

## Information Overload, Retrieval Strategies and Internet User Empowerment

Author: Dr. Christopher N. Carlson  
Affiliation: IWF Wissen und Medien gGmbH  
City: Göttingen  
Country: Germany  
Telephone: +49-551-5024-311  
Fax: +49-551-5024-322  
E-Mail: christopher.carlson@iwf.de

### Abstract:

Initial user benefits from search engine technology have been critically degraded over time by the rapid increase of Internet pages. Traditional retrieval strategies therefore yield increasingly poor results due to a dramatic increase in ballast in the results. Search engine users thus increasingly experience information overload. Technical approaches to dealing with this problem have caused an initial euphoria, yet have proven ineffective in solving the problem.

Enhancement of user empowerment in the area of Internet-based information retrieval must therefore be grounded in the augmentation of user capabilities. Alternative retrieval strategy approaches including a demonstration of their best areas of application are offered. Issues of information literacy and information anxiety are explored with regard to their relevancy to improving the retrieval skills of non-professional users.

Users must redefine their information needs and processing habits. Pre-filtering of perceived information requirements to reduce the amounts of information actively sought and acquired, while upgrading its quality, i.e. improving the precision/recall ratio, is a learnable trait. In terms of securing the future utility of inexpensive, universal-access online information exchange forums such as the Web, it is important that non-professional users learn to navigate successfully in an excessively information-rich environment.

-----

### 1. Data Smog - What's the problem?

Information overload is a phenomenon which has both objective and subjective causes. Objectively, the amounts of readily available information have increased exponentially in each of the last five decades. There is no indication at present that this rate of increase will not continue to apply in the foreseeable future. Plainly, nothing can (or should) be done about this. It is the logical result of a free information market coupled with technological progress.

The subjective component of information overload comes from our having more information available to us than we can readily assimilate; this is a perceived phenomenon - though it is clearly no less real on that account - and is sometimes referred to as "technostress"<sup>1</sup>. Perceived technostress induces a correlate perception that users are being controlled by ICT rather than

---

<sup>1</sup> Rosen, Larry; Weil, Michelle: "Technostress: Coping with Technology @ Work, @ Home, @ Play". John Wiley & Sons, 1997.

being empowered by it. Like any other kind of stress, technostress results in reduced intellectual performance and poor judgment; this is well-known to cognitive psychologists. In a sort of negative feedback loop, this partly causes, and is also partly a result of, haphazard and random use of ICT. Lack of a coherent conceptual knowledge management framework may also act as an aggravating factor.

Here are some pertinent facts which serve to illustrate the sheer force of numbers involved<sup>2</sup>:

3,062 - Number of U.S. newspaper and magazine articles published between 1997 and 1999 that talk about information overload

15,652 - Number of Web sites discussing information overload

2,892 - Number of titles in the Library of Congress in which the word *stress* appears

454 - Number of documents added to Lexis-Nexis each minute

40% - Percentage of workers who say their duties are interrupted more than six times an hour by intrusive communications

50% - Percentage of U.S. professionals who repeatedly receive messages that say the same thing

190 - Number of messages in all media sent and received daily by the average *Fortune* 1,000 office worker

80% - Percentage of information that is filed but never used

150 - Hours that the average person spends looking for lost information each year

71% - Percentage of workers who say their main job is tracking down information

44% - Percentage of managers who believe the cost of collecting information exceeds its value to their business

8 : 1 - Ratio of articles found on-line to those in newspapers

1 - Seconds it takes the World Wide Web to expand by 17 pages

7,349,000 - Projected increase in the number of URLs between 1997 and 2002

18,300,000 - Increase in the number of fax machines in the United States since 1987

2,809,000 - Increase in tons in the amount of paper used in offices from 1984 to 1998

---

<sup>2</sup> Source: "Data Data". Inc Magazine; January 1, 1999. URL: <http://www.inc.com/magazine/19990101/715.html> (accessed June 3, 2003)

In the historical perspective, more information has almost always been a very good thing. Information made possible the dissemination of culture, the development of commerce and technology - and it was one of the main driving forces behind the large-scale establishment of democracy and human rights. There was quite literally no downside to having more information. The dissemination of information empowered people. Oddly enough, the Information Age has been named for something which once conferred only benefits, and which is now increasingly seen as a problem.

