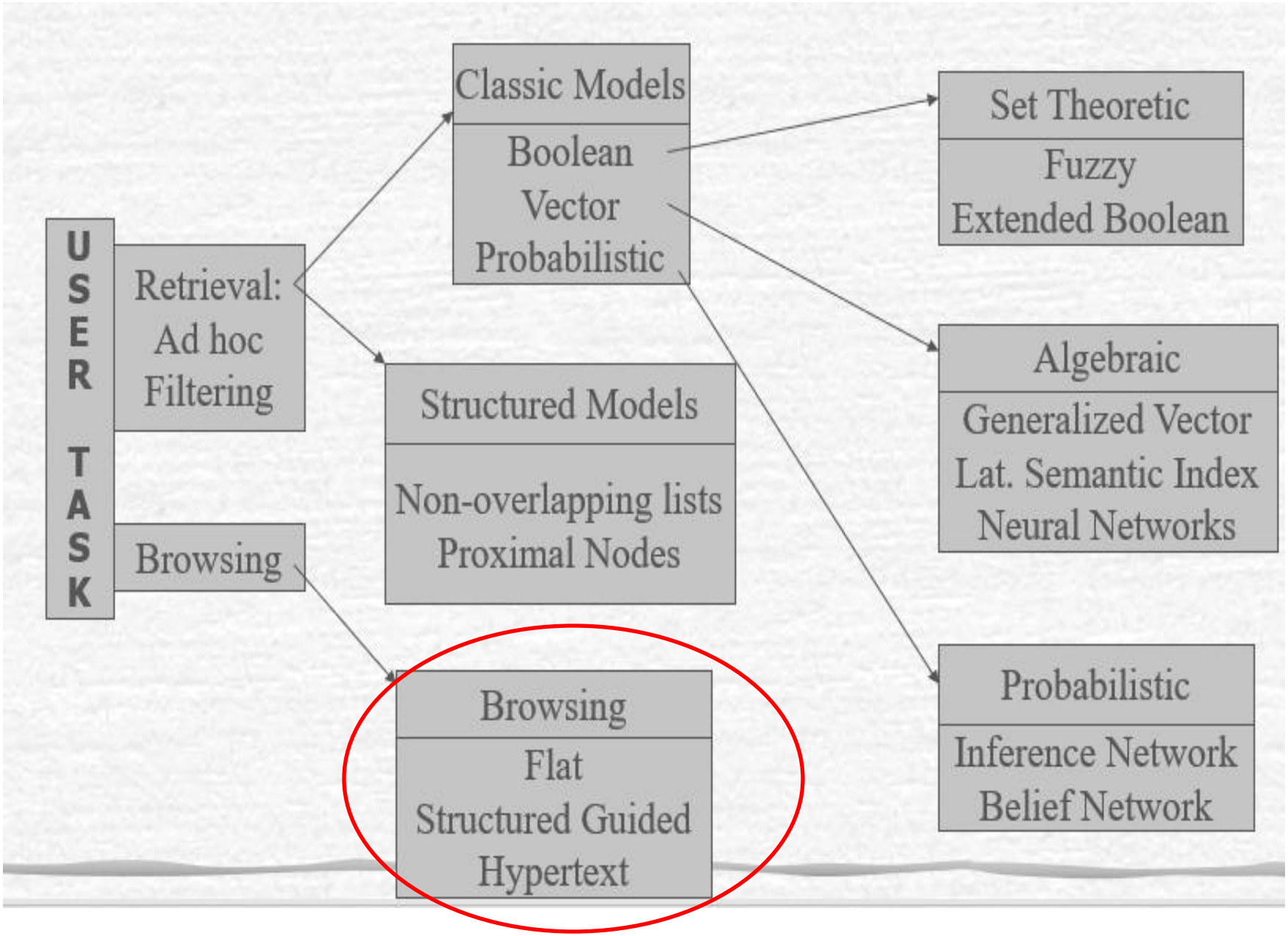
Information Retrieval

By
Dr Syed Khaldoon Khurshid


USER TASK
Retrieval: Ad hoc Filtering
Browsing
Classic Models
Boolean
Vector
Probabilistic
Structured Models
Non-overlapping lists
Proximal Nodes
Browsing
Flat
Structured Guided
Hypertext
Set Theoretic
Fuzzy
Extended Boolean
Algebraic
Generalized Vector
Lat. Semantic Index
Neural Networks
Probabilistic
Inference Network
Belief Network

Lecture of the week

- **Browsing**
  - **Flat**
  - **Structured Guided**
  - **Hypertext**

Importance

- **Information retrieval (IR) and its significance** in our digital age.
- IR systems help users find and access **relevant information** from vast collections of data.

In both Concepts

- **Understanding user behavior in IR is crucial for system design and optimization.**

Browsing and User Behavior

- Browsing can be defined as **an interactive search activity** in which the direction of the search is determined by the user on the basis of **immediate feedback** from the system being browsed.
- Most users of the information retrieval systems **exhibit browsing behavior** no matter what the underlying system structure.

Types of Browsing Behavior:

Different types of browsing activities users engage in are as follows:

1. **Exploratory browsing** is like a digital adventure where users start with a general idea and explore various paths to find interesting information.
2. **Directed browsing** is like having a specific destination in mind; users follow a predetermined path or query to find exactly what they're looking for.
3. **Opportunistic browsing** is a bit spontaneous; users switch between exploration and directed searches based on what catches their attention during the browsing process.

- So, **exploratory is about curiosity, directed is about precision, and opportunistic is a mix of both, adapting as users go along.**

Examples:

- **Exploratory Browsing:** Imagine you've just installed a new software application on your computer, and **you want to explore its features.** You click around the interface without a specific goal, trying different menus and buttons to see what the software can do. This is exploratory browsing, **where you're discovering the software's capabilities without a predefined task in mind.**
- **Directed Browsing:** Now, suppose you have a **specific task: you need to find and open a particular file on your computer.** You use the search function and enter the file name or location to quickly locate and access the specific file. This is directed browsing, **where you have a clear goal and follow a specific path to achieve it.**
- **Opportunistic Browsing:** While you're working on your specific task (directed browsing), **you notice a notification about a software update. You click on it out of curiosity to learn more about the update, even though it wasn't your initial goal**. This is opportunistic browsing, where **you seize unexpected opportunities to explore or gather additional information while engaged in a directed task**.

