Searching information in the electronic environment

A Compendium prepared

by

Alfred S. Sife, Frankwell W. Dulle and Grace E. P. Msoffe

Sokoine University of Agriculture

2010

Searching information in the electronic environment

A Compendium prepared

By

Alfred S. Sife, Frankwell W. Dulle and Grace E. P. Msoffe

Sokoine University of Agriculture

2010

PREFACE

Developments in Information and Communication Technologies (ICTs) have greatly changed the way information is being created, stored, distributed and accessed. Today, information searchers are faced with a new information environment - one characterized by proliferation of technology-driven information sources, media and formats. Recognizing such changes, this compendium has been prepared mainly to support academic and research activities at Sokoine University of Agriculture (SUA). The objective of the compendium is to impart information literacy skills to students, staff and other information users for their academic and professional careers and for lifelong learning. Although information literacy module is formally taught under CIT 200 (Computer applications) at SUA, such skills are necessary to all disciplines, to all levels and to both staff and students. The compendium aims at providing skills for retrieving information quickly and efficiently; evaluating the quality of information; understanding information-related ethical issues; and citing information sources appropriately. Simple to follow examples are presented throughout the compendium.

The contents of this compendium are based on the authors' experiences in providing library services, teaching information literacy courses, research in library and information science as well as on information gathered from various print and electronic resources. The authors believe that taking time to work through this material will greatly enhance one's information search skills. The authors acknowledge their indebtedness to various individuals and sources that have contributed towards the preparation of this compendium.

Chapter one introduces general techniques for searching information. Chapter two deals with online and offline databases such as library catalogues and e-journal databases. The use of web-based subject directories and search engines is explained in chapter three. Chapter four presents the concept of open access as a new mode of scholarly communication and the final chapter introduces quality control issues and citation methods.

A.S. Sife, F.W. Dulle and G.E.P. Msoffe
April 2010

Abbreviations

AGORA   Access to Global Online Research in Agriculture

CD-ROM   Compact Disc Read-Only-Memory

EMERALD   Electronic Management Research Library Database

HTML   Hyper Text Markup Language

ICTs   Information and Communication Technologies

OPAC   Online Public Access Catalogues

PDF   Portable Document Format

SNAL   Sokoine National Agricultural Library

SUA   Sokoine University of Agriculture

TEEAL   The Essential Electronic Agricultural Library

Table of contents

List of Figures

GENERAL INFORMATION SEARCH TECHNIQUES

1.1   Introduction

Information is a fundamentally important factor for success of many human activities. In the current information society, people from all walks of life require accurate, reliable and timely information to carry out various activities. Virtually every task, simple and complex ones, requires information of some kind. While some of the information can be readily available, other information requires extensive searching and consultation of multiple sources. In an academic environment, information is particularly important input for learning, teaching, research, and outreach activities. This implies that those individuals who have timely access to relevant information stand a higher chance of academic success.

Developments in Information and Communication Technologies (ICTs) have greatly changed the whole information environment. ICTs have led to rapid increase in the volume of information (information explosion); information is made available through multiple media and formats (print, electronic, graphical, audio, and textual); information is managed and stored in a variety of resources - CD-ROM[1]s, OPAC[2]s, digital libraries, and web-based databases; and that scholarly communication and publishing are increasingly taking place in electronic forms. While libraries increasingly get licensed access to electronic databases, at the same time the amount of information freely available on the Internet continues to increase. Yet, access to print information resources in libraries continues to be essential. Thus, the current problem is no longer one of not having enough information; it is just the opposite - too much information, in various formats, and of uneven quality.

Searching information in the electronic environment is, in most cases, time consuming and frustrating for those lacking the required skills. This is due to the enormous amount of information made available on the web; different ways in which information is stored and retrieved; diversity of search tools available; overall lack of organization of the web content; uneven quality of the web content; and the constant evolution of electronic sources of information. For that reason, information users require conceptual knowledge to convert

---

[1] Compact Disc Read-Only-Memory
[2] Online Public Access Catalogues

information needs into searchable queries, semantic knowledge to construct queries for a given system, as well as technical knowledge to enter queries as specific search statements. Consequently, information literacy skills become very important in such a complex information environment. The hit-or-miss process that worked for some users to obtain needed information from the libraries in the past is not effective today.

Broadly defined, information literacy entails a range of skills required for locating, retrieving, evaluating, and using information from various sources in an effective and efficient way. An information literate person is able to determine the needed information; locate and access needed information effectively and efficiently; evaluate information and its sources; use information to accomplish a specific purpose; and understand socio-economic and ethical issues surrounding information use. Without these skills, information searchers tend to spend many hours retrieving irrelevant documents that do not match their needs. While computer literacy is a pre-requisite for information literacy, a person who knows all computer hardware and software, and every keystroke may not be information literate. Information literacy is more than computer literacy as it focuses on the efficient use of information resources. The American Library Association (1989:1) describes information literate people as:

> *Those who have learned how to learn. They know how to learn because they know how knowledge is organized, how to find information and how to use information so that others can learn from them. They are people prepared for lifelong learning because they can always find the information needed for any task or decision at hand.*

1.2   Information search techniques

Information search is a process that aims at gathering recorded information from various sources for a defined purpose. For the searchers to retrieve information, their needs must be expressed in the form of queries that are interpretable by the information system. However, the ability of search tools to understand what a person wants is often very limited. Many search tools tend to look for occurrences of keywords, but do not understand what the keywords mean or why they are important to the searcher. In other words, search tools have not yet reached the point where humans and information systems understand each other well enough to communicate clearly. As a result, most queries tend to be quite brief and may not satisfy the normal rules of language syntax. From the information search viewpoint, the major task is to establish the relationship between people's information needs and documents

that meet those needs. A successful information search process can be summarized into the following steps:

- Search formulation (subject definition)
- Identifying appropriate sources of information
- Applying the right search tools and techniques
- Search evaluation (This is explained in chapter five)

1.2.1 Search formulation (Subject definition)

Obtaining information from electronic databases depends to a great extent on how precise the search queries are. Generally, poor queries return poor results while good queries return great results. This suggests that before one encounters the search exercise, it is important to thoroughly define the subject matter for a given topic. Thus, *search formulation* is the analysis of a topic to recognize key concepts and words that describe a topic and how they relate with each other. Search formulation results into appropriate terms and keywords to be used in the search process. The process also helps to understand fully what one wants to search and clarify unfamiliar terms.

During search formulation, the following linguistic diversities should be considered:

- Synonyms – different words that have similar meanings e.g. swine/pig.
- Variant spellings such as those in British and American spelling e.g. behaviour/behavior.
- Distinctive words or phrases e.g. cardiovascular system; food security.
- Plurals and other endings e.g. woman/women.
- Acronyms e.g. CaMV = Cauliflower mosaic virus.
- Broader, narrower and related terms e.g. livestock - cattle/ sheep/ donkey.
- Scientific and common names e.g. Pawpaw/*Carica papaya*
- Homonyms - words that share the same spelling and pronunciation but have different meanings e.g. mercury (metal)/mercury (planet).
- Compound words – words that may be split into two e.g. wastewater/ waste-water.
- Variations of a root word e.g. feminism, feminist, feminine.

3

One way to analyze a topic is by using a *concept mapping* technique. This is simply a way to display concepts and their relationships. The technique helps to organize ideas and define a topic. Fig. 1 illustrates how a concept map on a topic - *sustainable development* - is developed:

- First, write down the main idea in the center. This will be a starting point for the concept map. In this case, "sustainable development" is the central theme.
- Next, think of some issues related to the central theme. Link these issues around the central theme, and draw lines connecting them. In this example, there are three major categories of "sustainable development":
    - Economic development
    - Environmental protection
    - Social development
- For each issue, think of sub-concepts that relate to it and connect them to the appropriate issue. Each of the three issues has sub-ideas for possible research exploration.
- You can continue to brainstorm possible ideas. For example "*resource use*" can be divided into "*renewable resources*" and "*nonrenewable resources*".
- Not necessary that all concepts are included in the search, but this approach enables one to see relationships among topics and concepts.

