  
**C) Named entities like people, organizations, locations, and dates**D) Latent topics within the text

Question 3: Which NLP technique goes beyond overall sentiment and analyzes the sentiment associated with specific aspects or features mentioned in micro-texts, such as product reviews?

A) Sentiment Analysis  
**B) Aspect-Based Sentiment Analysis**C) Opinion Mining  
D) Named Entity Recognition (NER)

**Learning from this lecture:** Learners learned the concept behind Natural Language Processing Techniques for Micro-text Analysis.

**Lecture 9**

**Learning Objective:**  Learners shall be able to understand Google Analytics.

Emotion Classification: Emotion classification aims to categorize micro-texts into different emotional categories, such as happiness, sadness, anger, fear, or surprise. It provides insights into the emotional experiences expressed in micro-texts.

• Text Clustering: Text clustering groups similar micro-texts together based on their content. It helps in identifying patterns, themes, or clusters of related micro-texts, which can be useful for segmentation or summarization purposes.

• Language Detection: Language detection determines the language in which a micro-text is written. It is particularly helpful in multilingual contexts, where micro-texts may be in different languages.

• Intent Classification: Intent classification involves identifying the intention or purpose behind a micro-text, such as whether it is a question, request, complaint, or suggestion. It aids in understanding user intent and facilitating appropriate responses.

Named Entity Extraction: Named entity extraction involves identifying and extracting named entities from micro-texts, such as people's names, organizations, locations, or dates. It helps in building knowledge graphs or understanding key entities mentioned in micro-texts.

• Cross-lingual NLP: Cross-lingual NLP techniques enable the analysis of micro-texts in different languages, including translation, sentiment analysis, or entity extraction across language boundaries. They facilitate multilingual analysis and understanding.

These NLP techniques empower researchers, analysts, and organizations to gain valuable insights from micro-texts, enabling them to understand customer sentiment, track trends, perform market research, and make data-driven decisions. They leverage the power of language processing to unlock the information embedded within short and concise text data.

*Let’s check the take away from this lecture*

Exercise:

Question 1: Which NLP technique is used to group similar micro-texts together based on their content, helping in identifying patterns and themes?

A) Emotion Classification  
**B) Text Clustering**  
C) Language Detection  
D) Intent Classification

Question 2: What is the purpose of Intent Classification in the context of micro-text analysis?

A) Identifying the language of the micro-text  
B) Categorizing micro-texts into emotional states  
**C) Determining the intention or purpose behind a micro-text, such as a question or complaint**  
D) Extracting named entities from micro-texts

Question 3: Which NLP technique is particularly useful in multilingual contexts for identifying the language in which a micro-text is written?

A) Emotion Classification  
B) Named Entity Extraction  
C) Cross-lingual NLP  
**D) Language Detection**

**Learning from this lecture:** Learners learned the concept behind Natural Language Processing Techniques for Micro-text Analysis.

Objective Question 1: What is the primary purpose of web analytics? A) Analyzing weather patterns  
B) Tracking user behavior on a website  
C) Studying geological formations  
D) Monitoring stock market fluctuations

Answer: B) Tracking user behavior on a website

Objective Question 2: Which of the following is a key metric commonly tracked by web analytics tools like Google Analytics? A) Temperature  
B) Number of visitors  
C) Wind speed  
D) Traffic congestion

Answer: B) Number of visitors

Form 2: Subjective Questions

Module 4: Data Analytics

Lecture 1: Web Analytics Tools and Techniques

Question 1: Describe the significance of web analytics in optimizing user experience on websites. Provide examples of metrics that can be tracked and how businesses can benefit from analyzing this data.

Question 2: Imagine you are a digital marketing manager for an e-commerce website. Explain how you would use web analytics tools like Google Analytics to improve the website's performance and increase sales. Include specific strategies and metrics you would focus on in your analysis.