

Web 1.0 + Web 2.0

(Web 1.0)

5.1.1 Web Information Retrieval (Web IR)

Information retrieval: The science surrounding search engines is commonly referred to as information retrieval, in which algorithmic principles are developed to match user interests to the best information about those interests. Search and Information Retrieval on the Web has advanced significantly in last few years.

This section will provide a broad coverage of the important issues in information retrieval and search engines by including underlying models and current research directions and to help understand search engines, evaluate and compare them, and modify them for specific applications.

Search and Information Retrieval

By all measures, the Web is enormous and growing at a staggering rate, which has made it increasingly intricate and crucial for both people and programs to have quick and accurate access to Web information and services. Thus, it is imperative to have tools for efficient and effective resource and knowledge discovery. Search engines have assumed a central role in the World Wide Web's infrastructure as its scale and impact have escalated. The unabated growth of the Web and the increasing requirement of the search engine to anticipate and infer the information needs and provide relevant results, has fostered the development of the field of Web Information Retrieval (Web IR).

Search on the web is a daily activity for many people throughout the world. There are applications that involve search present in everyday tasks. The field of computer science that is most involved with R&D for search is information retrieval (IR)

get/retrieve

SAOSS R . SASOR

"Information retrieval is a field concerned with the structure, analysis, organization, storage, searching, and retrieval of information." (Salton, 1968). It is a general definition that can be applied to many types of information and search applications. The primary focus of IR since the 50s has been on text and documents.

The significance of documents and text in information retrieval can be depicted by its comparison with database records.

How Documents are different from Database Records?

A document is a significant text content that comprises of some structure. e.g., web pages, email, books, news stories, scholarly papers, text messages, Word™, Powerpoint™, PDF, forum postings, patents, IM sessions, etc.

On the other hand, database records (or tuples in relational databases) are typically made up of well defined fields (or attributes). e.g., bank records with account numbers, balances, names, addresses, social security numbers, dates of birth, etc.

It is always easier to compare fields with well defined semantics to queries in order to find matches than to text.

Example: In a bank database query finding records with balance > \$50,000 in branches located in Amherst, MA. finds matches easily by comparison with field values of records whereas, in a search engine query to find bank scandals in western mass, this text must be compared to the text of entire news stories. Therefore, comparing the query text to the document text and determining what is a good match is the core issue of information retrieval because exact matching of words is not enough. As there are many different ways to write the same thing in a "natural language" like English.

5.1.2 DIMENSIONS OF IR

Although central to Information Retrieval, It is more than just text, and web search as people are also involved in doing the IR work with different media, different types of search applications, and different tasks.

Other Media being the new applications that increasingly involve new media e.g., video, photos, music, speech etc. Just like text, the content here is difficult to describe and compare and so text may be used to represent them (e.g. tags)

CONTENT	APPLICATIONS	TASKS
Text	Web Search	Ad hoc search
Images	Vertical Search	Filtering
Video	Enterprise Search	Classification
Scanned docs	Desktop Search	Question Answering
Audio	Forum Search	
Music	P2P Search	
	Literature Search	

5.1.3 WEB IR TASKS

Web Information Retrieval research is typically organized in tasks with specific goals to be achieved. Existing tasks have changed frequently over the years due to the emergence of new fields.

Some of the main tasks as well as new and emerging ones are as follow:

1. **Ad-Hoc Retrieval:** It ranks documents using non-constrained queries in a fixed collection. This is the standard retrieval task in Web IR. The purpose in this search is to find relevant documents for an arbitrary text query .It is created in order to get information when the need arises.
2. **Filtering:** It processes a stream of information to match one's static set of likes, tastes and preferences. This method identifies relevant user profiles for a new document and selects documents using a fixed query in a dynamic collection. For example, "Retrieve all documents related to 'Research in India' from a continuous feed".
3. **Topic Distillation :** Finding short lists of good entry points to a broad topic. For example, "Find relevant pages on the topic of Indian History".
4. **Homepage Finding :** Find the URL of a named entity. For example, "Find the URL of the Indian High Commission homepage".
5. **Adversarial Web IR :** Develop methods to identify and address the problem of web spam, namely link spamming that affect the ranking of results.
6. **Summarization :** It is a concept which involves techniques for finding a compact description of a dataset and produces a relevant summary of a single or multiple documents.
7. **Visualization :** any technique used for creating images, diagrams, or animations to communicate a message and Develop methods to present and interact with results.
8. **Question Answering** Retrieve small snippets of text that contained an answer for open-domain or closed-domain questions so as to give a specific answer to a question

9. Categorization/ Clustering Grouping documents into pre-defined classes/ adaptive clusters. This process analyzes a particular input and assigns it to (one or more) category thus identifying relevant labels for documents.

5.1.4 WEB IR MODELS

Retrieval models form the theoretical basis for computing the answer to a specified query. A Retrieval Model is a formal representation of the process of matching a query and a document. The model of Web IR can be defined as a set of premises and an algorithm for ranking documents with regard to a user query. A Web IR model is a quadruple $[D, Q, F, R(q_i, d_j)]$ where D is a set of logical views of documents, Q is a set of user queries, F is a framework for modeling documents and queries, and $R(q_i, d_j)$ is a ranking function which associates a numeric ranking to the query q_i and the document d_j .