Relevant search terms can also be obtained from:
- Printed and online dictionaries, encyclopedias and thesauri.
- Keywords and descriptors used in key journal articles.
- Subject headings and lists of controlled vocabulary in individual databases.

As a general guideline, the broader the term, the greater are the chances of getting information, both relevant and irrelevant. Narrower terms usually give fewer results but most relevant. It is therefore recommended to start searching with specific terms and if not getting enough information, then go for broader terms.

1.2.2 Identifying appropriate sources of information

After obtaining appropriate subject headings, terms and keywords, one should identify the right information source(s) to use for searching. It is important to begin a search with the most relevant source for the subject area and search it thoroughly before moving onto another

source. In other words, it is important to consult specific information sources suited to particular types of information. A comprehensive search requires one to first, scan secondary sources such as library catalogues and bibliographic materials. This implies that information searchers should be familiar with a range of information sources in order to pick the right starting place.

Fig. 1: A concept map on "sustainable development"
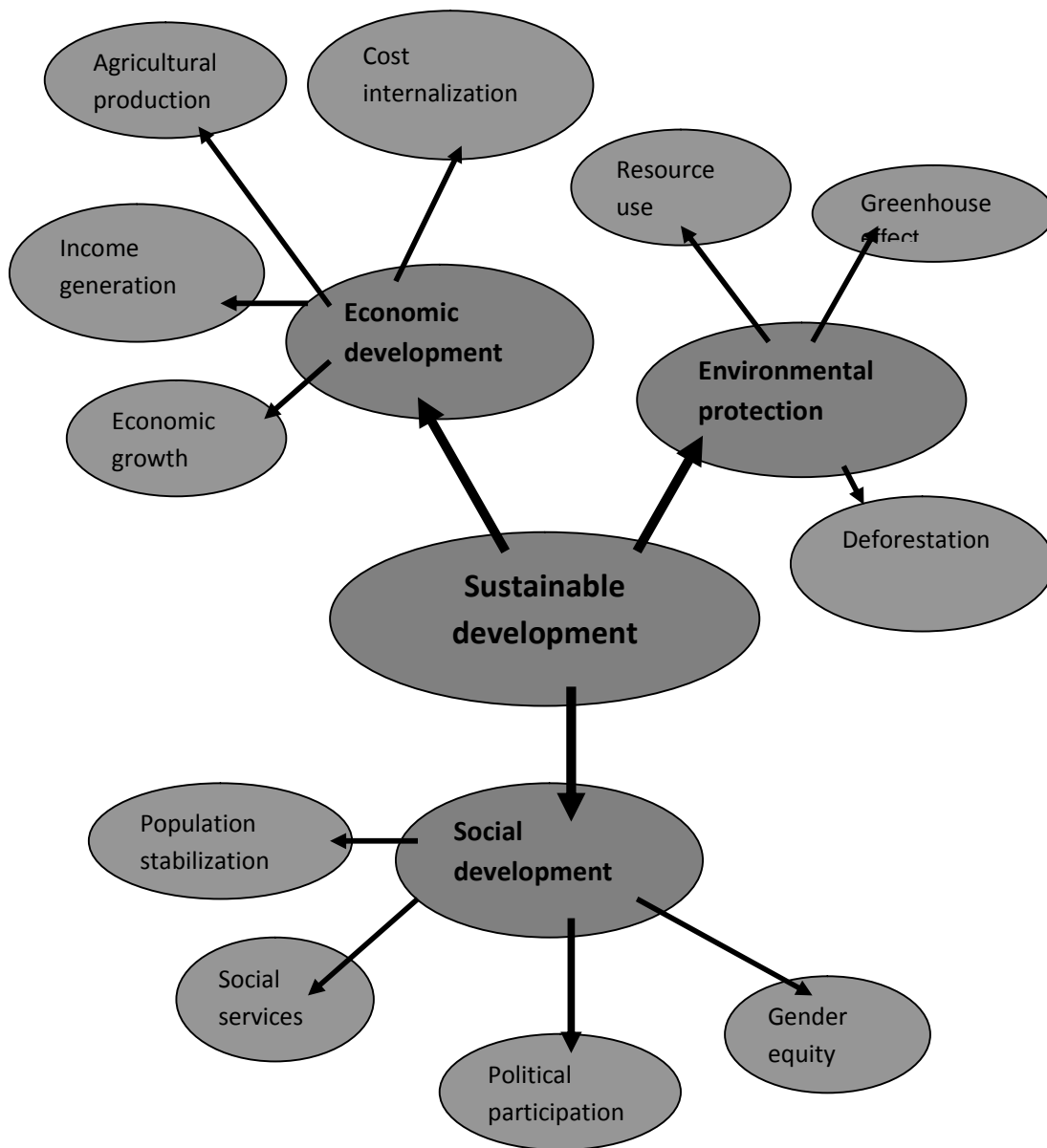
1.2.3 Applying the right search tools and techniques

The subject headings, terms and keywords developed in the first step will be employed for actual searching using appropriate search tools. In electronic databases, actual searching can be executed by entering keywords in search boxes or through browsing the organized topics. In most electronic databases, queries can be searched either as:

- **Phrase query**,
- In a combined form using **Boolean operators,**
- By **Truncation (Wildcards)**.

1.2.3.1 Phrase searching

Phrase searching is the use of quotation marks (" ") to locate documents containing a distinctive string of words associated with a topic. Phrases are constituent terms that are naturally married such as *sustainable development* and *poverty reduction strategies*. When using phrases, spaces between words are as important as any other character. For example, if a double space is included between any two words that make a phrase, the search will not produce the same results as those where there is a single space. There is variability in the way databases treat phrases. Some databases provide options for phrases, others do not, but almost all allow one to enter a phrase in quotes, ignoring the quotations if not supported. In most databases, if a phrase is not specified in a search statement, the default search is one in which any of the words may be present. This can lead to thousands of useless hits.

1.2.3.2 Boolean Searching

Information searching in most electronic databases is based on the principles of Boolean logic. This is the logical relationship among search terms, and is named after the British mathematician and logician George Boole. Boolean logic is used to construct search statements in which search terms are joined together by using logical **operators** in specified syntax. Boolean operators improve search results by broadening, narrowing or excluding search results. Typically, the Boolean operators are **OR, AND, NOT (AND NOT).** The underlying principle of Boolean logic is the set theory. For example, the **AND** operator is equivalent to the **intersection** set operation while the **OR** operator is equivalent to the **union** set operation.

Sometimes symbols are used to represent Boolean logical operators. In some databases, plus sign (+) is similar to AND operator and minus sign (-) is similar to NOT operator. Absence of a symbol is also significant, as the space between keywords is, in many cases, defaults to AND operator. Some databases offer a search template which allows the user to choose the Boolean operator from a menu. Usually, the logical operator is expressed with substitute language rather than with the operator itself.

AND: This operator retrieves records in which **ALL** of the search terms are present. In other words, search terms on both sides of this operator must be present somewhere in the document in order to be scored as a result. The AND operator tends to narrow a search and make it focused. In Fig. 2, records that contain the words *environment* and *development* will be retrieved. The more concepts combined in a search with AND logic, the fewer records will be retrieved as in the case of *environment* AND *development* AND *poverty*.