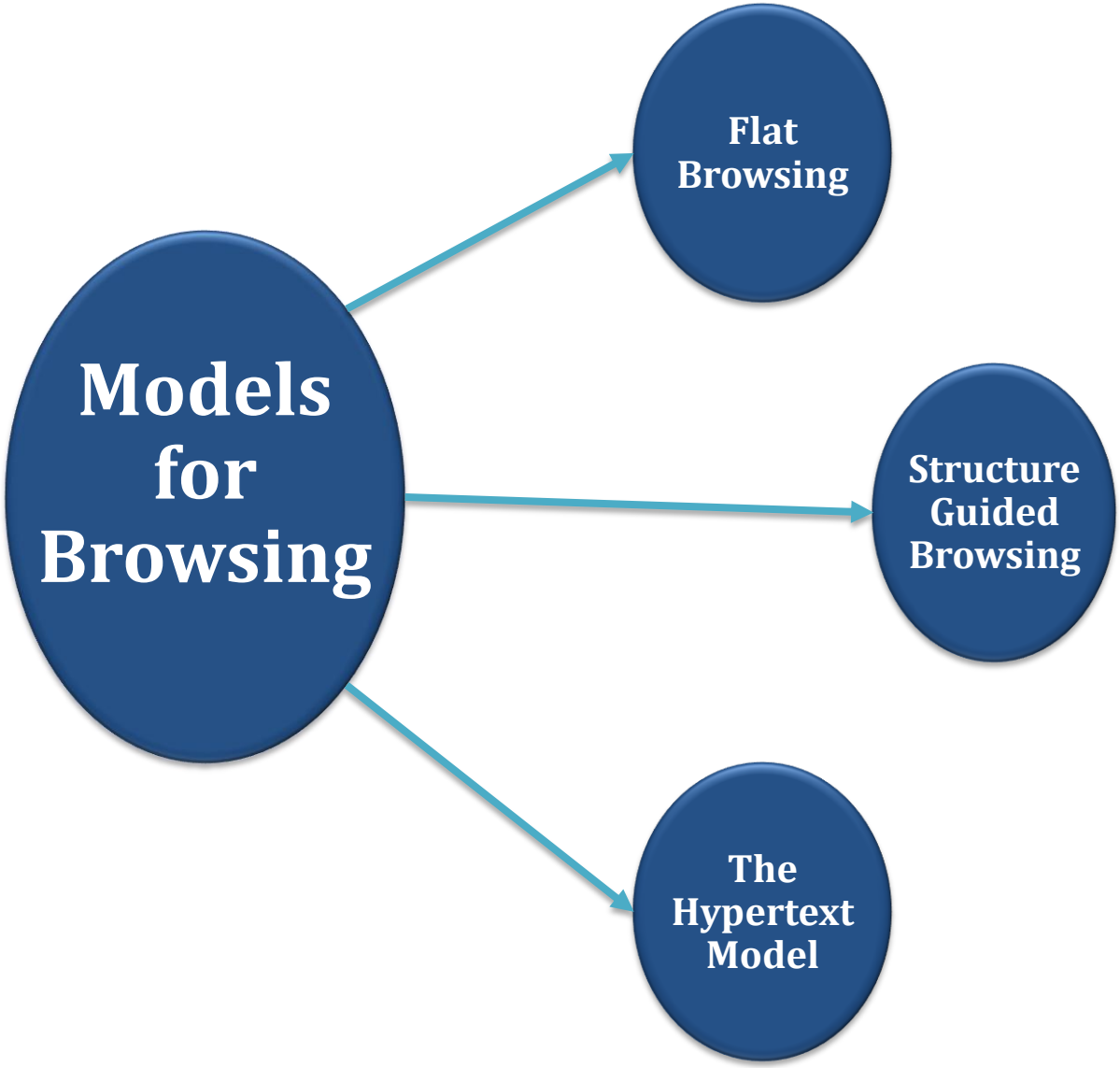
Examples of Features and Tools of Browsing

A few examples of browsing tools and features in popular Information Retrieval (IR) systems:

- **Google Search:**
  - **Search Filters**: Google provides filters such as "Images," "Videos," and "News" to help users refine their search results based on content type.
  - **Time Filters**: Users can filter search results by specific timeframes, such as "Past Hour" or "Custom Range," to find the most recent information.
- **Amazon (for Product Search):**
  - **Faceted Search**: Amazon uses facets like "Brand," "Price Range," and "Customer Ratings" to allow users to narrow down product search results based on specific criteria.
  - **Real-time Inventory**: The availability of products and pricing is updated in real-time, ensuring users see the most current information.
- **YouTube:**
  - **Video Thumbnails**: Users can preview video content through thumbnail images before clicking on a video, aiding in content exploration.
  - **Related Videos**: YouTube suggests related videos based on user behavior and video content, encouraging users to explore more videos.

Examples of Features and Tools of Browsing

- **eBay:**
  - **Category Navigation**: eBay employs a hierarchical category structure, allowing users to browse products by category, subcategory, and specific attributes.
  - **Saved Searches**: Users can save specific search queries and receive notifications when new matching listings are available.
- **Twitter:**
  - **Hashtags**: Users can explore trending topics and discover related tweets by clicking on hashtags within tweets.
  - **Real-time Updates**: Twitter provides real-time search results, ensuring users stay up-to-date with the latest tweets on a particular topic.
- **Library Databases (e.g., PubMed for Medical Research):**
  - **Faceted Searching**: Academic databases like PubMed offer facets such as "Publication Year," "Study Type," and "Authors" to refine search results for scholarly articles.
  - **Citation Metrics**: Users can view citation counts and related articles to explore the impact and relevance of academic papers.
- **These browsing tools and features enhance the user experience by allowing users to filter, explore, and refine search results interactively. Features like filters, facets, real-time updates, and related content recommendations facilitate efficient and effective information retrieval in various contexts.**


Models for Browsing
Flat Browsing
Structure Guided Browsing
The Hypertext Model

Flat Browsing

Document Representation:

- Imagine you have a bunch of documents. You can think of these documents as either dots on a piece of paper (like stars on a map) or as a list in a straight line. This helps you **organize and see the documents better.**

Relevance Feedback:

- When you're looking for information in these documents, you might read one document and then another. Relevance feedback is like when you're glancing through these documents to find what you need. You may also **add new words to your original search to help you find more relevant information**, like making your search smarter.