The model is characterized by four parameters:

1. Representations for documents and queries, which define the models structure and its empirical architecture.
2. Matching strategies for assessing the relevance of documents to a user query, which involves learning parameters from query and evaluating the query according to the pattern found.
3. Methods for ranking query output i.e how to represent the output of the ranking query and how to analyse it.
4. Mechanisms for acquiring user-relevance feedback.

Retrieval models can describe the *Computational process*, for example, how the documents are ranked and note that how documents or indexes are stored is implementation. The Retrieval models can also attempt to describe the *User process*, for example, the information need and interaction level. The Retrieval variables are usually depicted by queries, documents, terms, relevance judgments, users & information needs. They can have an explicit or implicit definition of relevance.

5.1.4.1 DIMENSIONS OF IR MODEL

There are two prominent dimensions for defining different kind of IR model:

- Computational process : This kind of Process Deals with the IR models on the Mathematical Basis. According to the this dimension, the models can be classified into three types:
 - Set theoretic models
 - Algebraic models
 - Probabilistic models.

In the following sections, we describe instances of each type.

I) Set theoretic models

In these kind of models Documents are represented by sets that contain terms and Similarities are derived using set theoretic operations. Implementations of these models include the Standard Boolean Model, the Extended Boolean Model and the Fuzzy Model. The strict Boolean and fuzzy-set models are preferable to other models in terms of computational requirements, which are low in terms of both the disk space required for storing document representations and the algorithmic complexity of indexing and computing query document similarities.

II) Algebraic models

Documents are represented as vectors, matrices or tuples algebraic models. These are transformed using algebraic operations to a one-dimensional similarity measure. Implementations include the Vector Space Model and the Generalized Vector Space Model. The strength of this model lies in its simplicity. Relevance feedback can be easily incorporated into it. However, the rich expressiveness of query specification inherent in the Boolean model is sacrificed.

III) Probabilistic models

In probabilistic models, Document's relevance is interpreted as a probability. Documents and queries similarities are computed as probabilities for a given query. The probabilistic model takes these term dependencies and relationships into account and, in fact, specifies major parameters such as the weights of the query terms and the form of the query document similarity. Due to its simplicity and efficient computation, the Vector Model is the most widely used model in IR. The model requires term occurrence probabilities in the relevant and irrelevant parts of the document collection, which are difficult to estimate. However, this model serves an important function for characterizing retrieval processes and provides a theoretical justification for practices previously used on an empirical basis (for example, the introduction of certain term-weighting systems).

- **User Process:** This kind of process deals with the models on the Relevance Basis. According to this dimension, the models can be classified into four types:
 - Classical models
 - Topical relevance models
 - User relevance models
 - Linear feature-based model

In the following sections, Instances of each type are described.

I. Classical models:

It has Query languages, Indexing (Boolean) and also Introduces ranking and weighting (Vector Space). It can be classified into following types.

1. Generic Document Model :

2. The Boolean Model: It is a Simple model based on Boolean algebra in which

-Term weights are binary $w_{i,j} \in \{0, 1\}$

If $w_{i,j} = 1$ — term present,

else if $w_{i,j} = 0$ — term not present

- Queries are Boolean expressions E.g., $q = k_a \wedge (k_b \vee \neg k_c)$

-Documents are considered relevant if the query evaluates to 1(true)

3. The Vector Space Model : It is Simple model, based on linear algebra in which Term weights are not binary and it allows computing a continuous degree of similarity between queries and documents Thus, allows ranking documents according to their possible relevance.

- Documents are represented as vectors $\sim d_j = (w_{1,j}, w_{2,j}, \dots, w_{t,j})$ where $w_{i,j}$ is the weight of term i in document j .
- Queries are also vectors $\sim q = (w_{1,q}, w_{2,q}, \dots, w_{t,q})$.
- Vector operations can be used to compare queries \times documents (or documents \times document)

5. The Probabilistic Model: It Models the IR problem in a probabilistic framework and estimates the probability of document d_j being relevant to the end user .

It assumes that: 1. the relevance probability depends only on the query and the document

2. there is a subset R of relevant documents.

3. index terms are independent

II. Topical relevance models:

Topical relevance is relevance to a subject (topic) and In its most simple form, matching words in documents and queries. With search engines, topical relevance is mainly used in conjunction with backlinks (incoming links). Websites that carry similar content are said to have topical relevance. Backlinks from websites that are topically relevant have more impact on a website's position in search results than backlinks from sites that are not related.

III. User relevance models:

The user issues a (short, simple) query. The system returns an initial set of retrieval results. The user marks some returned documents as relevant or nonrelevant. The system computes a better representation of the information need based on the user feedback.