Fig. 2: Boolean operator AND
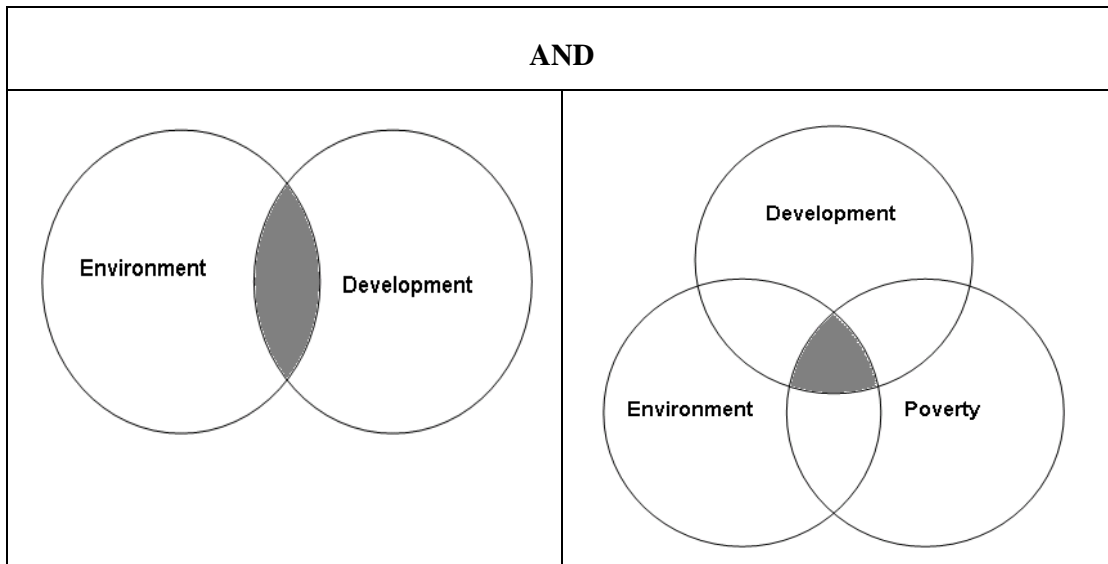
The **AND** operator only requires that the terms immediately on both sides must all appear in the document. It neither specifies the location of the terms in the document with respect to one another nor the logical sequence of the search terms. This operator can be used to attach several required terms together, all of which must be present for the outcome to be successful. For example, the query *environment* **AND** *development* **AND** *poverty* would only return documents that contained all the three terms.

OR: This operator is equivalent to the UNION operator in set theory. When used, the logical operator **OR** retrieves documents that contain **EITHER** of the search terms and usually broadens search results. It is used when looking for synonyms**,** spelling variations, or foreign spellings. It should be noted that the **OR** operator is not strictly equivalent to a sum but do approximate the total of the two separate queries. Documents which contain both terms still get counted as a single document. For example, in Figure 3, records in which one of the search terms "*atmospheric pollution*" OR "*air pollution*" is present will be retrieved. The more terms combined in a search with OR operator, the more records will be retrieved as in the case of *atmospheric pollution* OR *air pollution* OR "*smog*". Unless used in a controlled way to expand results, most often as part of a parenthetical argument dealing with synonyms, it is not recommended to use the **OR** operator.

Fig. 3: Boolean operator OR
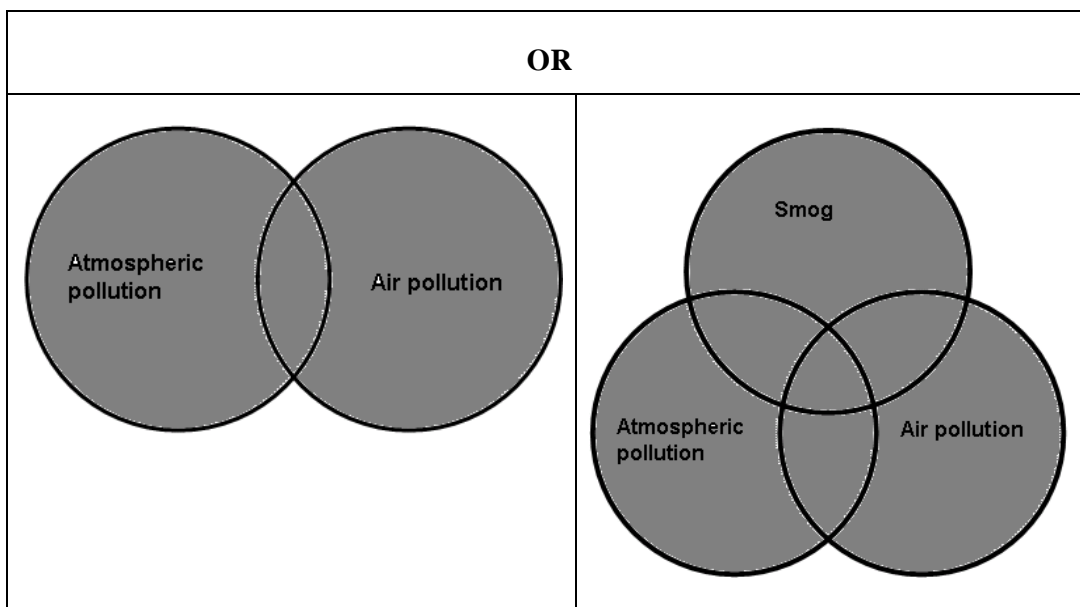
NOT: This operator excludes documents that contain the keywords a searcher wouldn't like to appear in the document. Documents containing the term after this operator (NOT) are rejected from the results. It is used when lots of search results with unwanted terms are anticipated. **NOT** is a unary operator meaning that it only works on the term or phrase that immediately follows the operator. It does not evaluate terms or phrases on both sides of the

operator. In Fig. 4, records that contain the word *pollution* but not *"greenhouse effect"* will be retrieved. In some databases **AND NOT** is used in place of NOT.

Fig. 4: Boolean operator NOT
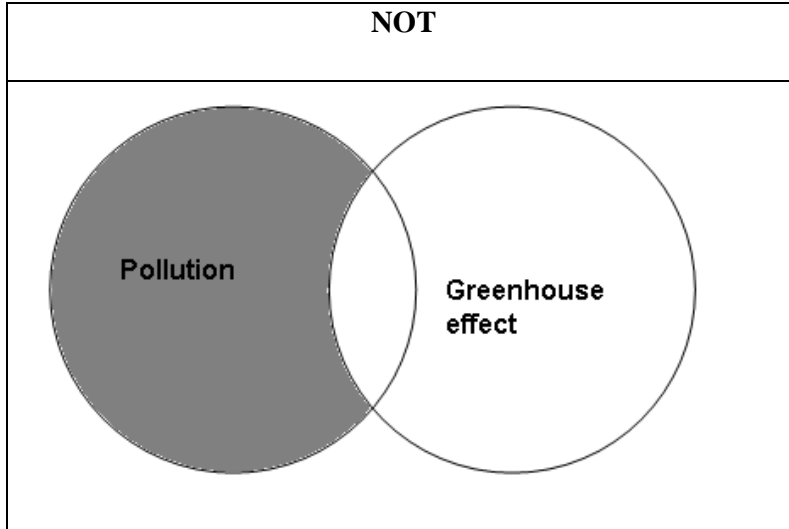
NEAR: The **AND** operator requires that the terms on both sides can appear anywhere in the document in order to get a successful result. The **NEAR** operator requires the two terms to be within a specified word count of one another to be counted as a successful result. Generally, most search engines that support the **NEAR** operator have a set value of ten words maximum distance. Some engines also use **ADJ** (for adjacent) as the equivalent operator to **NEA**R. The **NEAR** operator does not care which of the phrases or terms on either side of the argument comes first or not, just that the two phrases or terms are within the specified distance. The **NEAR** operator is important to ensure that search terms occur within the same sentence or same paragraph. It can remove large Web sites that contain a lot of terms.

BEFORE and **AFTER:** These operators work in the exact same manner as the **NEAR** operator, only that one can specify which terms need to come first or second. In the case of the **BEFORE** operator, the first term must occur before the second term within the specified word distance. In the case of the **AFTER** operator, the first term must occur after the second term within the specified word distance. These operators do provide greater control to searches.