The signal-to-noise ratio is an often-used metaphor for describing information overload. It was originally an engineering term from the audio industry and was used to describe the proportion of desired sound, i.e. music, to unwanted sounds such as the crackling noise of an old vinyl long-playing record. In the context of the Information Age, the term is used to describe the proportion of useful information found to all information found. Apart from the increase in information in absolute terms, there is, however, the additional problem of the relative decline in relevancy or pertinence of returned documents. The ease and low-cost of online publishing - once one has a text in a machine-readable format, and once one has a homepage, the *additional* cost of putting the text online is quite negligible - has led to a predictable glut of information which can, in all fairness, only be described as trivial or useless. (In fact, there are any number of Webpages that actually self-classify themselves as "Useless Knowledge Page" and the like<sup>3</sup>. These pages provide people with such information as that the Indian epic poem the "Mahabhrata" is eight times longer than "The Iliad" and "The Odyssey" combined.<sup>4</sup> But even without this deliberate attempt to achieve irrelevancy, it is plain that a query about any given matter stands in a sort of competition with vast amounts of information which would be suitable to answering some other query, but not this particular one.

This means that the signal-to-noise ratio has been critically degraded - with dramatic effects on precision and recall.

Much the same thing has happened with e-mail. The technology itself is extremely vulnerable to uncontrolled proliferation. As with Web publishing, there are no real additional costs to e-mailing once one has a provider and a machine-readable text. Sending a mail to dozens of people is as cheap and as easy as sending the same mail to one recipient only. One need only have the e-mail address in one's directory. By combining addresses into personalized mailing lists, it is possible with only a very few keystrokes to send large amounts of unwanted mail to people one barely knows. Given that this is so, it was only a question of time until someone invented spam - unsolicited commercial bulk e-mail.

History records the following significant date<sup>5</sup>: On April 12, 1994, Laurence Canter and Martha Siegel, a married team of Arizona lawyers, took spamming to an entirely new level of abuse when they posted to over 6,000 Usenet newsgroups an unsolicited commercial offer to help immigrants enter an upcoming "Green Card lottery." Prior to that, spamming had been a fairly sporadic and even desultory phenomenon, mostly limited to off-topic postings in newsgroups.

---

<sup>3</sup> 29 on the Google engine alone (accessed June 18, 2003)

<sup>4</sup> <http://www.coolquiz.com/trivia/directory/directory.asp?dir=Miscellaneous> (accessed June 18, 2003)

<sup>5</sup> Schenk, David: "Data Smog: Surviving the Information Glut". 1997. URL: <http://www.salemstate.edu/~tevangs/overload.htm> (accessed June 16, 2003)

Not ten years later, spam has become a major problem in terms of information overload. According to a recent survey<sup>6</sup>, spam is up fivefold over the past 18 months, leaving the electronic mailboxes of Internet users jammed with billions of unwanted commercial e-mails. AOL blocks 780 million pieces of junk e-mail daily, or 100 million more e-mails than it delivers.

## 2. Technology isn't solving the problem

Recent information and communication technologies (ICT) have produced an impressive arsenal to address the problems stemming from information overload. The most important technology-based retrieval aids are either integrable into Internet search engines or can be brought to bear on downloaded search results from search engines.

What aids are available and how do they work?