- It comprises of combination of evidence, features, query language (inference network, Inquiry)

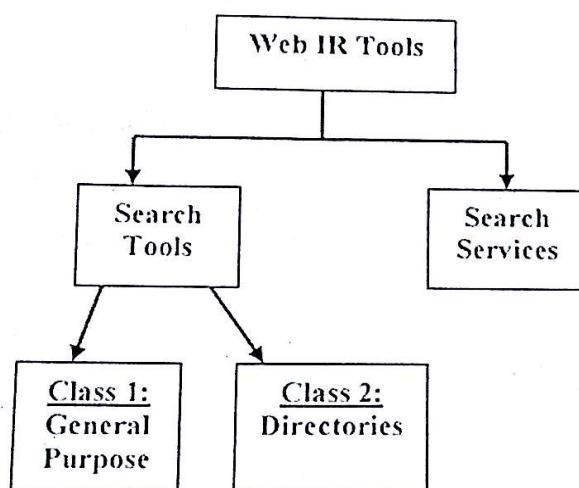
IV. Linear feature-based models:

As its name implies, the model's scoring function is comprised of a linear combination of features. One of the main benefits of such models is their ability to combine many different kinds of features in a straightforward manner. These are the prominent features of this model:

- Learning weights, arbitrary features, optimizing effectiveness measures (Ranking SVM, Linear Discriminant, MRF)
- "Learning to Rank", learning ranking rather than classification, preferences

5.1.5 WEB IR TOOLS

These are Automated methods for retrieving information on the Web and can be broadly classed as search tools or search services.



5.1.5.1 Search Tools

The Search tools employ robots for indexing Web documents. They feature a user interface for specifying queries and browsing the results. At the heart of a search tool is the search engine, which is responsible for searching the index to retrieve documents relevant to a user query. Search tools can be distinguished into two categories on the transparency of the index to the user. The two class categories are depicted along the following dimensions

- Methods for Web navigation,
- Indexing techniques,
- Query language or specification scheme for expressing user queries,
- Strategies for query-document matching, and
- Methods for presenting the query output.

- **Class1 search tools:** General Purpose Search Engine These tools completely hide the organization and content of the index from the user. Example: AltaVista, Excite, Google, Infoseek, Lycos

- **Class 2 search tools:** Subject Directories These feature a hierarchically organized subject catalog or directory of the Web, which is visible to users as they browse and search. Example: Yahoo!, WWW Virtual Library and Galaxy.

5.1.5.2 Search Services

The Search services provide users a layer of abstraction over several search tools and databases and aim at simplifying the Web search. Search services broadcast user queries to several search engines and various other information sources simultaneously. Then they merge the results submitted by these sources, check for duplicates, and present them to the user as an HTML page with clickable URLs. Example: MetaCrawler.

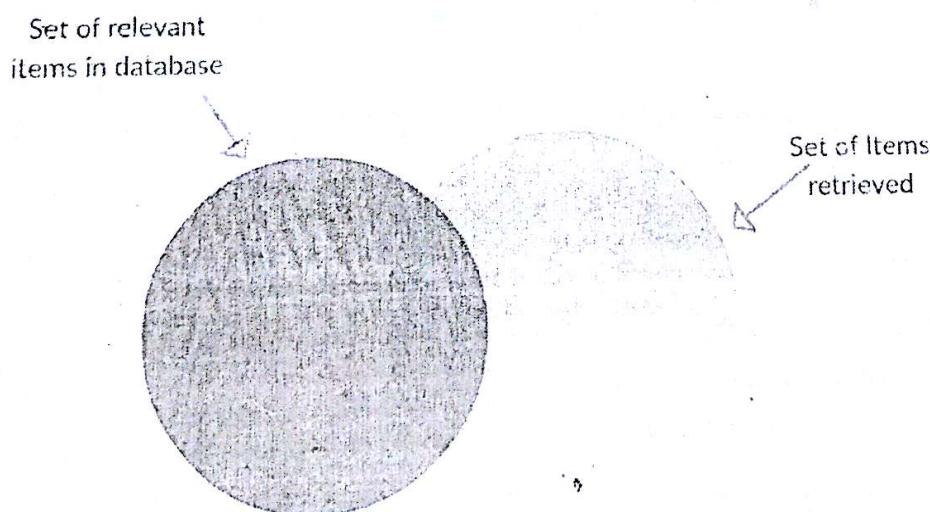
5.1.6 QUANTIFYING THE QUALITY OF WEB IR RESULTS

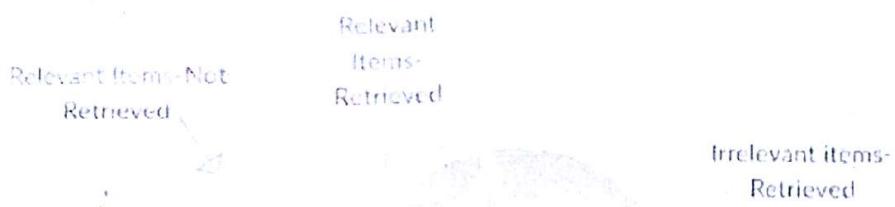
There are various ways to measure how well the retrieved information matches the intended information. The Web IR system might evaluate several aspects, namely, the assistance in formulating queries, the speed of retrieval, the resources required, the presentation of documents, the ability to find relevant documents, the appealing to users (market evaluation). The Evaluation is generally comparative. In an Information Retrieval scenario, the most common evaluation is retrieval effectiveness and the effect of indexing exhaustivity and term specificity on retrieval effectiveness can be explained by two widely accepted measures Precision & Recall.