Nested searching

Many databases allow creation of complex queries using Boolean operators, phrases and parentheses. This is called **nested searching**. Each bounded argument set off by parentheses is called a **Boolean expression**. Search services that support structured syntax do not always read from left to right. Instead, they read "inside-out", in order of precedence for the logical connectives. The combination within the parentheses is evaluated before being combined with the terms outside the parentheses. For example, in the Boolean expression: *environment* AND (*atmospheric pollution* OR *air pollution*), the phrases in the parentheses will be evaluated first.

1.2.3.3 Truncation

Truncation or **wildcard** searching is a useful way of retrieving all words having a particular word root without having to type them all separately. Truncation symbols differ depending on database, but mainly an asterisk (*) is used. Truncation can be performed in different ways:

- Right hand truncation - When * is placed after a word or part of a word, it locates documents containing words that start with the combination of letters. For example **agricultur*** will retrieve documents that contain the terms **agriculture**, **agricultural**, **agriculturalist,** etc.

- Left hand truncation is useful for searching complex (organic) compounds if a searcher does not know the root of the compound but knows the ending (functional group) e.g. ***ethylene**.

- Middle truncation. For example **wom*n** will retrieve documents that contain the terms **women** and **woman.**

- Some databases support truncation in phrases. For example, searching the query "*poverty * strategies*" will return documents that contain the phrases *poverty reduction strategies* and *poverty alleviation strategies.*

When using truncation, care should be taken to make word roots neither too short nor too common so that irrelevant records are not retrieved (e.g. **cat*** retrieves **cat, cats, catastrophe, catalogue, caterpillar** etc.). Fortunately, most web-based search engines have increasingly been automatically searching all words having a particular word root. Thus, there is no need to truncate words in such engines.

1.3 Common words

Search services ignore common words and characters known as "stop words." These include **a, about, an, and, are, as, at, be, by, from, how, i, in, is, it, of, on, or, that, the, this, to, we, what, when, where, which, with**. Though such words are essential connectors in language, search facilities ignore them because they are commonly found. Search services include these common words in their "stop-lists" so that whenever introduced in a query they are ignored. This means that search engines do not respond the same way humans do. So, for example if one types: *what is the meaning of biodiversity?* (without the quotation marks) in Google, the search engine will drop the words; **wha**t, **is**, **the** and **of**.

1.4 Relevance feedback

The process of searching information is often uncertain. It frequently happens that the initial query posed does not provide satisfactory results. There are several reasons for this case:

- A searcher may have only vague ideas of what information is needed (may not be very clear of what information he/she is searching for);
- A searcher may not be able to express his/her conceptual ideas into suitable queries;
- A searcher may not know the organization and content of the system or may not have good ideas of what information is available for retrieval;
- The vocabulary posed by a searcher may be quite different from that of a system;
- A searcher may pose a query that is broader or narrower than what is needed thus returning either far too many documents or not enough.

The idea of *relevance feedback* entails the use of retrieved records that seem relevant to obtain more records on a particular topic. It is a modification of the search process to improve accuracy by incorporating information obtained from prior relevance judgments. Relevance feedback exploits the notion that although it may be difficult to formulate a good query, it is easy to judge the retrieved documents. Seeing some retrieved documents may also lead information searchers to refine their understanding of the information they are seeking.

The basic procedure in the relevance feedback process as illustrated in Figure 5 is:

- The user issues a query.
- The system returns an initial set of retrieval results.
- The user examines some returned documents for relevance.

- The user adds words from known relevant document to the query (the user may get new search terms on the topic).
- The system displays a revised set of retrieval results.

Relevance feedback can go through one or more iteration of this sort. This means searchers can build their searches from what they have found. In the library, catalogues tend to offer alternative subject headings that one can search again to obtain other records on similar topics. One can also use these references to find other information. The relevance feedback concept can also be applied in the libraries as books tend to be shelved together in subject areas. Thus, if a useful book is found on the shelf, there will probably be others nearby.

Fig. 5: Relevance Feedback Example
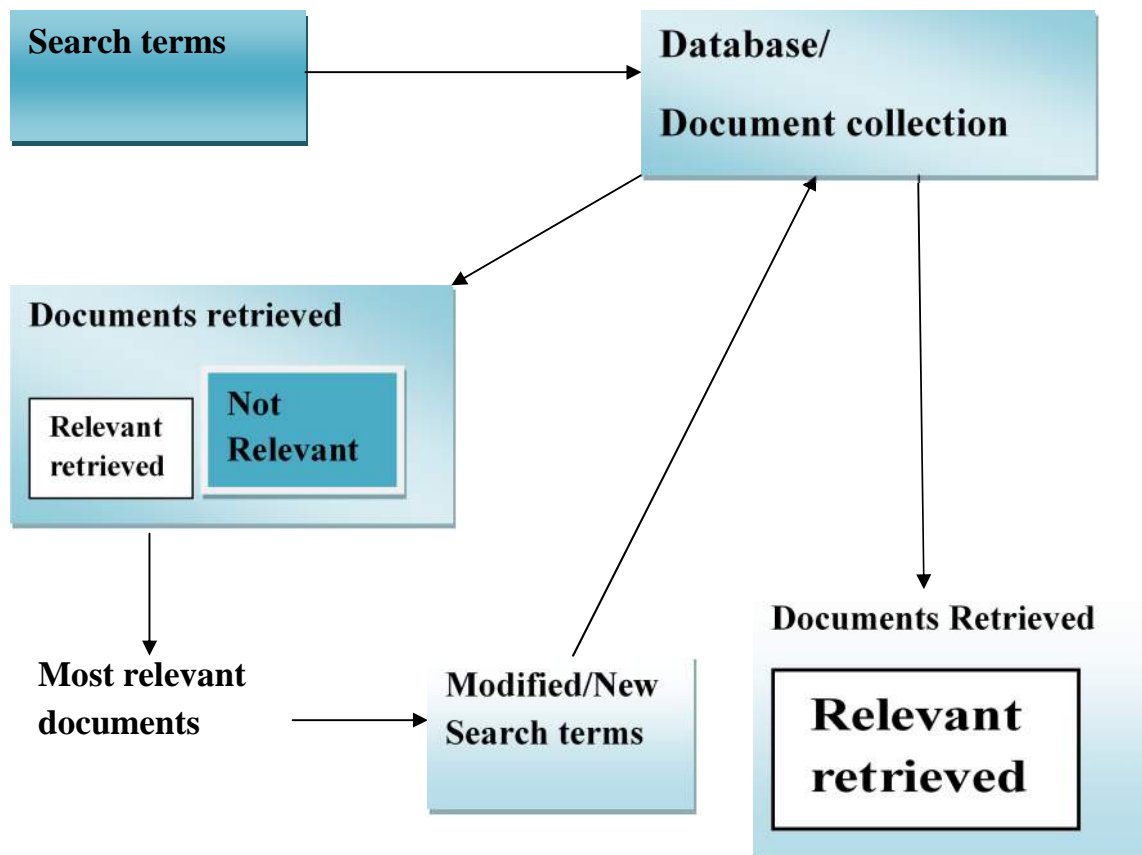
CHAPTER TWO

SEARCHING ELECTRONIC DATABASES

2.1 Overview

An electronic database is an organized collection of related records in a standardized format, searchable in a variety of ways, such as browsing or presenting search terms with the help of computer software. Electronic databases include but not limited to library catalogues, citation indexes, CD-ROMs, subject directories, e-journal databases as well as local databases attached to individual computers. These databases are mostly searched using computer software known as search engines. Following rapid advances in ICTs, search interfaces of electronic databases evolve rapidly as well. Changes to the search interface or the whole database may occur rapidly, though the search techniques remain the same. The following sections present examples of electronic databases and how to search information from them.

2.2 Library catalogues

Library catalogues are databases that contain organized bibliographic records describing the materials available or accessible in a particular library. The card catalogue which was being used in libraries for generations has been replaced by Online Public Access Catalogue (OPAC). OPAC is an online database of materials held by a particular library. To locate library materials such as books, searching can be carried out in two stages:

- Searching for bibliographic information,
- Identification of physical documents.