- intelligent agents - Agents are mostly tools for information retrieval. Some applications are intelligent information retrieval (IR) interfaces, mediated searching and brokering, and clustering and categorization. An agent-based approach means that IR systems can be more scalable, flexible, extensible, and interoperable, using agents that route information, broker requests and share metadata.
- ranking algorithms - The methodology by which search engines calculate positioning results. Ranking algorithms can be influenced by a wide variety of factors including domain name, spiderable content, submission practices, HTML code and link popularity. Search engine ranking algorithms are closely guarded and constantly updated to attempt to filter out those sites which attempt to manipulate the results.
- cluster analysis - Cluster analysis is an exploratory multivariate statistical method that attempts to find the "natural" groupings of objects based on attribute information about the objects. In typical cluster analysis with multivariate data, objects are usually the variables, and records (cases) are the attributes. In genomic analysis, when clustering arrays, the arrays are the objects and genes are the attributes. When clustering genes, genes are the objects and arrays are the attributes. The rule of thumb is that whatever is being clustered (variables, records, arrays, genes) is the object. The end result of cluster analysis is a cluster image display (CID) containing the dendograms (tree diagrams) showing the grouping of arrays and genes according to the order in which they were joined during clustering.
- web mining / data mining - Data mining is the data-driven discovery and modelling of hidden patterns in large volumes of data. Data mining differs from the retrospective technologies above because it produces models - models that capture and represent the hidden patterns in the data. Via data mining, a user can discover patterns and build models automatically, without knowing exactly what she's looking for. The models are both descriptive and prospective. They address why things happened and what is likely to happen next. A user can pose "what-if" questions to a data-mining model that can not be queried directly from the database or warehouse.
- web graph algorithms - These proceed from a recurrent phenomenon on the Web: For any particular topic, there tend to be a set of "authoritative" pages focused on the topic, and a

---

<sup>6</sup> Vise, David A.: "AOL Joins Microsoft In a Reply to Spam" Feb. 21, 2003. URL: <http://www.washingtonpost.com/ac2/wp-dyn?pagename=article&node=&contentId=A38150-2003Feb20&notFound=true> (accessed June 19, 2003)

set of "hub" pages, each containing links to useful, relevant pages on the topic. This observation motivated the development of the search algorithms: Given a set of pages  $v$  and the interconnections  $e$  between them, the algorithm ranks the pages in  $v$  by their quality as hubs, and as authorities. The notions of good hubs and authorities are captured by numerical values with specific definitions and update rules. The hubs, nodes and authorities are rendered as a graph.

- personalization, recommendations and collaborative filtering - Personalization algorithms apply known facts about users to customize information services. Known user preferences are derived from log files of prior behaviour or from specific statements made in user profiles - or both. Recommendations and collaborative filtering extend this principle to preferences of other users. The database knows that significant numbers of users interested in "A" are also interested in "B". This fact can be used to give "B" a higher ranking than it would normally have based on the explicit formulation of the search argument, when it is found together with "A". A well-known example of recommendation and collaborative filtering is the Amazon website, where insets inform users that persons who have purchased the book one is currently looking at have also tended to buy certain other books as well. It is assumed analogously that similar user preferences will be consistently similar.

Search engines actually support very different retrieval strategies. To illustrate this, let us view a comparison of the retrieval functions of some common search engines<sup>7</sup>:

### **Alta Vista**

Techniques and features:

Boolean - must use and, or, not, near (10 words) in Advanced Search

Allows user-influenced results ranking

Ranking: title words or first few words

- Closer to each other
- Document has more of the words
- More copies of the words throughout

Parentheses for nesting

Can restrict to field (qualifiers)

### **Excite**

Techniques and features:

Concept based searching-use statistical strength of interrelationships between words

Creates its own knowledge base (or internal thesaurus)

QBE - "similar documents"

Boolean searches

Keyword searches

---

<sup>7</sup> Lager, Mark: "Spinning a Web Search". URL: <http://www.library.ucsb.edu/untangle/lager.html>. 1996, updated several times until December 1999. (accessed June 6, 2003)

Relevance - marked with red X  
Robot is called Architext

### **Infoseek**

Techniques and features:  
Weight terms (required, desirable, undesirable)  
Similar pages - QBE  
Boolean operators  
Natural language  
Search mechanisms

### **Lycos**

Techniques and features:  
Probabilistic retrieval  
Indexes top 100 words and 20 lines of abstracts  
Keyword searching  
Boolean searching  
Automatic truncation  
Adjacency 0.0 - 1.0  
Results categorized  
Terms in bold  
Relevancy: early on vs. farther down

### **Magellan**

Techniques and features:  
Reviewed by writers  
Boolean searching  
Green light for information for all age groups  
Web, ftp, gopher, newsgroups, telnet sites  
Browse directory or Use search engine  
Relevancy = frequency of words  
Browse button  
Robot named Verity  
Lists up to 20 pages at the bottom of the screen

### **Open Text**

Techniques and features:  
Boolean searching  
Field operators: anywhere, summary, title, first heading, URL  
Query-by-example

(This overview is several years old and thus likely somewhat out of date. It should not be taken as current information but rather as an illustration of the principle that different search engines

have different functional principles. As a result of this, the selection of the most suitable search engine for a given query depends to a large extent on the actual object of the query. For a current comparison of the performance features of the most popular search engines cf. the Search Engine Showdown<sup>8</sup> website.)

