CHRIST
(DEEMED TO BE UNIVERSITY)
BANGALORE, INDIA

Unit 1

Overview of Web Technologies and HTML 5

Search Engines

- A search engine is a software program that provides information according to the user query.
- It helps to locate information on World Wide Web.
- User can search for any information by passing query in form of keywords or phrase.
- It then searches for relevant information in its database and return to the user.

Search Engine Components

- **Web Crawler**
  - It is also known as spider or bots. It is a software component that traverses the web to gather information.
- **Database**
  - All the information on the web is stored in database. It consists of huge web resources.
- **Search Interfaces**
  - This component is an interface between user and the database. It helps the user to search through the database.

Search Engine Working

- Search engines make use of Boolean expression AND, OR, NOT to restrict and widen the results of a search.
- The search engine looks for the keyword in the index for predefined database instead of going directly to the web to search for the keyword.
- Search engines generally work on three parts that are crawling, indexing, and ranking

Web Crawler

- Web crawler is a software to search for the information in the database.
- The crawler scans the web and creates a list of all available websites.
- Then they visit each website and by reading HTML code they try to understand the structure of the page, the type of the content, the meaning of the content, and when it was created or updated.
- Once web crawler finds the pages, the search engine then shows the relevant web pages as a result. These retrieved web pages generally include title of page, size of text portion, first several sentences etc.

Indexing

- Information identified by the crawler needs to be organized, Sorted, and Stored so that it can be processed later by the ranking algorithm.
- Search engines keep information like
  - The Title and description of the page
  - The type of content
  - Associated keywords
  - Number of inwards and outwards link, and a lot other parameters required for the ranking algorithm.

Indexing

- Indexing process performs the following three tasks:
  - **Text acquisition** - It identifies and stores documents for indexing.
  - **Text transformation** - It transforms document into index terms or features
  - **Index creation** - It takes index terms created by text transformations and create data structures to suport fast searching.

Ranking

- Ranking is the position by which the website is listed in any Search Engine.
- Ranking works with the following three steps:
  - 1. **Analyze user query** – This step is to understand what kind of information the user is looking for.
  - To do this it **breaks down the user's query** into a number of meaningful keywords.
  - search engines are clever enough to **interpret spelling** mistakes also.

Ranking

- 2. **Finding matching pages**
  - This step is to look into their index and find the best matching pages, for example, if you search dark wallpaper then it gives you the result of images, not text.

- 3. **Present the results to the users**
  - A typical search results page includes ten organic results in most cases it is enriched with other elements like paid Ads, direct answers for specific queries, etc.

Examples

Search Engine	Description
Google	It was originally called BackRub. It is the most popular search engine globally.
Bing	It was launched in 2009 by Microsoft. It is the latest web-based search engine that also delivers Yahoo's results.
Ask	It was launched in 1996 and was originally known as Ask Jeeves. It includes support for match, dictionary, and conversation question.
AltaVista	It was launched by Digital Equipment Corporation in 1995. Since 2003, it is powered by Yahoo technology.
AOL.Search	It is powered by Google.
LYCOS	It is top 5 internet portal and 13th largest online property according to Media Matrix.
Alexa	It is subsidiary of Amazon and used for providing website traffic information.

Search Engine Optimization

- SEO means the process of improving your website's visibility in search engines whenever people search for:
    - Products ones sell.
    - Services provided.
    - Information on topics in which user have deep expertise and/or experience.

Web Services

- A web service is a standardized software module Written in varied programming languages, supports compatibility between networks :
  - To carry out a specific set of functions.
- A client invokes a web service by submitting an XML request, which the service responds with an XML response.

Web Services

- Web services include any software, application, or cloud technology that provides standardized web protocols (HTTP or HTTPS) to interoperate, communicate, and exchange data messaging – usually XML (Extensible Markup Language) – throughout the internet.

- The basic web services platform is XML + HTTP. All the standard web services work using the following components −

  - **SOAP (Simple Object Access Protocol)**
  - **UDDI (Universal Description, Discovery and Integration)**
  - **WSDL (Web Services Description Language)**

SOAP - Simple Object Access Protocol

- SOAP is a XML-based protocol for accessing web services, which is a language independent, and it can be used with Java, .Net or PHP language on any platform.