Searching the OPAC enables the user to obtain a classification number (**call number** or **class mark**), which is a unique identifier for the library material. Taking an example of the Sokoine National Agricultural Library (SNAL), catalogued library materials are arranged according to the **Library of Congress Classification System** (Table 1). This system uses a combination of letters and numbers to determine the location of documents on the shelves. The **CALL NUMBER** (e.g. **TA330.D3**) consists of one or two letters (TA in this case), indicating the main subject or classification, the number (330), is for the special aspect of the document, and the third letter (D) is the code of the author or title of the document.

Table 1: Library of Congress Classification System

Code	Description	Code	Description
A	General Works	H	Social Science
	General encyclopedia, reference books, dictionaries etc.	**HA**	Statistics
B	Philosophy-Religion	**HB-HJ** Economics, Business, Commerce, Management	
	B-BJ Philosophy, including BF, Psychology	**HM-HX** Sociology	
	BL-BX Religion	J	Political Science
C	Auxiliary Sciences of History	**JA-JC** Political science	
	CB History of civilization (general)	**JF-JQ** Constitutional history and public administration	
	CC Archeology	**JS** Local government	
	CS Genealogy	**JX** International law	
	CT Biography (general)	K	Law
D	History: General & Old World	**KFK** 1200 Kentucky law	
	D World history, including world wars	L	**Education**
	DA-DR Europe	M	Music
	DS **Asia**	N	**Fine Arts**
	DT Africa	**NA** Architecture	
E-F	History of America	**NB** Sculpture	
	E- America (general)	**NC** Graphic arts	
	E 151-857 United States	ND **Painting**	
	F Local U.S. history	**NK** Décorative arts	
	F 446-460 Kentucky	P	Language & Literature
	F 1001-1140 Canada	**P** Philology and linguistics	
	F 1201 Other countries, Latin America	**PA-PN** Language and literatures	
G	Geography, Anthropology, Folklore	**PE** English language dictionaries	
	G Geography (general)	**PR** English literature	
	GB-GC Physical Geography	**PS** American literature	
	GN Anthropology	**PZ** Juvenile literature	
	GR Folklore		
	GV Recreation		
Q	Science	T	Technology
	QA Mathematics	**TA-TH** General and civic engineering	
	QA76 Computer science	**TJ-TP** Mechanical, electrical, and chemical engineering	
	QB Astronomy	**TR** Photography	
	QC-QH Physical and earth sciences	**TS** Manufactures	
	QK-QR Life sciences	**TT** Handicrafts, arts and crafts	
R	Medicine	**TX** Home economics	
	RT Nursing	U	**Military Science**
S	Agriculture	V	Naval Science
		Z	Bibliography & Library Science

Source: http://www.lscc.cc.fl.us/library/guides/lcclass.htm

The current SNAL OPAC uses Web-based Library Integrated System (WEBLIS) software. The library's catalogue is accessible online at **www.suanet.ac.tz/lib**. The following are a few search tips for SNAL catalogue:

Simple search:

- Enter your search terms, separate each term with a semicolon (;). The semicolon is interpreted by the system depending on the choice of Boolean operator.
- You may use Boolean operators: + for OR, * for AND, ^ for NOT.
- If you do not know what to enter as a search term, press the "List" button. This will give you a list of keywords where you can select relevant terms.
- Click UNDO button to delete last input and CLEAR to empty the form (Fig. 6)

Advanced search:

- To use more than one field in the search process. Thus, a record can be searched using a call number, authors, keywords and other fields as shown in Fig. 7.

Thesaurus supported search:

This function provides information on the logic of the classification of the documents. Users can navigate through a thesaurus, view the connections between various terms, and use them appropriately. During navigation one can indicate specific words relevant to the topic.
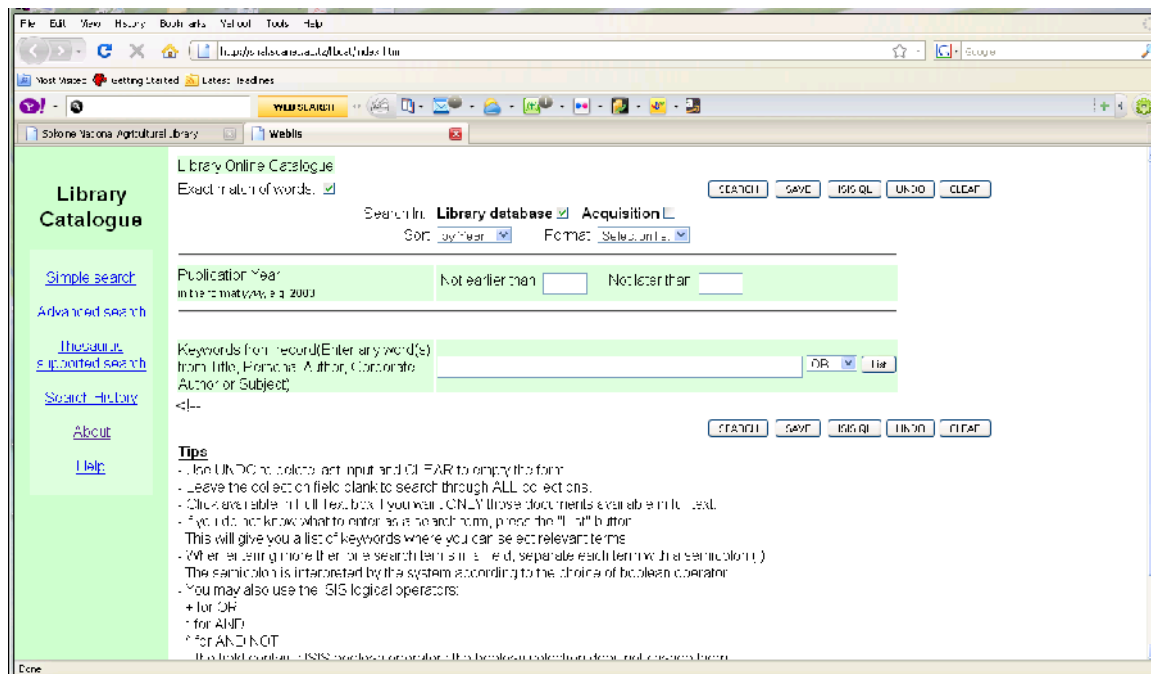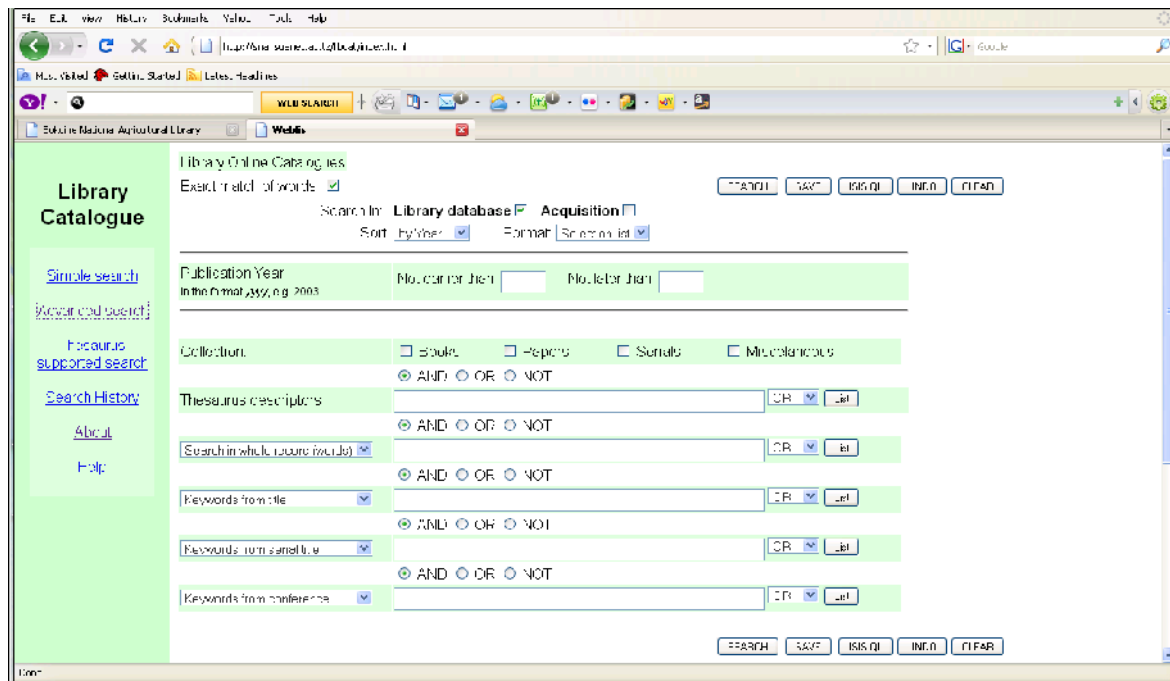
- Enter the appropriate root of the term in the box.
- Click on blue button(s) ■ to select the term(s) for searching.
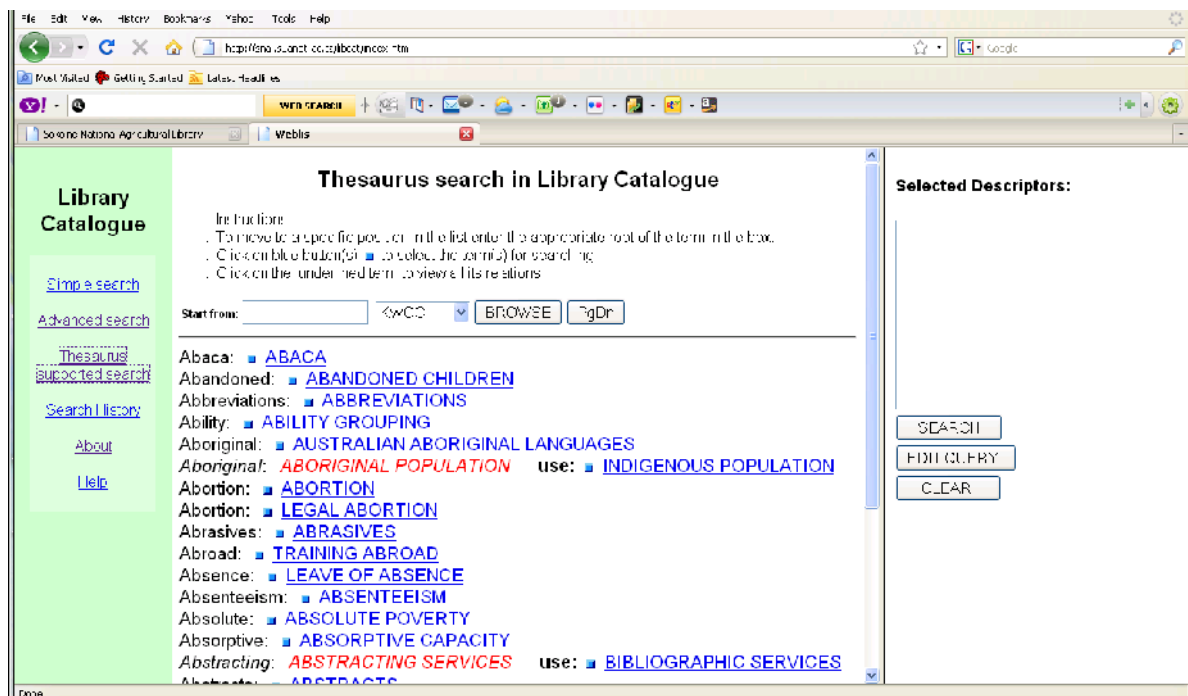- Click on the underlined <u>term</u> to view all its relations (Fig. 8)