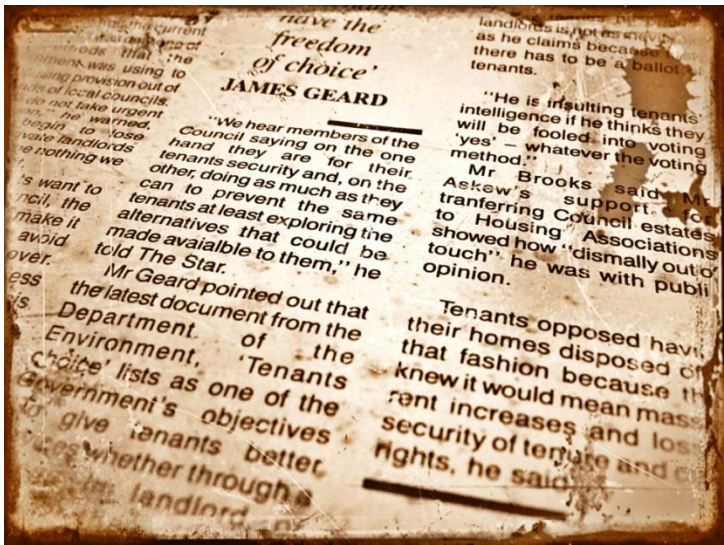
Flat Browsing

Flat Exploration:

- Sometimes, you explore one document very thoroughly, like reading a web page from top to bottom. **Flat browsing is needed when you want to thoroughly explore one document, like reading a web page from start to finish, to find specific information.**
- But the disadvantage is that you might not know where you are in the bigger context (like not knowing which part of a website you're on).
- So, it's like **having a map of documents, looking for what you need, and sometimes reading one document very carefully.** But be careful because you might get lost in the big picture!

Simple Example:

A **real-time searching example of flat browsing** is when you visit a news website and scroll through a single article, reading it from beginning to end to get the latest information on a specific topic or news event.



Flat browsing can be represented mathematically as:

$$Browsing = \sum_{i=1}^{n} P_i$$

Where:

- $Browsing$ represents the action of flat browsing.
- n is the number of documents explored.
- P_i represents the level of attention or exploration given to each document i, which can vary from document to document.

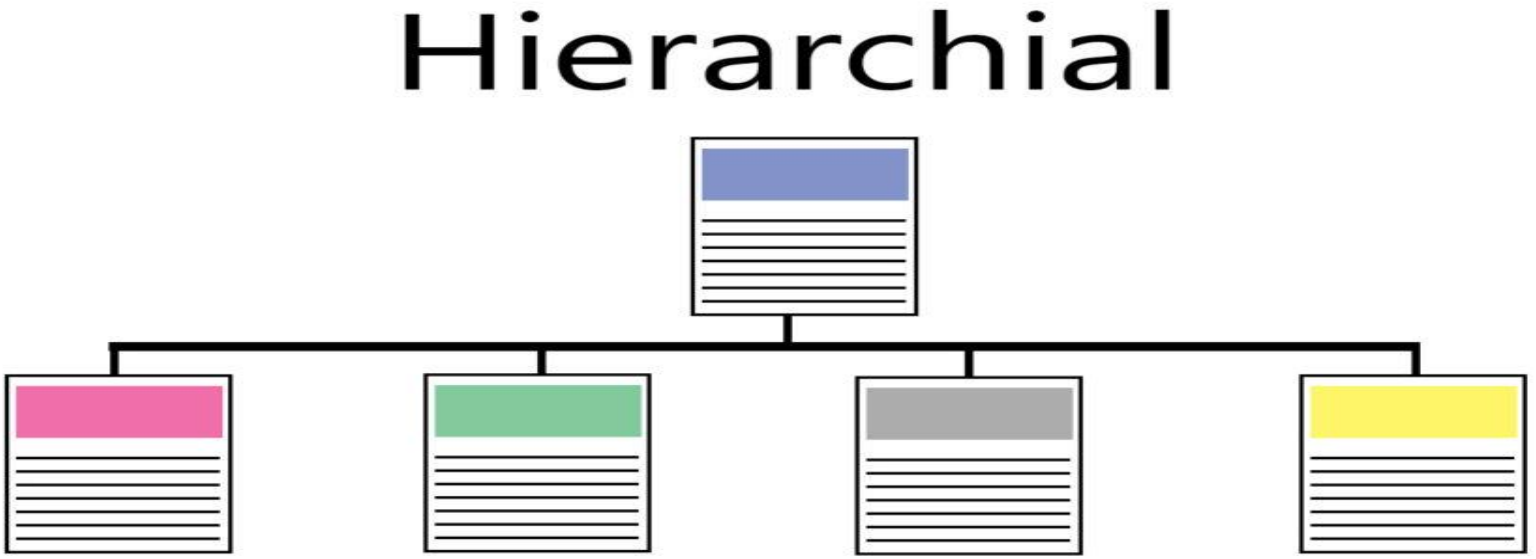
Computation:

- Suppose you're exploring three web pages (documents) on a topic, and you spend different amounts of time on each page:
  - You spend 30 minutes reading the first page (P1 = 30 minutes).
  - The second page doesn't have much information, so you only spend 10 minutes on it (P2 = 10 minutes).
  - The third page is very detailed, and you read it for 45 minutes (P3 = 45 minutes).
  - Now, let's calculate the total flat browsing time:
  - Browsing= P1+P2+P3= 30+10+45= 85