- SOAP is a protocol or in other words is a definition of how web services talk to each other or talk to client applications that invoke them.

Web Service Description Language (WSDL)

- WSDL is an XML based interface description language ,used for describing the functionality offered by a web service, with the extension **.wsdl.**

- It provides the **machine-readable description** of how the service can be called, what parameter it expects, and what data structure it returns.

Universal Description, Discovery, and Integration (UDDI)

- It is an XML-based registry for businesses word wide to list themselves on the internet, that defines a set of services supporting the description and discovery of the business, organizations, or other web service providers.

- A UDDI works in the following manner:
  - A **service provider registers** its business with the UDDI registry.
  - The consumer looks up the business and service in the UDDI registry.
  - The consumer binds the service with the service provider and uses the service.

UDDI…

- **White Pages**: The white pages contain basic information such as company name, address, phone number, and other business identifiers such as tax numbers.

- **Yellow Pages**: The yellow pages contain detailed business data organized by relevant business classification.

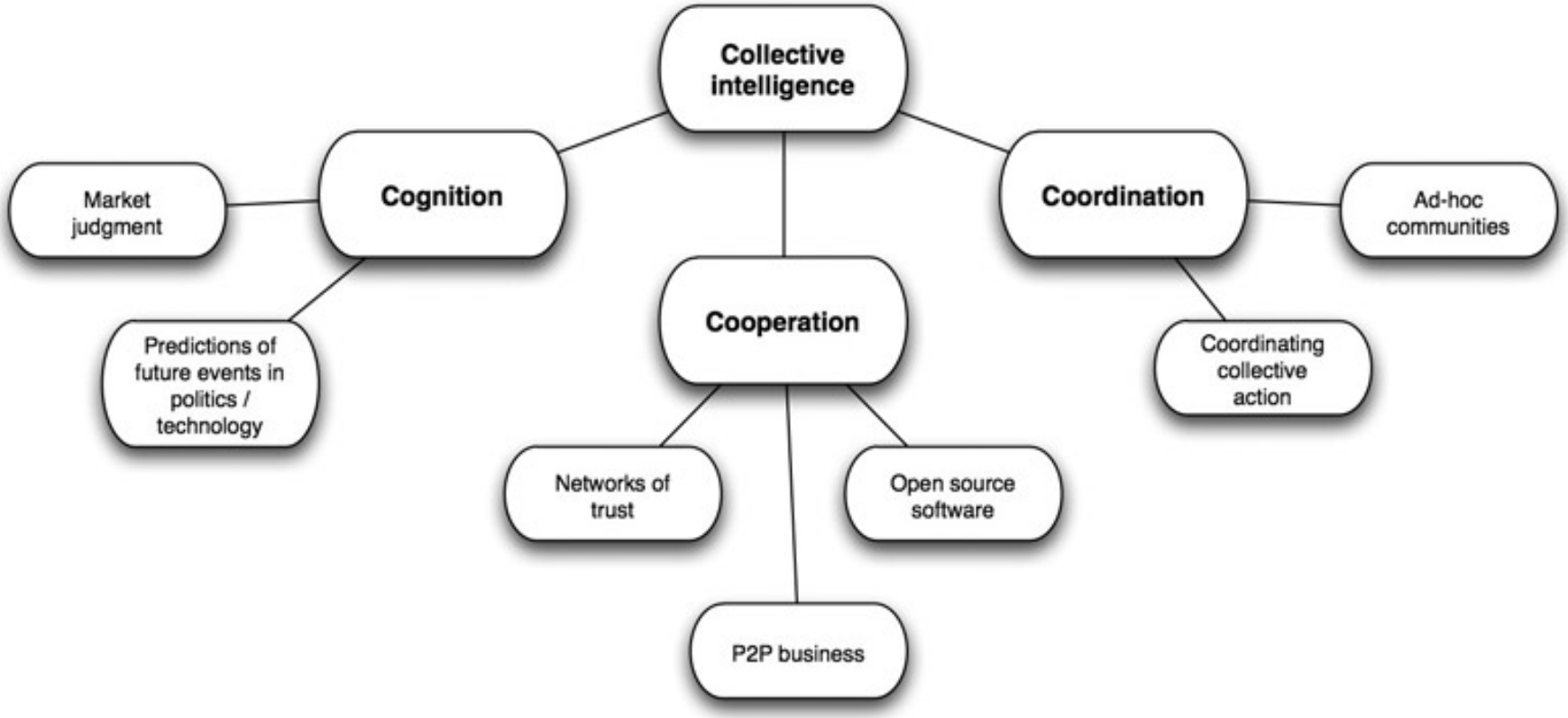
- **Green Pages**: The green pages contain information about the company's crucial business process, such as operating platform, supported programs, and other high-level business protocols.