Precision and Recall

Precision and Recall are the basic measures used in evaluating search strategies and they can be explained with the help of the diagrams below:

In the figures, there is a set of records in the database which is relevant to the search topic and the Records are assumed to be either relevant or irrelevant (these measures do not allow for degrees of relevancy). The actual retrieval set may not perfectly match the set of relevant records.





RECALL is the ratio of the number of relevant records retrieved to the total number of relevant records in the database. It is usually expressed as a percentage.

$$\text{RECALL: } \frac{A}{A+B} \times 100\%$$

Where

A is Number of relevant records retrieved and
B is Number of relevant records not retrieved

A perfect Recall score of 1.0 means that all relevant documents were retrieved by the search (but says nothing about how many irrelevant documents were also retrieved).

PRECISION is the ratio of the number of relevant records retrieved to the total number of irrelevant and relevant records retrieved. It is usually expressed as a percentage.

$$\text{PRECISION: } \frac{A}{A+C} \times 100\%$$

Where

A is the Number of relevant records retrieved and
C is the Number of irrelevant records retrieved

A perfect Precision score of 1.0 means that every result retrieved by a search was relevant (but says nothing about whether all relevant documents were retrieved)

Recall and precision are inversely related and Much of this relationship has to do with language. If the goal of a search is comprehensive retrieval, then the searcher must include synonyms, related terms, broad or general terms, etc. for each concept.

He may decide to combine terms using Boolean rather than proximity operators. In addition, some secondary concepts may be omitted. As a consequence of these decisions, precision will suffer. Because synonyms may not be exact synonyms the probability of retrieving irrelevant material increases. Broader

terms may result in the retrieval of material which does not discuss the narrower search topic. Using Boolean operators rather than proximity operators may increase the probability that the terms won't be in context. Unfortunately, if the searcher doesn't use these techniques he won't achieve high recall.

EXAMPLE

Problem: Assume the following:

- A database contains 100 records on a particular topic
- A search was conducted on that topic and 50 records were retrieved.
- Of the 50 records retrieved, 30 were relevant.

Calculate the precision and recall scores for the search.

Solution: Using the designations above:

- A = The number of relevant records retrieved,
- B = The number of relevant records not retrieved, and
- C = The number of irrelevant records retrieved.

In this example

$$A = 30, B = 70 \text{ (} 100 - 30 \text{)} \text{ and } C = 20 \text{ (} 50 - 30 \text{)}$$

$$\text{Recall} = (30 / (30 + 70)) * 100\% \Rightarrow 30/100 * 100\% = 30\%$$

$$\text{Precision} = (30 / (30 + 20)) * 100\% \Rightarrow 30/50 * 100\% = 65\%$$

5.1.7 Some Other Measures:

There are a number of more advanced and specific types of precision and recall measures that are used as modern evaluation measures.

Fallout is a measure of how quickly precision drops as recall is increased. Fallout is defined as the probability to find an irrelevant among the retrieved documents.

R-precision is the precision at R where R is the number of relevant documents in the collection for the query. It is the precision after R retrieved documents, where R is the number of relevant documents that exists for that query. An R-precision of 1.0 is equivalent to perfect relevance ranking and perfect recall. However, a typical value of R-precision which is far below 1.0 does not indicate the actual value of recall (since some of the relevant documents may be present in the hit-list beyond point R).

Initial precision is the precision at recall 0% in the interpolated precision-recall graph. It is an indication of relevance ranking of the top few hits. Similarly, one can define a final precision that is the precision at 100% recall. Final precision indicates how far down one need to go in the hit-list to find all relevant documents.

Precision at 0.5 Recall is the precision after half the relevant documents have been retrieved and Average Precision is the average of precision scores at every relevant document in the retrieved set.

Recall (1000) is the recall after 1000 retrieved documents. This is more practical than true recall over all documents since modern systems can return a huge number of results.

5.1.8 ISSUES IN IR

- Relevance

A relevant document contains the information that a person was looking for when they submitted a query to the search engine. Many factors influence a person's decision about what is relevant: e.g., task, context, novelty, style. There is topical relevance (same topic) vs. user relevance (everything else). There are retrieval models that define a view of relevance. There exist ranking algorithms that are used in search engines which are based on retrieval models. Also, most of the models describe statistical properties of text rather than linguistic i.e. counting simple text features such as words instead of parsing and analyzing the sentences

- Evaluation

Initially, originated in the Cranfield experiments in the 60s, there are experimental procedures carried out and measures are taken for comparing system output with user expectations. These IR evaluation methods are now used in many fields where they typically use test collection of documents, queries, and relevance judgments. Recall and precision are two examples of effectiveness measures.