Structure Guided Browsing

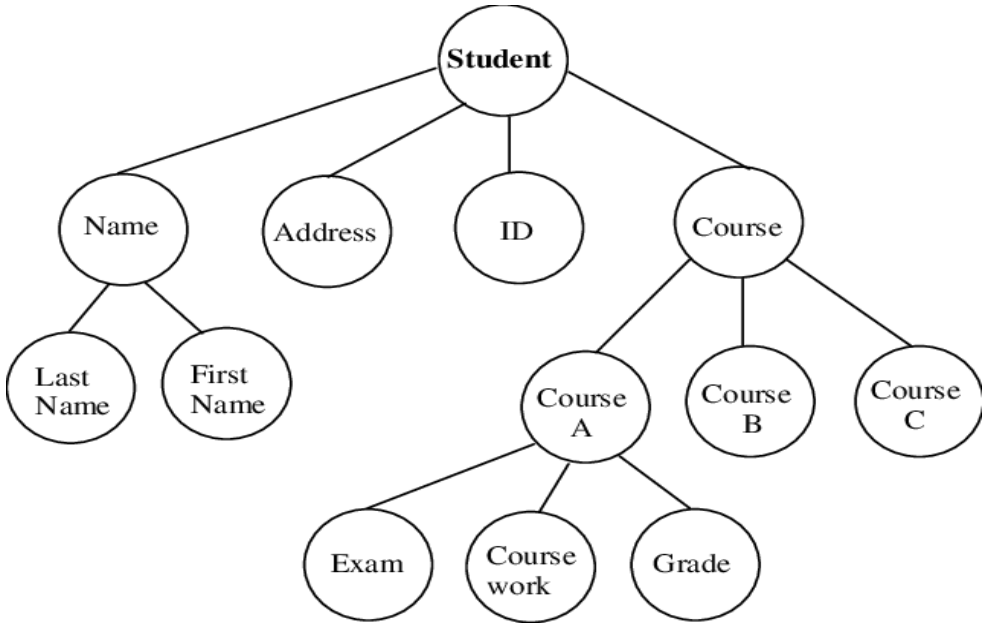
- **Documents organized in a structure as a directory.**
  - Directories are hierarchies of classes which group documents covering related topics

    e.g.: “Yahoo!” provides hierarchical directory



Structure Guided Browsing

- **Same idea applied to a single document**
  - Chapter level, section level, etc.
  - The last level is the text itself (flat!)
  - A good UI needed for keeping track of the context
  - e.g.: the adobe acrobat .pdf files



Applying Structure to Document Navigation:

The idea of applying hierarchical structures **to improve document navigation.**

The approach can be used for individual documents, making it easier to navigate lengthy content.

Hierarchy Levels and User Interface:

- The different levels of hierarchy, such as chapters and sections, within a document.
- The importance of a user-friendly interface (UI) to help users keep track of their location within the document.

Real-World Examples:

- There are websites or platforms that implement structure-guided browsing for better user experience.

Examples

1. Adobe Acrobat Reader:

Adobe Acrobat PDF files often include a table of contents and bookmarks that enable users to navigate through documents with ease. Users can jump to specific chapters, sections, or pages.

2. Wikipedia:

Wikipedia articles are organized into sections with table of contents navigation. Users can click on section headings to jump to specific parts of an article, aiding in structured browsing.

Examples

3. **eBook Readers(e.g., Kindle):**

   eBook readers provide features like chapter navigation, bookmarks, and highlights. Readers can easily move between chapters and sections, enhancing the reading experience.

4. **Academic Journals and Research Databases (e.g., IEEE Xplore, PubMed):**

   Research articles often have structured layouts with sections like abstracts, introductions, and references. Users can navigate directly to these sections for efficient information retrieval.

Examples

5. **Online Documentation and Help Centers (e.g., Microsoft Docs, GitHub):**

   Technical documentation and help centers are organized with a hierarchical structure. Users can explore topics, subtopics, and specific content sections to find relevant information.

6. **E-Learning Platforms (e.g., Coursera, edX):**

   Online courses are structured into modules and lessons. Learners can easily access different course components and track their progress.

Examples

7. File Management Systems (e.g., Google Drive, Dropbox):

Cloud storage platforms use folder hierarchies to organize files. Users can navigate through folders and subfolders to locate and manage their documents.

- These websites and platforms leverage structure-guided browsing to improve user navigation and provide a more organized and efficient user experience, especially when dealing with large and complex content.

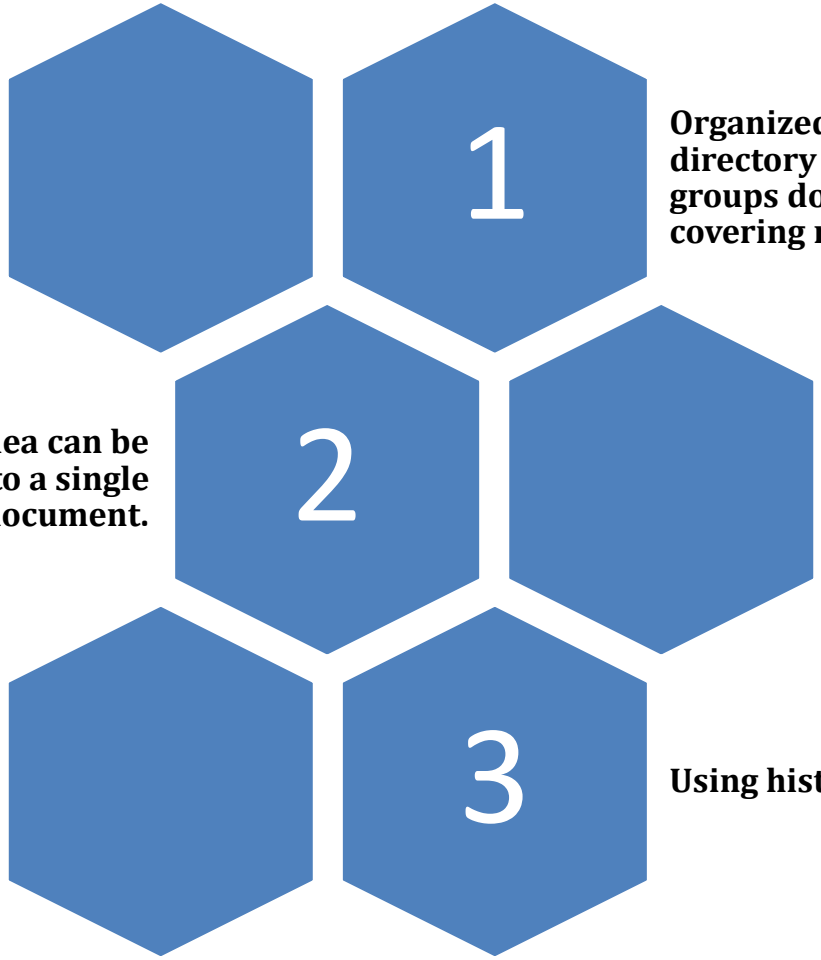
Benefits and Challenges:

- Highlight the benefits of structure-guided browsing, such as improved user navigation and understanding of complex content.
- Address the challenges, like designing effective UI elements, to ensure a seamless browsing experience.