Collective Intelligence

- **Collective intelligence (CI)** is shared or **group intelligence (GI)** that emerges from the collaboration, collective efforts, and competition of many individuals and appears in agreement decision making.

- Collective IQ is a measure of collective intelligence, although it is often used interchangeably with the term collective intelligence.

Collective Intelligence…

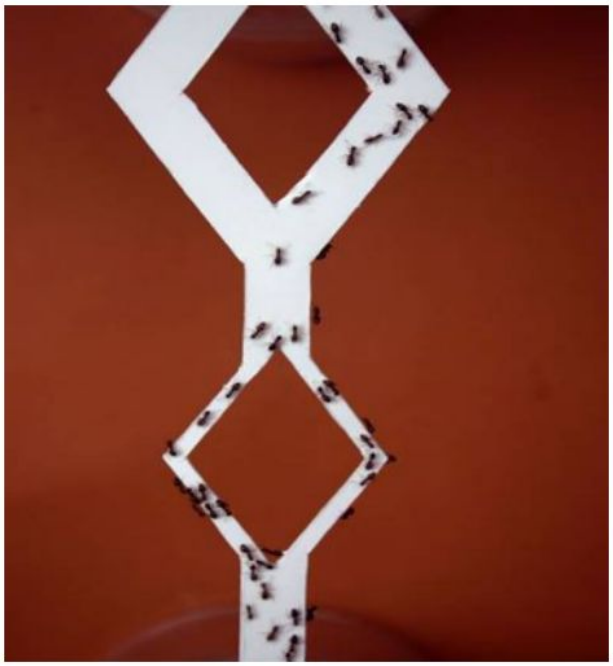


Collective Intelligence…

- It can be understood as an emergent property from the interactions among:
  - Data-information-knowledge
  - Software-hardware
  - Individuals (those with new insights as well as recognized authorities) **that continually learn from feedback** to produce just-in-time knowledge for better decisions than these three elements acting alone

- Or it can be more narrowly understood as an emergent **property between people and ways of processing information.**

Collective intelligence
Fishes, Birds



HTML vs HTML5

- Both HTML and HTML5 are hypertext markup languages, primarily **used to develop web pages** or applications.
- HTML5 is the latest version of HTML and supports new functionalities like multimedia, new tags and elements as well as new APIs.
- HTML5 also supports audio and video.

HTML vs HTML 5…

HTML	HTML5
HTML does not provide native audio and video support.	HTML5 provides native audio and video support.
HTML only supports vector graphics if used in conjunction with different technologies like Flash.	HTML5 supports SVG (Scalable Vector Graphics), Canvas, and other virtual vector graphics.
HTML doesn't allow users to draw shapes such as circles, triangles, and rectangles.	HTML5 allows users to draw shapes such as circles, triangles, and rectangles.
HTML only uses browser cache and cookies to store data temporarily.	HTML5 uses web SQL databases, local storage, and application cache for storing data temporarily.
Longer document type declaration.	Shorter document type declaration.
Longer character encoding declaration. Uses the ASCII character set.	Shorter character encoding declaration. Uses the UTF-8 character set.
Compatible with almost all browsers.	Only compatible with newer browsers, considering there are many new tags and elements which only some browsers support.
Built based on Standard Generalized Markup Language (SGML).	HTML5 has improved parsing rules providing enhanced compatibility.