Fig. 6: SNAL catalogue – simple search

Fig. 7: SNAL catalogue – advanced search

Fig.8: SNAL catalogue – thesaurus supported search

2.3   Electronic journals

A journal is a periodical that contains scholarly information on research whose intended audience includes scholars, practitioners and experts in the subject field. Electronic journals (e-journals) are scholarly journals accessible in a computerized format. Following developments in ICTs, a lot of journals are currently available electronically via the web, CD-ROMs or other digital databases. Some journals only exist in electronic format and others have both print and electronic versions. The medium distinguishes an electronic scholarly journal from its print counterparts but the process of developing content for both print and electronic peer-review scholarly journals is generally the same. The digital medium allows the publishing process to occur at a faster pace than in print by providing authors with information quickly to revise their work to meet editorial standards.

E-journals on the Internet may be freely available; available only after registration; or available only on subscription. These e-journals may be available at their own websites; through the websites of individual publishers or professional societies; through consolidated services that give access to titles from different publishers; or through a combination of these. The following sections describe three e-journal databases namely; Access to Global Online Research in Agriculture (AGORA), Electronic Management Research Library Database (EMERALD) and the Essential Electronic Agricultural Library (TEEAL).

AGORA

Access to Global Online Research in Agriculture (AGORA) is a programme that provides free or low cost access to major scientific journals in agriculture and related sciences to public institutions in developing countries. It provides access to several journals from the world's major academic publishers. AGORA is a partnership between the Food and Agriculture Organization (FAO), science and technology publishers, and other partners.

Opening AGORA
- Start web browser (e.g. Internet Explorer).
- Type: **www.aginternetwork.org** in the "Address" box

Logging In to AGORA
Abstracts and other free resources can be accessed without logging onto AGORA. To access full-text articles, one should follow these steps:

17

- On the AGORA main page, click on Journals.
- On the Journals page, click the Log In button. Click Yes or OK if any message about secure connections or security certificates appears.
- At the Log In page, type in the given user name and password.
- This will launch the user authentication page
- The AGORA main page will launch again. If it does not, click AGORA link.
- After the new AGORA main page launches, click the Journals link again.

Fig. 9: AGORA Home page
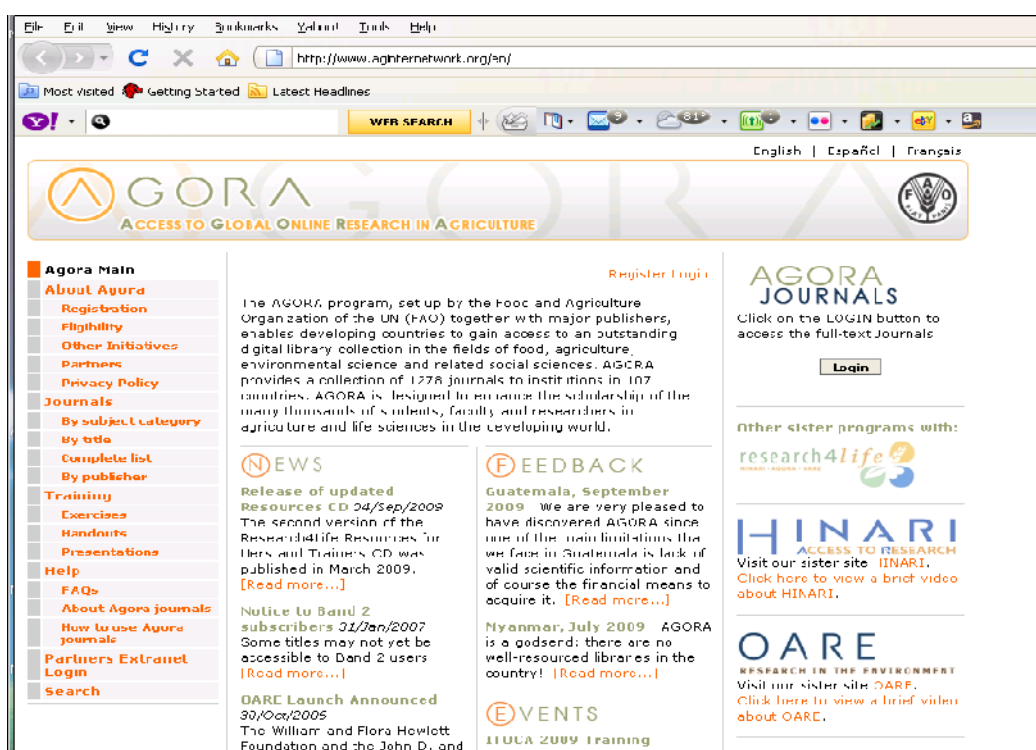
Finding Journals by Title

- Under Browse Journals by Title on the Journals page, choose the first letter of the journal title to see a list of all titles beginning with that letter or
- Choose "View list of journals" to see all available journals from A-Z or
- Enter exact or partial title in "Search Journal Title" box at top of Journals page.