Conclusion:

- Summarize the key points, underlining the significance of structured document navigation.
- Encourage the adoption of structure-guided browsing techniques to enhance user interaction with documents.

Structure Guided Browsing



Organized in a directory structure. It groups documents covering related topics.

The same idea can be applied to a single document.

Using history map.

Mathematical expression for the Structure Guided Browsing

- The algorithm can be represented as follows:

  ***StructureGuidedBrowsing*(D, H, UI) = N**

Where:

- ***StructureGuidedBrowsing*** represents the algorithm for guided browsing.
- **D** represents the document or content being browsed.
- **H** represents the hierarchical structure of the document, including levels like chapters, sections, etc.
- **UI** represents the user interface elements and tools designed for structured navigation.
- **N** represents the user's navigation path and actions within the structured document.

Mathematical expression for the Structure Guided Browsing

- This simplified expression represents the algorithm's input (document and structure) and user interaction through the user interface, **resulting in a navigation path within the document.**

Computed Example:

- Let's populate the mathematical expression for Structure Guided Browsing with real-time values:
- Suppose we are using the Structure Guided Browsing algorithm for an online research paper (D) that has a hierarchical structure (H) with three levels: Abstract, Sections (1-5), and References. The user interface (UI) includes features like a table of contents, section headings, and hyperlinks.

StructureGuidedBrowsing(*D, H, UI) = *UserNavigatedPath

- In this scenario:
- **D** represents the online research paper.
- **H** represents the hierarchical structure: Abstract, Sections (1-5), and References.
- **UI** includes elements like a table of contents, section headings, and hyperlinks.
- **UserNavigatedPath** represents the user's path within the document, such as

"Abstract -> Section 3 -> References."

- The algorithm takes these inputs and provides the user's navigation path within the document, showing how the user interacts with the structured content.

Week 4: Day 1 Completed.

Hypertext Model: the next...

Hypertext Model

Introduction:

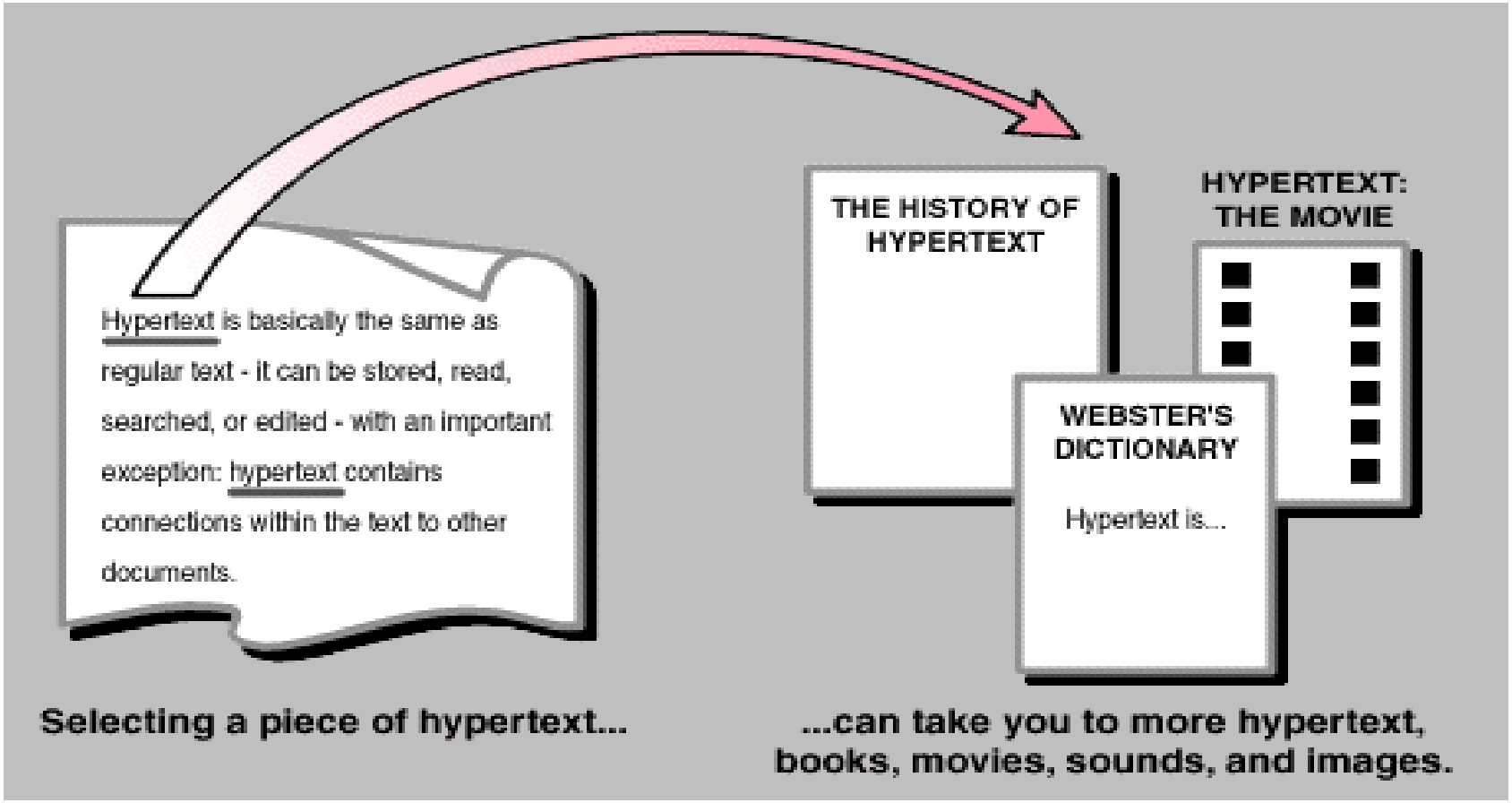
- Reading is a fundamental skill that allows us to access and understand written information. When we pick up a book, magazine, or any written material, our natural inclination is to start at the beginning and read through to the end. This **linear reading approach works well for many types of content, providing a structured and sequential understanding of the material.** **However**, **there are situations where this traditional reading method may not be the most efficient or effective way to absorb information.**

Hypertext Model

- This is where the concept of **non-linear reading**, often facilitated by the Hypertext Model, comes into play. In some cases, we need the flexibility to **jump between sections, chapters, or topics as per our specific needs.**
- **The Hypertext Model offers an alternative reading experience, allowing us to explore text in a non-sequential manner.**

The Hypertext Model

- **A high level interactive navigational structure allowing users to browse text non-sequentially.**



The Hypertext Model:

1. High-Level Interactive Navigational Structure:

- This refers to a way of reading that's not like starting at the beginning and going in order.
- Instead, it's like having a map where you can jump to different parts of the text as you like.