HTML 5….

- It supports
  - **SVG (Scalable Vector Graphics)** is used to define vector-based graphics for the Web. It defines the **graphics in XML format**. Every element and every attribute in SVG files can be animated.

- HTML5 allows JavaScript to run within a web browser. (Supports ReactJS)

  - **Standard Generalized Markup Language** - SGML is a standard that defines how to specify the generalized markup languages for various documents.
  - It is not a document language in itself but rather a description regarding how to define them.

HTML 5….

- In HTML5**, web SQL databases** are used to store data temporarily. Previously, only the browser cache was used.

- New elements have been added – time, summary, aside, **audio**, command, and data.

HTML 5…

- It has introduced new multimedia features which supports both audio and video controls by using **<audio> and <video>** tags.
- Enrich semantic content by including <header> <footer>, <article>, <section> and <figure> are added.
- **Drag and Drop-** The user can grab an object and drag it further dropping it to a new location.
- **Geo-location services-** It helps to locate the geographical location of a client.
- **Web storage facility** - store data on the web browser.
- Uses SQL database to store data offline.
- Allows drawing various shapes like **triangle, rectangle, circle**, etc.
- Capable of handling incorrect syntax.

HTML Tags

Tag	Description
<!--...-->	Specifies a comment
<!DOCTYPE>	Specifies the document type
<a>	Specifies an anchor
<abbr>	Specifies an abbreviation
<acronym>	Deprecated: Specifies an acronym
<address>	Specifies an address element
<applet>	Deprecated: Specifies an applet
<area>	Specifies an area inside an image map
<article>	This element is used to define an independent piece of content in a document, that may be a blog, a magazine or a newspaper article.
<aside>	It specifies that article is slightly related to the rest of the whole page.
<audio>	It is used to play audio file in HTML.

HTML Tags…

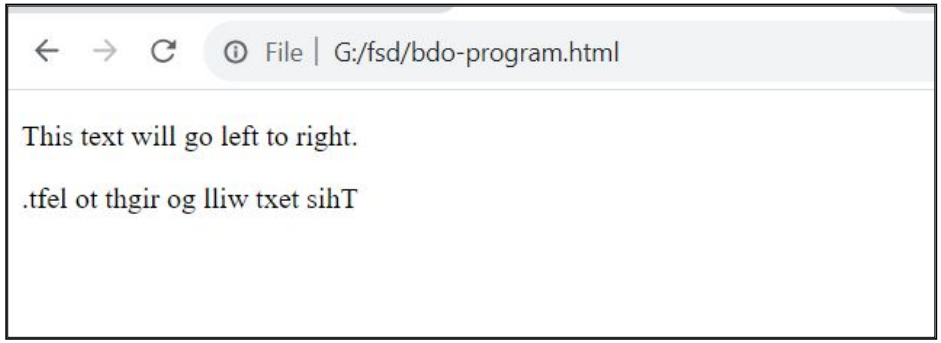
<bdi>	The bdi stands for bi-directional isolation. It isolates a part of text that is formatted in other direction from the outside text document.
<canvas>	It is used to draw canvas.
<data>	It provides machine readable version of its data.
<datalist>	It provides auto complete feature for textfield.
<details>	It specifies the additional information or controls required by user.
<dialog>	It defines a window or a dialog box.
<figcaption>	It is used to define a caption for a <figure> element.
<figure>	It defines a self-contained content like photos, diagrams etc.
<footer>	It defines a footer for a section.
<header>	It defines a header for a section.

HTML Tags…

<main>	It defines the main content of a document.
<mark>	It specifies the marked or highlighted content.
<menuitem>	It defines a command that the user can invoke from a popup menu.
<meter>	It is used to measure the scalar value within a given range.
<nav>	It is used to define the navigation link in the document.
<progress>	It specifies the progress of the task.
<rp>	It defines what to show in browser that don't support ruby annotation.
<rt>	It defines an explanation/pronunciation of characters.
<ruby>	It defines ruby annotation along with <rp> and <rt>.
<section>	It defines a section in the document.
<summary>	It specifies a visible heading for <detailed> element.
<svg>	It is used to display shapes.
<time>	It is used to define a date/time.
<video>	It is used to play video file in HTML.
<wbr>	It defines a possible line break.