- Users and Information Needs

Evaluation on the basis of search is user-centered where using keyword queries may often be poor descriptions of actual information needs. Here, interaction and context are important aspects for understanding user intent. So, query refinement techniques such as query expansion, query suggestion, relevance feedback may improve ranking.

"The web we know now, which loads into a browser window in essentially static screens, is only an embryo of the web to come."

This is stated by Darcy DiNucci in her 1999 article "fragmented future" in which she coined term web 2.0 to refer to the ever changing state of the web. This term was then popularized by Timothy O'Reilly at the O'Reilly media web 2.0 conference in late 2004.

In the last few years, there is a variety of developments on the world wide web and a evolutionary shift in the way it is used. Web 2.0 is all about users and content, instead of just surfing on the internet. The use of web is being shifted from passive consumption of content to more active collaboration. It focuses on user created content, usability and interoperability of data by the end users.

Web 2.0 is a catch all term used to describe this improved functionality of web. Although the term suggest a new version of world wide web but it does not refer to any actual change in technical specifications but rather it reflects the changes in the ways of utilization of web.

Web 1.0 vs Web 2.0

Web 1.0 websites functioned as a static outlet for information and products where content is posted by programmers to be viewed by users while Web 2.0 websites function as applications, or programs that can be updated from any computer and user actively contribute to the content of the sites.

The relationship of web 1.0 to web 2.0 is roughly the equivalence of pong to 'the matrix'. Web 1.0 is essentially a prototype- a proof of concept, the concept of interactive content universally accessible through a standard interface that is well known as web 2.0.

Web 2.0 websites are designed in such a way that they allow more user participation, that encourages users to add value to the application as they use it. Now the user are actively participating who actually creates the content rather than passively receiving it.

Instead of merely retrieving information from web 2.0 site, a user's participation is increased by allowing him to contribute to site's content by creating and maintaining a profile or account on the site and following other user's content or commenting on published articles.

The characteristics of web 2.0 are rich experience , user participation , dynamic content and scalability. Further characteristics such as openness, freedom and collective intelligence by way of user participation make essential attributes of web 2.0.

Advantages of web 2.0

1. With web 2.0, information can be pulled from a number of different places and it can be personalized to meet the needs of a single user.
2. web 2.0 makes the internet a true democratic system,digital democracy as information flows freely and people can express their ideas without fear of repression.
3. Internet has become the greatest medium of communication surpassing telephone and printing press.

Disadvantages of web 2.0

1. Internet creates dependence. Nowadays people have become heavily dependent on internet for almost every kind of information.
2. Security is a bigger issue with web 2.0,as many web services are offered free, that could be easily targeted by hackers hence compromising security and privacy of users.

Prominent examples of web 2.0 include social networking sites and social media sites(eg. Facebook , twitter) ,blogs , wikis, video sharing sites(eg. youtube) and web applications(apps).

5.2.1. Social web and Big data

With million of people, everywhere using social network sites to stay connected with their friends, discover new people and activities and share user created contents, Social web is becoming way of life.

"Social networks and Social media has long been driven by two things: The connection between the people that use them and the information they share."

Social web defines how web 2.0 websites tends to interact much more with the end user as integral part of the website either by adding comments on the content, uploading new content or adding user generated content(eg. Personal digital photos).

While gleaning insight and contextualizing interactions in social environments is nothing new but the challenge in doing so is the amount of information in doing so is the amount of information generated.

Some of the data analytics platforms that stand include

Apache Hadoop as used by Facebook and Yahoo and other impressive names in the industry

Hadoop is popular and based in the cloud.

Amazon Web Services this company has an extraordinary way to sense data giving it a clear competitive edge.

Harmovert Data Platform. Pentaho has the ability to offer a 360 degree view of the consumer using social media

Making sense of it and discovering new patterns or knowledge out of it and understanding how it can be used in a particular context requires all new technologies and techniques specifically the emerging field of Big data.

Social web

- produces huge amount of data through user's active collaboration
- user participates to create content of site.

Big data

- This enormous Data is needed to be analyzed in right direction to produce meaningful information

New trend services

• pr
an
a
b
t
i

Social web is a set of applications or platforms which uses web 2.0 technologies to facilitate user to actively collaborate for the content of the site. It involves sharing, creating of information, commenting over shared information.

The social aspect of web 2.0 technologies is bringing together people with similar tastes. As people become active on web information about their social relationships become more available and hence providing better insight into their behaviour and interest..

People's behaviour that is what facebook analytics judge to enable personalization and better user experience. So the posts shown on a particular user's wall is vaguely result of his own activities describing his/her interests.

Most important aspect to understand for the success of any business that uses social web is that it is not the information but its interpretation that makes it so useful and then comes into picture the role of big data technologies which help the social media marketers manage , analyse and leverage large social data sets to gain actionable insights and a clear competitive advantage. Therefore rightful analysis of big data gives tracks to build new products and services to better meet the future needs and expectations of users.

"Big Data means that by looking at the interests, location and search history of a user, social web marketers can know their users better than ever and it allows them with knowledge of a user base to target them better for providing various kind of services and expand their business."