2. Nodes and Links:

- Imagine the text is like a big web.
- Nodes are like stops on the web. They could be chapters, sections, or web pages.
- Links are like the paths that connect these stops.

The Hypertext Model:

3. What Nodes Are:

- Nodes are where you find the actual content. They can be like chapters in a book, sections in an article, or even separate web pages.

4. Links in Nodes:

- Inside these nodes, there are links attached to specific words or phrases.
- These links act like signposts to guide you to related parts of the text.

The Hypertext Model:

5. Communication Between Writer and Reader:

- This whole setup is a way for the writer to communicate with the reader.
- Instead of reading from start to finish, readers can follow their interests.

6. Sequenced Organizational Structure:

- Most writing has a specific order. It's like a story that starts at the beginning and goes to the end.
- But not all writing is like that, and the Hypertext Model helps with those pieces.

The Hypertext Model:

7. Don't Expect to Understand Randomly:

- If you read randomly in a traditional book, you might not understand the story.
- Similarly, if you jump randomly in a hypertext document, you might miss the message.
- In a nutshell, **the Hypertext Model is like an interactive map for reading.** It breaks away from the usual start-to-finish reading and lets you explore different parts of the text easily. It's like a conversation between the writer and the reader, and it's great for content where the order isn't as important. But remember, it's still important to **follow some kind of logical path to fully understand the message.**

The Hypertext Model

Written text is usually conceived to be read sequentially.

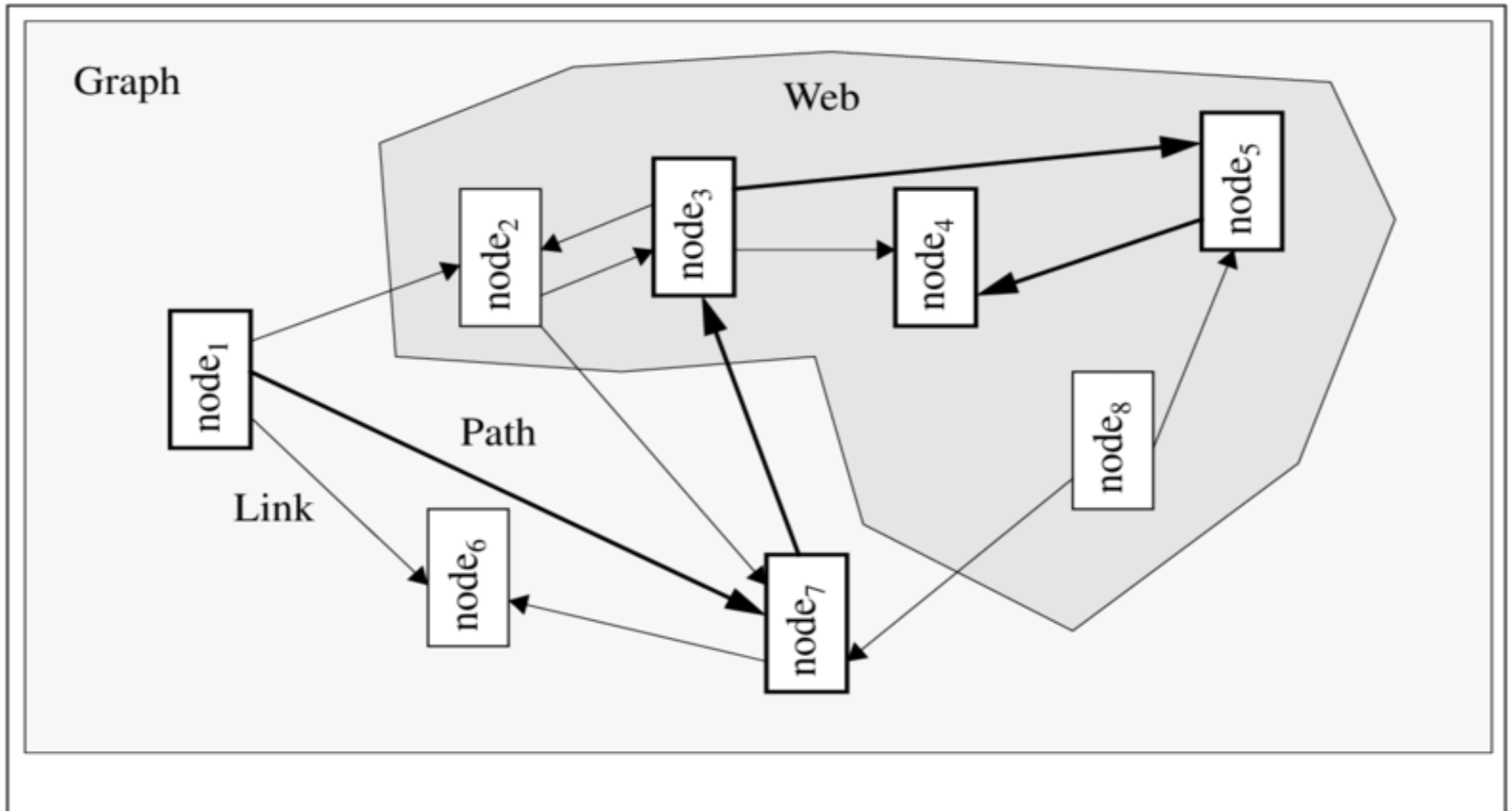
The reader should not expect to fully understand the message conveyed by the writer by randomly reading pieces of text here and there.