Example 1

```
<!DOCTYPE html>
<html>
<head>
<title>Text Direction Example</title>
</head>
<body>
<p>This text will go left to right.</p>
<p><bdo dir="rtl">This text will go right to left.</bdo></p>
</body>
</html>
```



Example 2

```
<!DOCTYPE html>
<html>
<head>
<title>Double Quote Example</title>
</head>
<body>
<p>Amit is in Spain, <q>I think I am wrong</q>.</p>
</body>
</html>
```

This text will go left to right.

.tfel ot thgir og lliw txet sihT

Amit is in Spain, “I think I am wrong”.

HTML 5 Tags

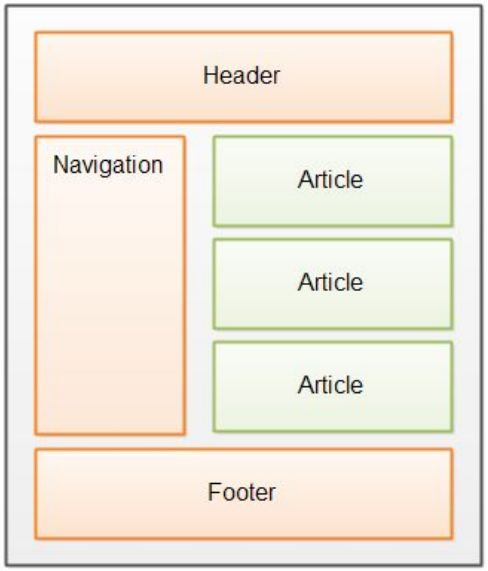
Tags (Elements)	Description
<article>	Represents an independent piece of content of a document, such as a blog entry or newspaper article
<aside >	Represents a piece of content that is only slightly related to the rest of the page.
<audio>	Defines an audio file.
<canvas>	This is used for rendering dynamic bitmap graphics on the fly, such as graphs or games.
<command>	Represents a command the user can invoke.
<datalist>	Together with the a new list attribute for input can be used to make comboboxes
<details>	Represents additional information or controls which the user can obtain on demand
<embed>	Defines external interactive content or plugin.
<figure>	Represents a piece of self-contained flow content, typically referenced as a single unit from the main flow of the document.

HTML 5 Tags…

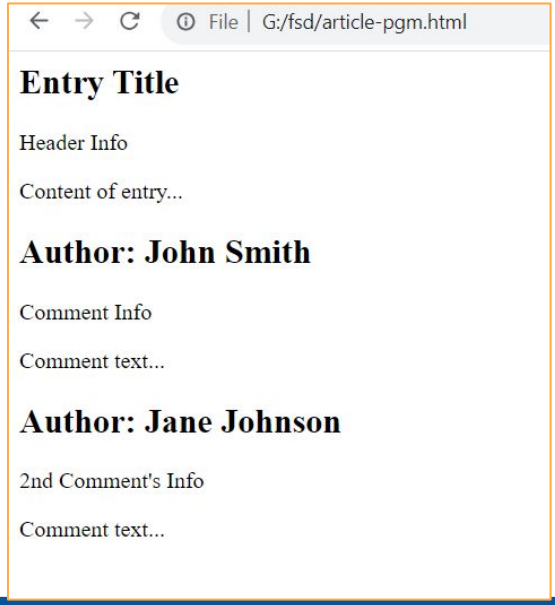
<footer>	Represents a footer for a section and can contain information about the author, copyright information, et cetera.
<header>	Represents a group of introductory or navigational aids.
<hgroup>	Represents the header of a section.
<keygen>	Represents control for key pair generation.
<mark>	Represents a run of text in one document marked or highlighted for reference purposes, due to its relevance in another context.
<meter>	Represents a measurement, such as disk usage.
<nav>	Represents a section of the document intended for navigation.
<output>	Represents some type of output, such as from a calculation done through scripting.
<progress>	Represents a completion of a task, such as downloading or when performing a series of expensive operations.
<ruby>	Together with <rt> and <rp> allow for marking up ruby annotations.
<section>	Represents a generic document or application section
<time>	Represents a date and/or time.
<video>	Defines a video file.
<wbr>	Represents a line break opportunity.