Browse Journals by Publisher

- Choose a publisher from the Browse Journals by Publisher drop-down menu, which lists all journals in AGORA from that publisher.

Retrieving Full-Text Articles

After locating the specified journal using one of the methods described above,

- Click on the Journal Title link to go to the publisher's website.
- Choose the issue and volume required.
- Choose the article needed.
- Click on the link to the full-text file (it can be PDF or html).

Browsing Journals by Subject

- Choose a subject from "Browse Journals by Subject" category to see a list of journal titles covered in that subject area.
- Choose a journal and then browse or search for citations in that journal.

EMERALD

The Electronic Management Research Library Database (EMERALD) (**http://www.emeraldinsight.com/**) publishes journals in the fields of management, library services and engineering. Access to journals is mainly through fee-based subscription.

Browsing

- **To view a list of the journals**
  Select the Browse tab then select "Journal Title" from the Browse by options and click "Browse" to list the journals.
- **For a specific title of the journal that contains the needed article**
  Select the browse tab and check the "All Journals". Select the journal from the alphabetical links and then select the issue required. A list of the articles contained within that issue will appear.
- **Browsing by subject to find an article**
  Check the "Subject" button and browse. A list of subject headings will appear. Click the journals that pertain to a given subject and select the journal that suits the needs.

Quick Search

- Quick search is used to find articles on a given subject. **S**earch against All Fields (excluding full text) will bring back the most results. Searching All Fields (including full text), will retrieve more results.

Advanced Search

- Advanced Search is used for more restrictive searches. For example, if the author's name is known, but are not sure of the exact article title, one can enter the author's name in the "Author(s)" field and a couple of words from the title in the "Article Title" field.

Fig. 10: Emerald home page
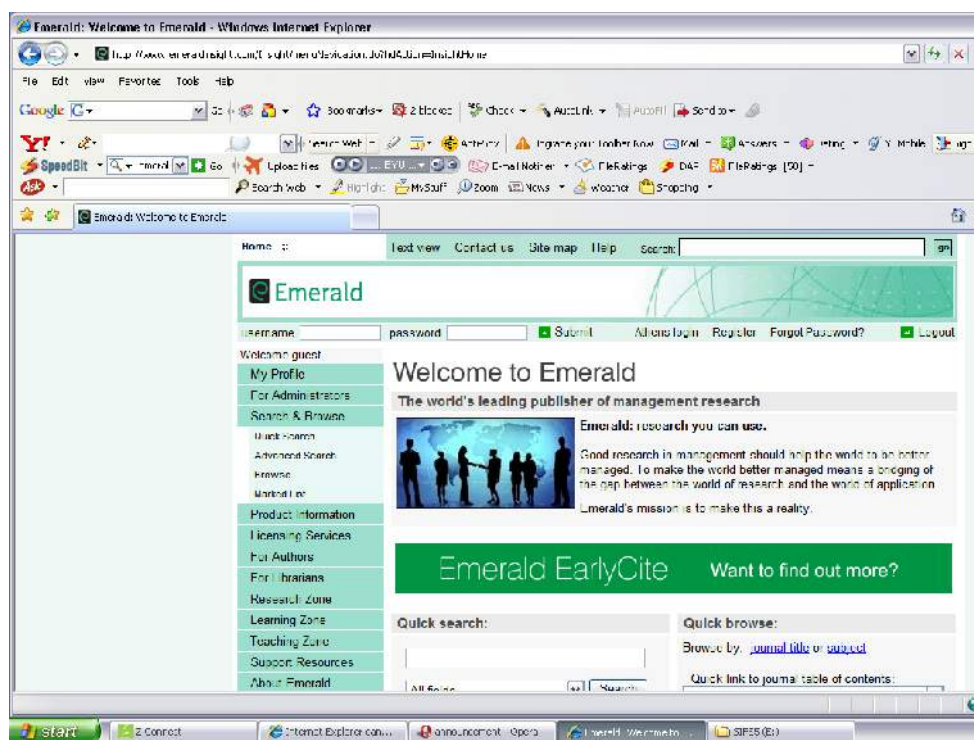
The Essential Electronic Agricultural Library (TEEAL)

The Essential Electronic Agricultural Library (TEEAL) is a collection of scientific journals in agricultural and related sciences. These include biotechnology, aquaculture, crop and soil science, natural resources, agricultural engineering, nutrition and food science, agroforestry, and veterinary medicine. TEEAL is a project of Cornell University's A. R. Mann Library with the cooperation of over 60 publishers and societies, and with the support of the Rockefeller Foundation and other international funding agencies. The database is on small

hard drive that can be connected to one or many networked computers through the local area network (LAN).

Searching TEEAL

To begin, double-click the TEEAL icon on the desktop, or go to the Start Menu then Programs to select TEEAL. One can either search or browse the journal collection.

Quick Search

Use the Quick Search box on the home page to quickly find search terms

- Type in any keyword or multiple keywords, such as subject terms, all or part of an author name, article title, journal title, location, or year.
- Click the **Find It** button.
- The Search Results page will open to display the results.
- Use Booleans, wildcards or quotation marks to expand or narrow the search.
- To search specific fields, click **Advanced Search**.

Fig. 11: TEEAL quick search
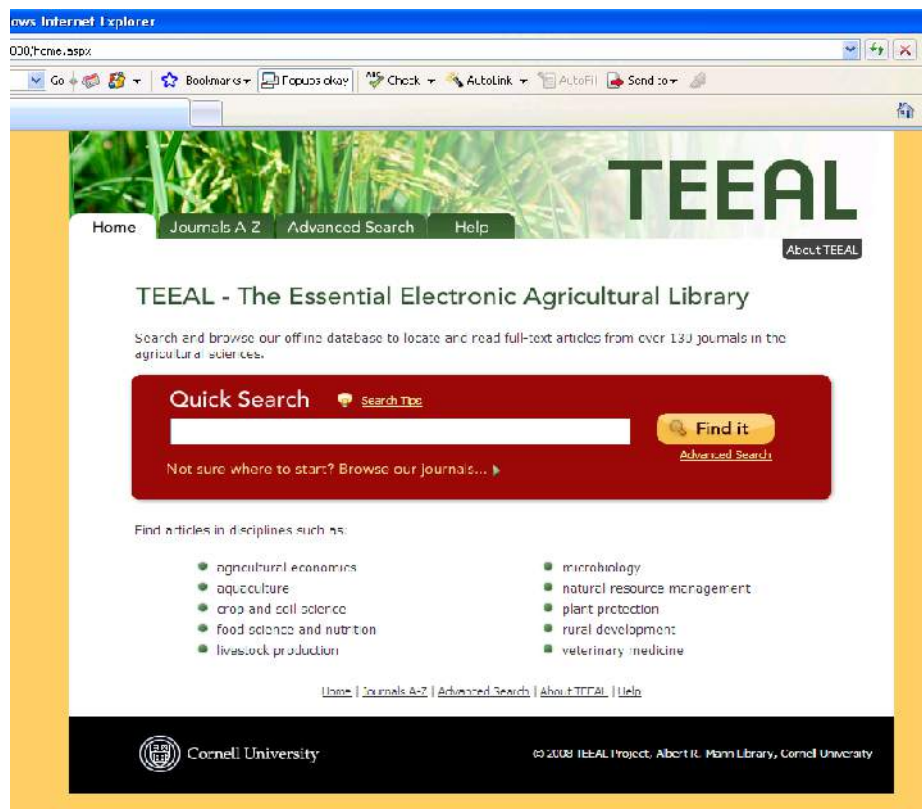
Advanced Search

Advanced Search allows one to select specific fields, such as Article Title, Author, Descriptors, or Date to narrow the search.