A national incident, a weather phenomenon, an issue with distribution networks, unhappy customers – this is information that the organisation needs to act on swiftly in order to undertake the necessary action. In evaluating the advantages and disadvantages of social networking, it's best to err on the side of caution and protect your privacy.

Advantages:

1. Real time information sharing makes social websites necessary part of most people's day today life.
2. Commonality of interest can be widely explored through social web.
3. Worldwide connectivity makes users well informed.

Disadvantages:

1. Invasion of privacy and risks of frauds and identity theft.
2. Cyber bullying is a serious problem with social networks which may expose individuals to many forms of harassment or even inappropriate contacts.

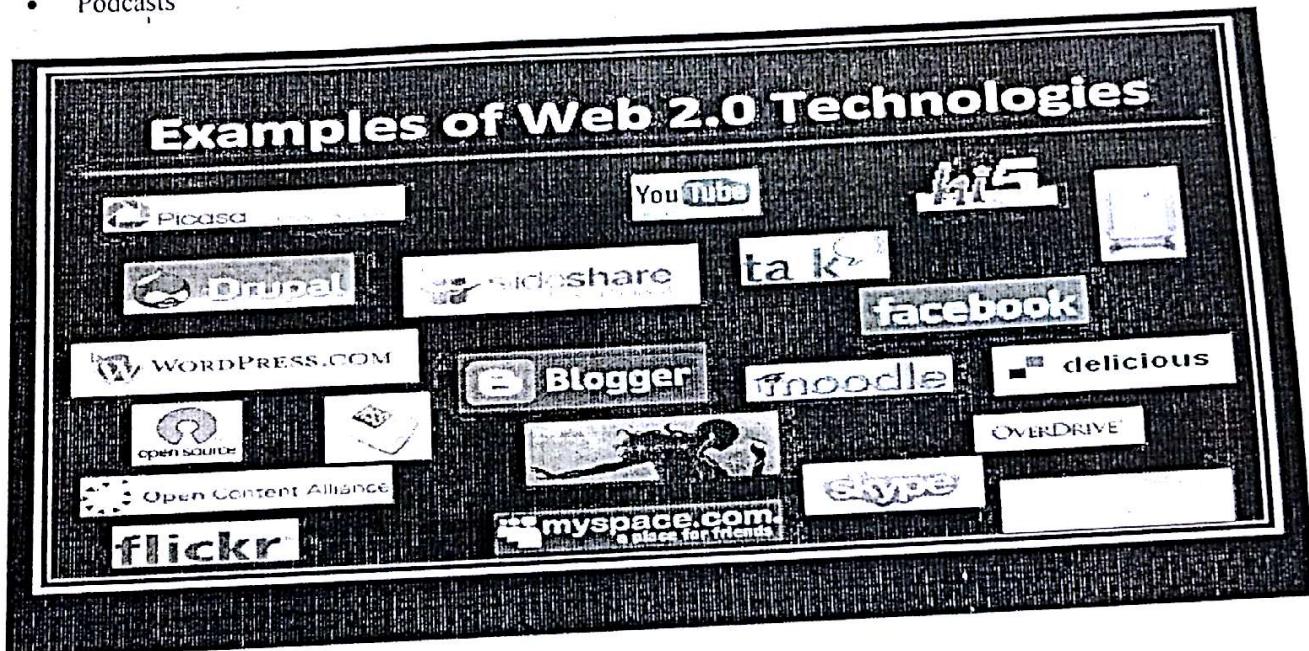
3. Backlashing is very common nowadays on social media. When potentially offensive content is posted online, the amount of feedback can be excessive and is often brutal. This is particularly true with highly opinionated subjects like politics and religion.

5.2.2 Web 2.0 technologies

Web 2.0 encourages a wider range of expressive capability, facilitates more collaborative ways of working, enables community creation, dialogue and knowledge sharing and creates a setting for learners to attract authentic audiences by various tools and technologies.

Few of the most popular web 2.0 tools are

- Weblog
- Social networks
- Podcasts



- Wikis
- Microblogs
- Social bookmarking
- E-portfolios

1. Blogging

Being a social creature, human beings have a natural tendency to express themselves and having audience for the same is overwhelming. The World Wide Web makes it possible for you to publish your thoughts (or whatever else you'd like) and distribute them out to the entire world (of Internet-connected computer users).

Nowadays, there are several good, reliable blogging tools available for free on the Web. You can set up your account and start blogging away within minutes. To understand this interactive tool of web 2.0, one must be familiar with the basic terminology of blogging.

- Weblog or blog : A Weblog, or "blog", is a personal journal or newsletter on the Web. Some blogs are highly influential and have

- Over 70 million blogs have been created since 2000
- Blogosphere growth over 120k blogs being created

enormous readership, while others are mainly intended for a close circle of family and friends. The power of Weblogs is that they allow millions of people to easily publish their ideas, and millions more to comment on them.