Also worth mentioning in this context is a discipline known variously as information design or information architecture which has evolved to meet the challenge of using information to provide meaningful communication. The discipline embraces such diverse fields as business administration, computer science, cognitive psychology, graphic and typographic design, and technical communication. Information architecture is the structuring of data to meet the informational and management requirements of an organization or group of people<sup>9</sup>.

Although these various technologies have some palliative and/or remedial effect, they do not actually solve the problem.

With regard to anti-spam filters and other technology-based remedies for the unwanted e-mail problem, these too have proven to be of limited effect only. Here the problem is also aggravated because spammers carefully study the functional methodologies so that they can better circumvent them.

Also, though it does not strictly belong to this subject, no great reliance should be placed upon increased standardisation of Web content data structures or on a more consistent application of metatagging, since these things - though desirable in themselves - are often impossible or impracticable under the actual production conditions of the WWW.

### 3. What really helps?

The first - and perhaps most important - step in dealing effectively with information overload is for users to achieve a better understanding of the Internet as an information resource. Realistic expectations about information availability serve to reduce technostress and its companion phenomena disappointment and frustration.

The question about the availability of information on the Internet is about having appropriate expectations of the types of information that can be found. Because of the vast quantity of data that is available, it is easy to think that any information one could want is available. Upon reflection, though, it is understandable why this is not true. Possible reasons that a particular resource or piece of information is not available are<sup>10</sup>:

- Publishing firms and authors who are paid to create and disseminate information are unlikely to circumvent the marketplace of information and make the same information available free of charge via the Internet.

---

<sup>8</sup> Notess, Greg: "Search Engine Showdown. The Users' Guide to Web Searching". URL: <http://www.searchengineshowdown.com/> (accessed June 16, 2003)

<sup>9</sup> For further information on this issue see Victor, Stephen P.: "Instructional Applications of Information Architecture". URL: [http://library.thinkquest.org/50123/info\\_arch.html](http://library.thinkquest.org/50123/info_arch.html) (accessed June 3, 2003)

<sup>10</sup> Cf. Hinchcliffe, Lisa J.: "The Electronic Library: The World Wide Web". URL: <http://alexia.lis.uiuc.edu/~janicke/ElecLib.html>. Updated: May 29, 1997. (accessed June 6, 2003)

- Providing information and maintaining access to it is time-consuming and costly. Thus, organizations or institutions that are willing to freely distribute their information may not be able to do so via the Internet for financial reasons.
- The demand for a particular type of information often determines its availability. Much of the information which could be made available never will be since the demand for it is not very great. For example, election results for November 1994 are widely available on the Internet whereas the results of the 1952 elections are less likely to be available.
- Text-based information is easily disseminated via the Internet; however, information in numerical, graphical, audio, video, etc. form is more difficult to provide access to and thus is not available as often as text is. For example, movie reviews are easier to locate than video clips from movies.
- Some information is just plain not available, regardless of the means available to disseminate it, e.g., how many cameras did Robert Flaherty use in filming "Nanook of the North"?

This leads us to the phenomenon usually referred to as the Deep Web, and sometimes as the Hidden Web. Search engine indices are for the most part generated by spiders, crawlers or other similar (semi-)intelligent agents. Obviously, only static pages can be found, and - for the most part - only pages are found that are on major domain name servers, or that have been linked by other pages that have been found.

In a White Paper written by a member of the BrightPlanet group, the following findings were published<sup>11</sup>:

- Public information on the deep Web is currently 400 to 550 times larger than the commonly defined World Wide Web.
- The Deep Web contains 7,500 terabytes of information compared to nineteen terabytes of information in the surface Web.
- The Deep Web contains nearly 550 billion individual documents compared to the one billion of the surface Web.
- More than 200,000 Deep Web sites presently exist.
- Sixty of the largest Deep Web sites collectively contain about 750 terabytes of information - sufficient by themselves to exceed the size of the surface Web forty times.
- On average, Deep Web sites receive fifty per cent greater monthly traffic than surface sites and are more highly linked to than surface sites; however, the typical (median) Deep Web site is not well known to the Internet-searching public.
- The Deep Web is the largest growing category of new information on the Internet.
- Deep Web sites tend to be narrower, with deeper content, than conventional surface sites.
- Total quality content of the Deep Web is 1,000 to 2,000 times greater than that of the surface Web.
- Deep Web content is highly relevant to every information need, market, and domain.
- More than half of the Deep Web content resides in topic-specific databases.