Nodes in Hypertext Model:

- In the context of the Hypertext Model, nodes are like **individual stops or chunks of information** within a larger document. Think of nodes as self-contained units of content that you can explore in a non-linear way. Here's a simple breakdown:
- Nodes are like **sections or chapters in a book**, parts of an article, or even separate web pages in an online document.
- Each node contains specific information related to a particular topic, idea, or theme.
- Nodes are designed to be relatively self-contained, meaning **you can understand them without necessarily reading everything before or after.**
- They act as **entry points to different parts of the document and are connected by links.**

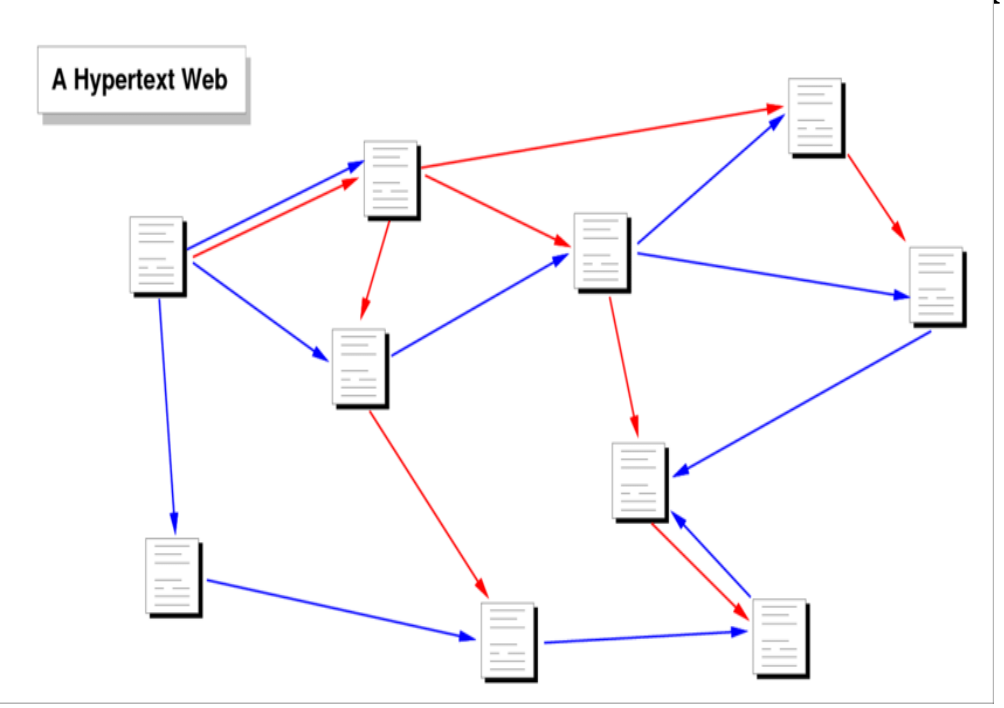
Examples of Nodes:

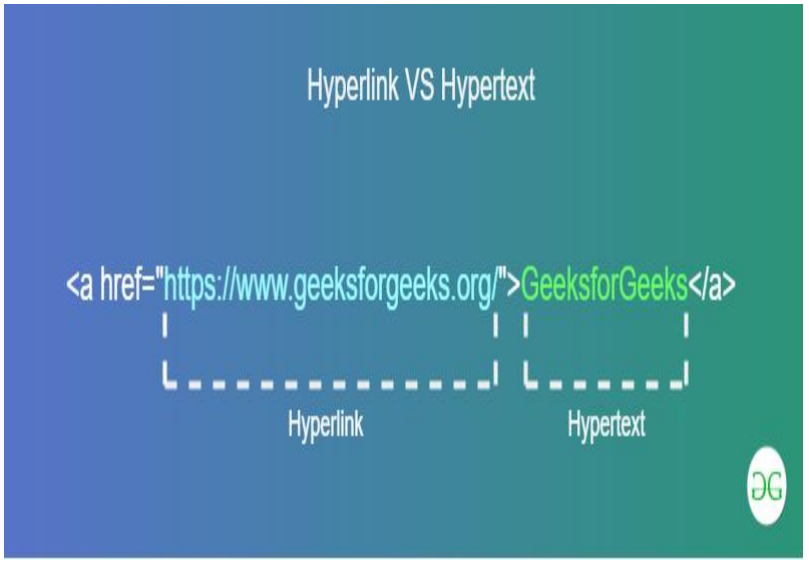
- In a book, **each chapter can be a node**. You can jump from one chapter to another without reading them in order.
- In an article, sections like "Introduction," "Methods," "Results," and "Conclusion" can be **nodes**.
- On a website, **each web page or subsection can be a node**. You can click on links to move between these nodes.



Links in the Hypertext Model:

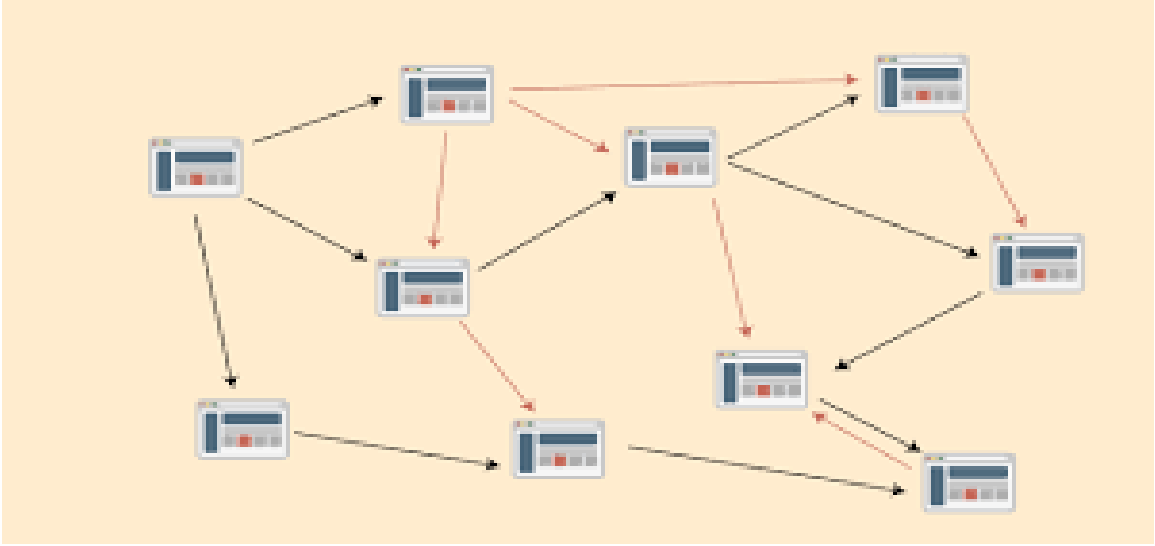
- Let's explain how links work within nodes in the Hypertext Model:
- In the Hypertext Model, links are like **interactive signposts or pathways** that are attached to specific words or phrases within the nodes.
- These links are often highlighted or underlined, making them stand out from the rest of the text.