- **Blogger:** A blogger is someone who writes a blog.
- **Blogosphere:** Blogosphere is a word used to describe the online community of bloggers and their writings.
- **Permalink:** A permalink is the permanent identifier to a specific Weblog post or article. Bloggers love permalinks — they provide an easy way to capture specific references to posts or articles about which bloggers are writing.
- **Blogroll:** A blogroll is a list of blogs and bloggers that any particular blog author finds influential or interesting. Blogrolls indicate which online community a blogger is attracted to or belongs to, and they are part of the conversations of the blogosphere.

Few popular blogging resources are: wordpress, blogger, Technorati, problogger (blogging for beginners), about.com, edublogs etc.

2. Social networking

Social networking Websites provide a "virtual community" for people interested in a particular subject or just to "hang out" together.

Examples of Social Networks



Members create their own online profile page with biographical data, pictures, likes, dislikes and any other information they choose to post. Users can communicate with each other text, voice, chat, instant message, videoconference, and blogs, and the services typically provide a way for members to contact friends of other members, thus enabling everyone's "network" to grow.

Youtube is a special kind of social networking site which is basically a video sharing tool of web 2.0 along with features of social networking sites such as liking, sharing and commenting on a particular video. YouTube videos are "streamed" to users on the YouTube site and can also be "embedded" into other Websites and blogs with ease.

Internet with the fact that a podcast can be subscribed by its listeners, so that when new podcasts are released, they are automatically delivered, or fed, to a subscriber's computer or mobile device.

- Pod means a mobile playback device such as ipod or any other mp3 player and casting derived from broadcasting, hence podcasting is distribution of audio or video files over internet for playback devices

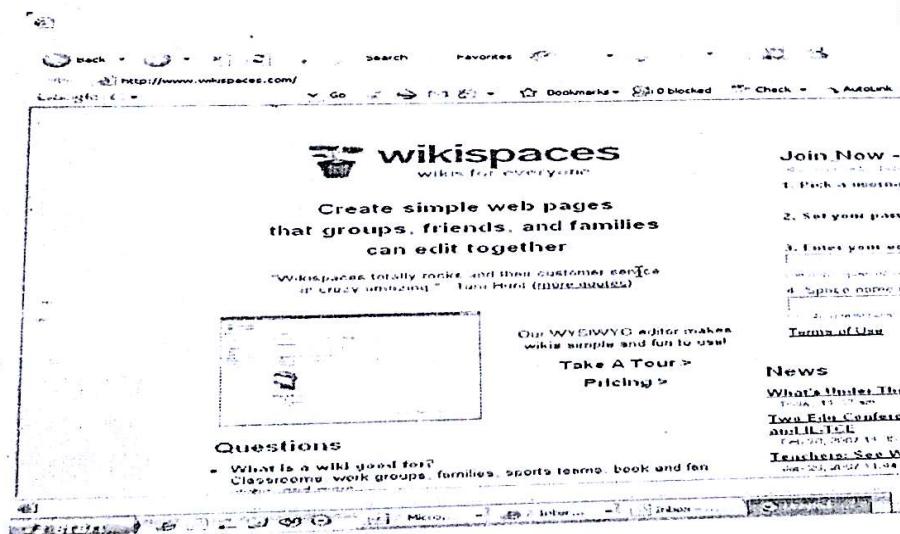
- If you are listening to a podcast on your computer then it will be played on the video or audio player installed already but if you want to subscribe to a podcast feeds then you'll need to install "podcatcher" software on your computer. The most popular podcatcher software is iTunes.
- The key difference between a podcast and a plain old audio file is the distribution model. Most podcasts are shared (syndicated) using the RSS format - Real Simple Syndication. Through RSS, anyone can subscribe to and catch podcasts, which will be automatically downloaded and managed by a podcatching program like iTunes.

4. Wikis

According to The Wiki Way, "Open editing has some profound and subtle effects on the wiki's usage. Allowing everyday users to create and edit any page in a Web site...encourages democratic use of the Web and promotes content composition by nontechnical users."

A wiki is a website that which allows its users to actively collaborate and modify its content and structure simply from the web browser. The collaborative encyclopedia "Wikipedia" is the most popular example of a wiki today.

- A single page in a wiki Website is referred to as a wiki page.
- The entire collection of wiki pages, which are usually well interconnected by hyperlinks, is "the wiki". A wiki is essentially a database for creating, browsing, and searching through information.
- A defining characteristic of wiki technology is the ease with which pages can be created and updated.
- Vandalism of wikis is a common problem i.e. due to its open nature anyone with internet and a computer can change wiki content to something offensive, adding nonsense or deliberately adding incorrect information) can be a major problem.
- Following are the platforms to create a wiki: wikispaces, pbwiki, wetpaint etc.
- Some popular wikis are Wikipedia, wikiwikiweb, memory alpha, wikivoyage and susning.nu etc.