---

<sup>11</sup> Bergman, Michael K.: "The Deep Web: Surfacing Hidden Value". In: The Journal of Electronic Publishing; Vol. 7, Issue 1; August 2001. URL: <http://www.press.umich.edu/jep/07-01/bergman.html> (accessed June 4, 2003)

- A full ninety-five per cent of the Deep Web is publicly accessible information - not subject to fees or subscriptions.

Exacerbating this effect is the relatively small amount of database overlap between the most popular search engines. Greg Notess' most recent overlap survey<sup>12</sup> showed fully 50% of relevant pages found by only one search engine (out of ten that were deployed in the survey). An additional 21% were found by only two search engines. This despite the fact that Notess also found substantial database growth in the two years prior to his latest survey update (March 6, 2002) Incidentally, this is quite suggestive as an indirect indicator of the extent of total Internet page growth, since if the ten most common search engines have all experienced substantial database growth and if at the same time database overlap is stagnant or even decreasing, then plainly the actual number of potentially findable pages on the surface Web must have increased by an even larger proportion.

Given that the surface Web is dramatically smaller than the Deep Web and that at the same time the average search engine holds only a very small part of the surface Web in its database, then plainly even a perfectly formulated search argument is going to miss almost all actually available and relevant documents if a user's retrieval strategy is limited to the use of a single search engine.

Though it may seem paradoxical at first blush, the phenomenon of missing so much relevant information when one knows perfectly well that it is actually there is every bit as stressful as the more common information overload effect of receiving an un-sought-after information glut.

It is therefore quite important to understand how search engines work in general and where their databases come from. Since different search engines also employ different tools to optimize retrieval results or support some types of retrieval strategy more than others, it is also useful to know how particular search engines work. Despite the fairly prevalent phenomenon of the "favorite" search engine, it is simply not the case that any one search engine is the best for all or even most kinds of searches. User empowerment is greatly augmented by knowing the strengths and weaknesses of several search engines - including meta-search engines - and being able to deploy them in an effective manner.

What do search engine returns tell us about precision and recall? What alternate retrieval strategies are suggested by the returns on an initial search argument?

Precision vs. recall is the main paradigm for evaluating quality in information retrieval. Ideally, of course, we would like to have 100% precision and 100% recall on every query. The dilemma caused by information overload is that on many queries, even a fairly low recall value will almost inevitably lead to an effective precision value close to zero. Unless it is among the first 10 results displayed, a single pertinent document representing  $R = 100\%$  will effectively vanish if it is mixed in among 1000 hits, the remaining 999 of which are false positives. Even retrieval strategies which deliberately accept dramatically reduced recall percentages regularly lead to ballast amounts which hardly permit a competent evaluation of relevancy.

---

<sup>12</sup> G. Notess, *ibid.* URL: <http://www.searchengineshowdown.com/stats/overlap.shtml> (accessed June 16, 2003)

Users need to enhance their skills in evaluating retrieval results. Over time, librarians and other information professionals have developed a set of criteria<sup>13</sup> which can be used to evaluate the usefulness of information resources. Here are some helpful evaluation criteria:

- **Format** - Information resources on the Internet are generally structured in one of three ways: file archives, gopher files, or hypertext documents which correspond to the Internet tools of ftp, gopher and World Wide Web. The process for determining the availability of a particular information resource varies based on which type of resource is desired. Also, in some situations it will not be possible to access a resource in a particular format. For example, some Internet providers do not provide browsers or viewers for the World Wide Web.
- **Scope** - The scope of a particular information resource is a measure of the intended coverage of the source, the actual coverage of the topic it provides and the currency of the information it contains. Often times, the coverage of a topic by an information resource is greatly influenced by the audience for which the resource was created. Internet documents titled "readme" or "about this..." often provide information about the scope of a particular resource; however, it is often only possible to guess at the scope by browsing around a given information resource.
- **Relation to Other Works** - As might be expected of a large electronic network with no centralized organizing or regulating body, the Internet contains many resources that overlap with one another. An individual collecting the addresses of Internet resources is likely to acquire a large file which will quickly become unmanageable. Recognizing the relationships between particular resources and discarding duplicative addresses is one method of electronic information management. In addition, many Internet resources have print counterparts which may or may not contain more information than the Internet information resource. An awareness of this fact will help in locating the most complete and current information available on a given topic.
- **Authority** - Knowing the educational and/or occupational background of the creator or compiler of an information resource can help in determining the reliability and accuracy of the resource and the information it contains. Among other things, personal home pages on the World Wide Web, campus directory entries and information retrieved through finger may reveal useful and relevant information about an information provider.
- **Treatment** - Determining the intended audience of a particular Internet resource will reveal the intended treatment of the information contained in the work. Two types of distinctions can be particularly useful: scholarly vs. general public and expert vs. novice (student) Examining the objectivity of the resource will also help determine the accuracy and reliability of the information provided.
- **Arrangement** - The Internet has no overriding organizational scheme or structure. Many resources are arranged alphabetically, just like many print sources; however, many other organizational structures exist including academic department, corporate structure, type of resource, and subject categories. How well a resource is arranged will impact how easy it is to use.
- **Cost** - While the Internet is often touted as a free resource, there are many hidden costs, and others that are not so hidden. First, the Internet was developed and continues to be maintained, in many cases, through U.S. government funding. Accessing the Internet

---

<sup>13</sup> Smith, Linda C. (1991) "Selection and Evaluation of Reference Sources" in Richard E. Bopp & Linda C. Smith (eds.), *Reference and Information Services: An Introduction*. Englewood, CO: Libraries Unlimited, p.240

requires, minimally, a computer, modem and an Internet access provider. In some areas, freenets provide access without charge to members of the local community; however, freenets are not widely existent. Many Internet users have access to the Internet through an academic institution or other organization. If desired, individuals can purchase Internet access through commercial Internet providers. The cost of doing so can vary greatly. And, though the hype that surrounds the Internet may not reflect this, not all of the information resources on the Internet are free. There are numerous fee-based databases and other services for which users must pay if they want to use them. Finally, many people discover that the Internet is costly in non-monetary ways as well. Finding and/or providing information on the Internet is sometimes a frustrating and time consuming experience, requiring much patience and energy. For some people, this is not a problem; for others, it is better to be less involved with the Internet.

Though these criteria can be used most easily to evaluate printed information and resources, they can also be applied to resources and information found on the Internet in order to determine the accuracy and usability of a particular Internet resource.

Based on the initial returns, users need to be able to revamp their queries so as to bring the returns more into line with their expectations and needs. Using Boolean operators to break search topics down into logical concepts, using more specific or unique terms to increase the relevancy of returned documents, using synonyms to expand the search, recruiting additional search terms from returned and pertinent documents and adapting one's search strategy to the specific question or using different databases or search engines according to the nature of a question are all useful techniques to reduce ballast and boost relevant returns.

Admittedly, it takes time to acquire the skills needed to apply these techniques successfully, but it also takes time to sift through hundreds, and maybe thousands, of false positive hits.

The simple fact of the matter is: Nothing about the Internet as it is is going to change significantly in the foreseeable future. It will continue to expand at exponential rates, pages will go on being structured to suit the tastes and needs of the page designers, metatags will remain unused or be applied inconsistently and idiosyncratically. Page recruitment for search engine databases will be a hit-or-miss affair with little database overlap between engines, while the greatest part of the Web will be unavailable to spiders and crawlers by virtue of being dynamically generated or because they require a specific log-in to be accessed. Broken links and Error 404 will still be everyday occurrences. Further technological aids to address the information overload problem will, of course, continue to be developed, but these will still have only palliative or remedial effects. Information overload will go on being perceived by users as a problem, it will continue to induce feelings of stress and it will quite certainly remain a problem in terms of retrieval precision and recall.