Example:

- Imagine you're reading an online article about space exploration. In **one section (node),** you come across the phrase "Apollo 11 mission." This phrase is a link.
- When you click on "Apollo 11 mission," the link takes you to a new node that provides in-depth information about that historic mission, complete with images and further details.
- After reading about Apollo 11, you can easily return to the previous node by using a "back" button or link.



- In summary, **links in the Hypertext Model act as gateways to related content within the nodes.** They offer a dynamic and interactive way to explore information, enabling users to jump between topics, references, and details with ease.

- The Hypertext Model can be represented mathematically as follows:

Hypertext Model = (N, L)

Where:

- N represents the set of nodes within the hypertext, where each node contains specific content or information.
- L represents the set of links that connect these nodes, allowing users to navigate from one node to another.

- This mathematical expression encapsulates the fundamental components of the Hypertext Model, where nodes represent content units, and links enable non-linear navigation between these units.

Computed Example:

Hypertext Model = (N, L)

Now, let's populate it with values:

1. N (Nodes):

Node 1: "Introduction" (Contains introductory information about the topic)

Node 2: "Chapter 1" (A chapter within the document)

Node 3: "Chapter 2" (Another chapter within the document)

Node 4: "References" (A section with references and sources)

Node 5: "Conclusion" (The concluding section)

2. L (Links):

Link 1: Connects from Node 1 ("Introduction") to Node 2 ("Chapter 1")

Link 2: Connects from Node 1 ("Introduction") to Node 3 ("Chapter 2")

Link 3: Connects from Node 3 ("Chapter 2") to Node 4 ("References")

Link 4: Connects from Node 3 ("Chapter 2") to Node 5 ("Conclusion")

- **Hypertext Model consisting of nodes (representing different sections of a document) and links (representing connections between these sections).** Students can explore how users can navigate between nodes to access specific content within a document in a non-linear way.

Navigating Hypertext: Benefits and Challenges:

- **The Hypertext Model offers significant benefits in terms of flexibility and quick information access. However, it also presents challenges related to maintaining coherent understanding, dealing with overwhelming choices, staying contextually aware, and structuring content effectively.** Addressing these challenges while leveraging the model's benefits can lead to a more effective and engaging reading experience.

Some common applications of the Hypertext Model and how it enhances the user experience:

1. Web Articles and Blogs:

- **Example:** Online news articles, blog posts, and informative web content.
- **Enhancement:** Hypertext enhances the user experience by providing embedded links to related articles, references, or additional information. Readers can quickly access background details, sources, or related topics with a simple click. It offers a deeper and more informative reading experience.

Some common applications of the Hypertext Model and how it enhances the user experience:

2. E-books and Digital Textbooks:

- **Example:** Digital books and educational materials.
- **Enhancement:** In e-books, the Hypertext Model allows readers to navigate seamlessly between chapters, sections, and references. It promotes interactive learning by offering links to multimedia content, glossaries, and supplementary materials. This enriches the educational experience.

Some common applications of the Hypertext Model and how it enhances the user experience:

3. Online Documentation and Help Guides:

- **Example:** Software manuals, user guides, and troubleshooting documentation.
- **Enhancement:** Hypertext makes it easy for users to find specific information or solutions quickly. Clickable links guide users to relevant sections or troubleshooting steps, reducing frustration and enhancing user self-help. It streamlines the search for answers.

Some common applications of the Hypertext Model and how it enhances the user experience:

4. **Academic Research Papers:**

- **Example:** Online academic journals and research papers.
- **Enhancement:** Hypertext citations and references within research papers allow readers to access source materials and related studies effortlessly. This promotes transparency and facilitates scholarly discourse by enabling readers to explore the academic context easily.

Some common applications of the Hypertext Model and how it enhances the user experience:

5. **Interactive Tutorials and Online Courses:**

- **Example:** Interactive tutorials, online courses, and educational platforms.
- **Enhancement:** The Hypertext Model enables interactive learning by providing links to quizzes, assignments, video lectures, and supplementary resources. Learners can choose their learning paths, access additional explanations, and reinforce their understanding.

Some common applications of the Hypertext Model and how it enhances the user experience:

6. Online News Portals:

- **Example:** News websites.
- **Enhancement:** News articles often include links to related stories, background information, or multimedia content. Readers can get a comprehensive view of a news event by exploring various angles and sources through these links.

Some common applications of the Hypertext Model and how it enhances the user experience:

7. Encyclopedia and Knowledge Databases:

- **Example:** Online encyclopedias and knowledge databases.
- **Enhancement:** Hypertext allows users to explore a vast array of topics by clicking on cross-referenced links. This promotes self-directed learning and deepens users' understanding of complex subjects.

Empowering User-Centric Exploration with Hypertext

- In these contexts, the Hypertext Model significantly enhances the user experience by offering quick access to related information, enabling non-linear navigation, and providing a more interactive and informative reading or learning journey. **It empowers users to explore, discover, and engage with content on their terms, making digital information more accessible and engaging.**

Case Study Assignment-IV for IR:

Implement the architectures of Structure Guided Browsing and Hypertext Model in their respective chosen applications

REFERENCES:

[1] “INFORMATION RETRIEVAL MODELS,” *Ebrary*, 2013. https://ebrary.net/201793/psychology/retrieval_models (accessed Feb. 22, 2023).

[2] “Ch_2 Information Retrieval Models,” *Rutgers.edu*, 2023. https://aspoerri.comminfo.rutgers.edu/InfoCrystal/Ch_2.html (accessed Feb. 22, 2023).

[3] “Fig. 2. Horizontal taxonomy.,” *ResearchGate*, 2021. https://www.researchgate.net/figure/Horizontal-taxonomy_fig2_47397195 (accessed Feb. 22, 2023).

[4] “4.1. IR MODELS – BASIC CONCEPTS – Wachemo University e-Learning Platform,” *Wachemo-elearning.net*, 2023. https://wachemo-elearning.net/courses/information-storage-and-retirievalitec3081/lessons/chapter-four-ir-model/topic/4-1-ir-models-basic-concepts/ (accessed Feb. 22, 2023).


Any
Questions