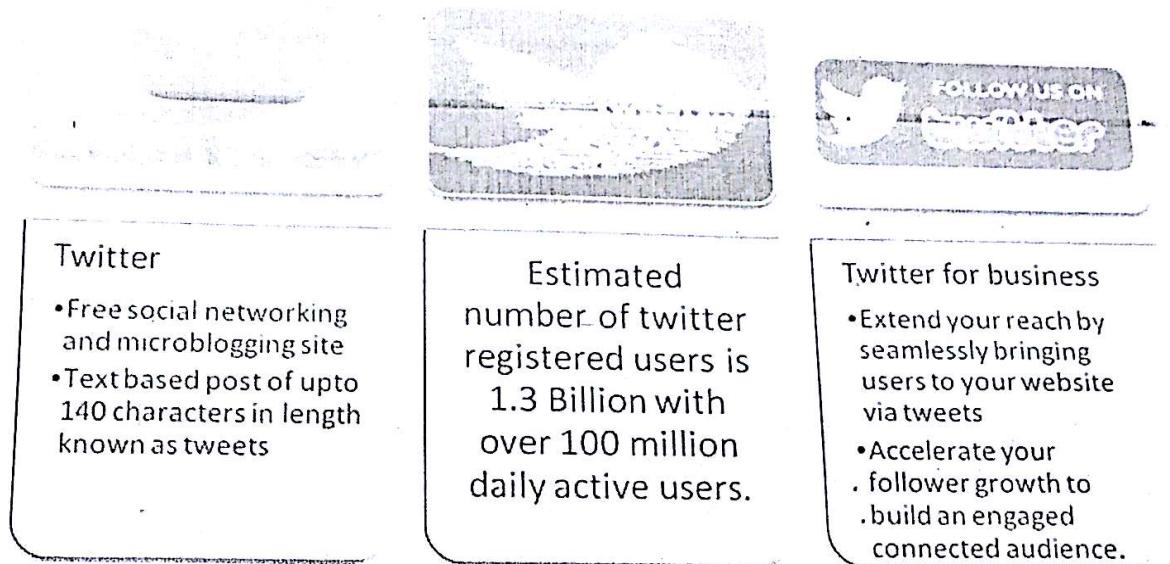


Home page of wikispace

The English Wikipedia has the largest user base among wikis on the World Wide Web and ranks in the top 10 among all Web sites in terms of traffic.

5. Micro-blogging
 Microblogging is the practice of posting small pieces of digital content—which could be text, pictures, links, short videos, or other media—on the Internet. Micro-blogging enable users to write brief messages, usually limited to less than 200 characters, and publish them via Web browser-based services, email, or mobile phones. The most popular micro-blogging service today is called Twitter. Micro-blogging is also known as 'mobile social-networking' or 'themed Instant Messaging'. It creates a sense of online community where groups of friends and professional colleagues connect to each other and frequently update content and follow each other's posts.

- **How it works:** This is one of the best examples of subscription services where the subscribers must typically create accounts, which are linked with cell phones, e-mail accounts, instant messaging, web pages—any medium they will use to send updates in order to post a microblog or to read those posted by others.
- These posts might consist of short text snippets (maximum number of characters specified by the application), a photo, an audio clip, or a few seconds of video, any of which can be shared publicly or with a selected group of subscribers.
- The posting of microblogs has enjoyed a popular upsurge in last few years year, with add-ons appearing regularly that enable more sophisticated updates and interaction with other applications.



Other microblogging sites are friendfeed, tumblr, plurk, yammer, shout'em, google talks etc.

6. Social Bookmarking

Social bookmarking is a way to store, organize, search, manage, and share collections of websites. In a social bookmarking system, users save links to websites that they want to remember and/or share.

- These bookmarks are usually public, but can be saved privately, or shared only with specified people or groups. People can access these bookmarks chronologically, by category or tags, or via a search engine.
- Many social bookmarking services provide web feeds (RSS) for their lists of bookmarks and tagged categories. This allows subscribers to become aware of new bookmarks as they are saved, shared, and tagged by other users.
- As these services have matured and grown more popular, they have added extra features such as ratings and comments on bookmarks, the ability to import and export bookmarks from browsers, emailing of bookmarks, web annotation, and groups or other social network features.

Several Popular social bookmarking sites include: Del.icio.us ,Digg , Technorati etc.

7. E-Portfolios

An E-portfolio is a digitized collection of artifacts including demonstration, accomplishments and resources that represents an individual, group or institution.

- It is a collection of work developed across varied contexts over time.
- It may include input text, electronic files, images, multimedia, blog entries, and hyperlinks.

There are three main types of e-portfolios:

- Developmental (e.g., working): It shows the advancement of skill over a period of time
- Assessment : It demonstrate skill and competence in a particular domain or area.
- Showcase: A showcase portfolio highlights stellar work in a specific area, it is typically shown to potential employers to gain employment. When it is used for job application it is sometimes called career portfolio.

Most e-portfolios are a mix of the three main types to create a hybrid portfolio.

Today, electronic portfolios are gaining popularity in:

- Schools and Higher education
- Job applications and Continuing professional development
- Assessment
- Recognition of prior learning

The development of technology is still growing up rapidly. Web 2.0 changes our use of Internet enormously, making internet a digital democracy .To sum up, Web 2.0 represents “for the user, by the user”.