Ultimately, the greatest utility of the Internet may well prove to be as a sort of gigantic Yellow Pages (i.e. classified directory) for experts. If users are able to identify the persons having the information or the expertise they seek, then it is usually possible to query them via e-mail or offline (by telephone, fax or even by letter).

#### 4. Enabling User Empowerment through Information Literacy

Information literacy - as opposed to mere computer literacy - can provide a major contribution to the enablement of user empowerment in the area of enhancing retrieval skills.

What does information literacy consist of and how can it be achieved?

According to the American Library Association (ALA), information literacy is a set of abilities requiring individuals to "recognize when information is needed and have the ability to locate, evaluate, and use effectively the needed information."<sup>14</sup> The ALA has developed a set of Information Literacy Standards<sup>15</sup> which state that an information literate individual is able to:

- Determine the extent of information needed
- Access the needed information effectively and efficiently
- Evaluate information and its sources critically
- Incorporate selected information into one's knowledge base
- Use information effectively to accomplish a specific purpose
- Understand the economic, legal, and social issues surrounding the use of information, and access and use information ethically and legally

Information literacy forms the basis for lifelong learning. It is common to all disciplines, to all learning environments, and to all levels of education. It enables learners to master content and extend their investigations, become more self-directed, and assume greater control over their own learning. Information literacy is related to information technology skills, but has broader implications for the individual, the educational system, and for society. Information technology skills enable an individual to use computers, software applications, databases, and other technologies to achieve a wide variety of academic, work-related, and personal goals. Information literate individuals necessarily develop some technology skills.

Teaching for information literacy represents a substantial challenge for educators, particularly at the secondary and tertiary levels. A brief survey of the current discussion about pedagogical strategies revealed that the following new methodological approaches are generally seen as being especially helpful in the facilitation of information literacy and information competence:

- Inquiry Learning - Students immerse themselves in the topic, context, or situation they are studying. They investigate the location, historical background, current situation and problems. They become mini-experts on the topic (Knowledge Attack) before beginning the inquiry process (Inquiry Learning Model). In this inquiry process students form a question that becomes the focus of their investigation. They form subsidiary questions, form hypotheses, plan and carry out their research, come to some conclusions and decide how they could make change happen.

---

<sup>14</sup> American Library Association. "Presidential Committee on Information Literacy. Final Report". (Chicago: American Library Association, 1989.)

<sup>15</sup> ALA Information Literacy Standards. URL:  
[http://www.ala.org/Content/NavigationMenu/ACRL/Standards\\_and\\_Guidelines/Information\\_Literacy\\_Competency\\_Standards\\_for\\_Higher\\_Education.htm#stan](http://www.ala.org/Content/NavigationMenu/ACRL/Standards_and_Guidelines/Information_Literacy_Competency_Standards_for_Higher_Education.htm#stan) (accessed June 23, 2003)

- Problem-based Learning - This has several distinct characteristics which may be identified and utilized in curriculum design: Reliance on problems to drive the curriculum - the problems do not test skills, they assist in development of the skills themselves; the problems are truly ill-structured - there is not meant to be one solution, and as new information is gathered in a reiterative process, perception of the problem, and thus the solution, changes; students solve the problems - teachers are coaches and facilitators; students are only given guidelines for how to approach problems - there is no one formula for student approaches to the problem; authentic, performance-based assessment is a seamless part of the instruction.
- Project-based Learning - This is an instructional approach that contextualizes learning by presenting learners with problems to solve or products to develop. For example, learners may research adult education resources in their community and create a handbook to share with other language learners in their program, or they might interview local employers and then create a bar graph mapping the employers' responses to questions about qualities they look for in employees.
- Service Learning - This approach attempts to combine learning practice, e.g. learning-by-doing, with volunteerism and community service, with a view to thereby enhancing student commitment and thus augmenting the learning results.

## 5. Conclusion

Plainly the problem of information overload is multi-causal in origin; it therefore follows that any solution must be polyvalent. In spite of increasing availability of curricular and pedagogical resources for the teaching of information literacy, it seems clear that user empowerment in terms of Internet retrieval will largely come from increased self-reliance and self-teaching of retrieval-relevant skills. Excessive emphasis on technology-based solutions should definitely be avoided.