HTML5 structure

Article Tag




```
<article>
  <header>
    <h1>Entry Title</h1>
    <p>Header Info</p>
  </header>
  <p>Content of entry...</p>
  <article>
    <header>
      <h2>Author: John Smith</h2>
      <p>Comment Info</p>
    </header>
    <p>Comment text...</p>
  </article>
  <article>
    <header>
      <h2>Author: Jane Johnson</h2>
      <p>2nd Comment's Info</p>
    </header>
    <p>Comment text...</p>
  </article>
</article>
```



Canvas tag

- The **<canvas> tag is used to draw graphics**, on the fly, via scripting (usually JavaScript).
- The <canvas> tag is transparent, and is only a container for graphics, one must use a script to actually draw the graphics.
- Any text inside the <canvas> element will be displayed in browsers with JavaScript disabled and in browsers that do not support <canvas>.

Syntax

<canvas id = "script"> Contents... </canvas>

Canvas tag

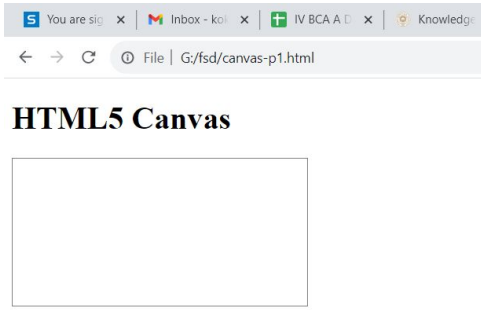
Attributes

The <canvas> tag accepts two attributes which are described below:

Attributes	Descriptions
height	This attribute is used to set the height of the canvas by taking the value in pixels and its default value is 150.
width	This attribute is used to set the width of the canvas by taking the value in pixels and its default value is 300.

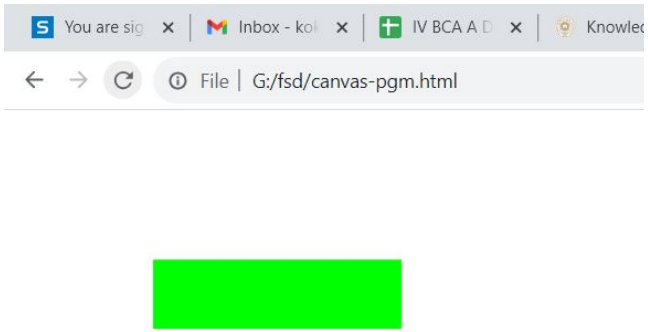
Canvas Example 1

```
<!DOCTYPE html>
<html>
<body>
<h1>HTML5 Canvas</h1>

<canvas id="myCanvas" width="300" height="150"
style="border:1px solid grey"></canvas>

</body>
</html>
```



Canvas Example 2

```
<canvas id="myCanvas">
Your browser does not support the canvas tag.
</canvas>

<script>
let canvas = document.getElementById("myCanvas");
let ctx = canvas.getContext("2d");
ctx.fillStyle = "#00ff00";
ctx.fillRect(100, 100, 180, 180);
</script>
```



Canvas Example 3

```
<!DOCTYPE html>
<html>

<body>
   <canvas id="arc2"
         width="200"
         height="200"
         style="border:1px solid black">
   </canvas>

   <script>
      var c = document.getElementById("arc2");
      var cx = c.getContext("2d");
      var grd = cx.createRadialGradient
         (100, 100, 5, 100, 100, 100);
      grd.addColorStop(0, "red");
      grd.addColorStop(1, "green");
      cx.fillStyle = grd;
      cx.fillRect(0, 0, 200, 200);
   </script>
</body>

</html>
```