Fig. 12: TEEAL advanced search

<u>Search by Date</u>

- Type the year range desired into the boxes labeled "Date From".
- Most journals in the current collection start in 1993 and onwards.

<u>Searching by Journal(s)</u>

- Click on a journal(s) in the "Available" list on the left-hand side of the "Select Journals to Search" box.
- Double-click on a journal(s) or use the Add/Remove buttons to insert/remove journals in the "Selected" list on the right-hand side of the box.

BROWSING

To browse through the collection of journals, click on the **Journals A-Z** tab in the toolbar at the top. Click on a letter to display all the journals starting with that letter.
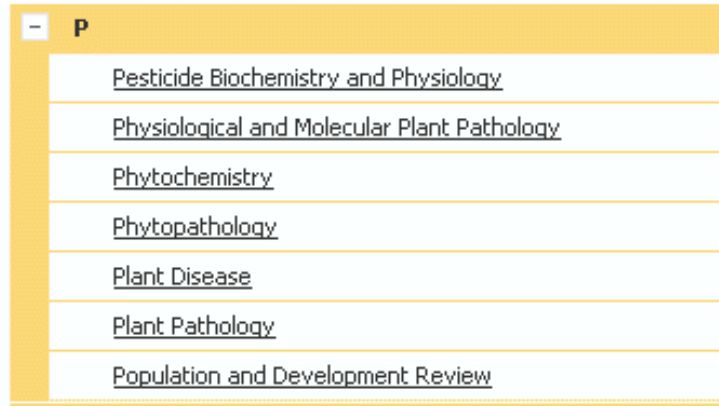
- Click on a journal to see all available years.
- Click on a particular year to see all available volumes and issues.
- Click on an issue to see all available articles.
- To view an article, click the **Full-Text Article in PDF** link located under the article title in the citation.

CHAPTER THREE

SUBJECT DIRECTORIES AND SEARCH ENGINES

3.1   Overview

From the vast quantity of information available on the Web, only a small portion is likely to meet someone's information needs. So, one should be able to locate specific information to meet his/her needs. One way to do this is to use *Subject Directories* and *Search Engines*. These are search facilities designed to allow one to search for, find and view sites of interest. Since the content of the Internet is not controlled, search facilities enable the searcher to impose some order to obtain the required information. The results from different search engines and the sites listed in various subject directories vary; reflecting the purpose, interests and approaches of the people behind a particular directory or a search engine. Thus, no search facility covers the entire Web and no two are exactly the same.

3.2   Subject directories

A subject directory is a catalog of websites collected and organized by humans. Universities, libraries, companies, organizations, and even volunteers tend to create subject directories to catalog portions of the Internet. These directories are organized by subject and consist of links to Internet resources relating to these subjects. The major subject directories available on the Web tend to have overlapping but different databases. Subject directories organize information resources in hierarchical subject categories - they start with a few main categories and then branch out into subcategories, topics, and subtopics. **INFOMINE** from the University of California is a good example of an academic subject directory. Commercial directories such as **Yahoo** have link to a wide range of topics and often emphasize entertainment, commerce, hobbies, sports, and travel.

Most directories offer two ways of searching:
- Navigating through subject categories to a specific topic.
- searching for keywords:
    - in the category headings
    - in the names and descriptions of the websites
    - in a search engine database associated with the directory

Some of the major Internet subject directories include:

- **Academic Info** (**http://www.academicinfo.net**) **-** gateway to college and research level Internet resources maintained by former librarian Mike Madin and a volunteer group of subject specialists.

- **BUBL Link** (**http://bubl.ac.uk**) **-** selective resources from the University of Strathclyde Library in Glasgow, Scotland

- **INFOMINE** (**http://infomine.ucr.edu**) **-** large collection of scholarly Internet resources collectively maintained by several libraries, including those from the University of California.

- **The Internet Public Library** (**www.ipl.org**) **-** large, selective collection from the University of Michigan.

- **Intute** (**www.intute.ac.uk**) - collection of resources for education and research, including collections in Science, Engineering & Technology, Arts & Humanities, Social Sciences and Health & Life Sciences

- **Librarians' Internet Index (http://lii.org**) **-** carefully chosen, organized, and annotated directory maintained by a large group of librarians in California

- **Subject Guides A to Z** (**http://www2.lib.udel.edu/subj)** - extensive collection of subject pages from the University of Delaware Library

- **The WWW Virtual Library** (**www.vlib.org**) **-** guides to many disciplines sponsored by the W3 Consortium

- The **Open Directory** (**http://dmoz.org**) organizes scholarly and popular websites.

- **About** (**www.about.com**) - large collection of topical collections gathered by company-certified subject specialists

- **Google Web Directory** (**http://directory.google.com**) - version of the Open Directory Project using the Google link ranking technology; Google search results are also included with directory results

- **LookSmart** (**www.looksmart.com**) - large collection of links to reviewed sites in thousands of categories

- **Open Directory Project** (**http://dmoz.org**) - significant resource collection compiled by thousands of volunteer editors owned by America Online and promising extensive expansion.

- **Yahoo! Directory** (**http://dir.yahoo.com**) - directory contained within Yahoo portal; lacks reliable site evaluation, so it is not an appropriate source for academic research.

3.3   Internet search engines

A search engine is a programme designed to help find information on a computer system. An Internet search engine allows the user to enter keywords relating to a topic and retrieve information about WebPages containing those keywords. Search engines "index" (record by word) each word within all or parts of documents. When a query is posed to a search engine, it matches the words against the records it has in its databases to present a listing of possible documents. Search engines differ in their search capabilities, search speed, the design of the search interface, the way in which they display results, and the amount of help they offer.

The factors used by search engines to rank their results include:

- Number of occurrence of search term(s) on the page - keywords that appear multiple times in a document's index tend to be ranked higher.
- Where the term(s) occurs - keywords that appear in the document's title, or perhaps metatag description or keyword description fields can be given higher weight than terms only in the document body.
- If the page contains all of the search terms
- If terms are found near or next to each other
- Sites are ranked higher if they have lots of other sites linking to them.
- Rare or less frequent keywords that do not appear as frequently in the engine's index database are often ranked more highly than common terms or keywords.
- The popularity of the target document

Meta-search engines

Meta-search engine send requests to several other search engines and/or databases and returns the results from each one. They are sometimes referred to as *parallel search engines*, *multithreaded search engines*, or *mega search engines*. They allow users to enter search terms only one time and access several search engines simultaneously. Since it is hard to catalog the entire web, the idea is that by searching multiple search engines, one is able to search more of the web in less time.

Search engines have a disadvantage in that they are indiscriminate. They collect relevant and irrelevant information. Furthermore, search engines contain very large databases and as a result, they often return thousands of results. It is necessary to apply techniques that narrow results and push the most relevant pages to the top of the results list.

Examples of general search engines:

- **AllTheWeb** (www.alltheweb.com)

- **AltaVista** (www.altavista.com)

- **Webbrain** (www.webbrain.com)

- **Google** (www.google.com)

- **Yahoo** (http://search.yahoo.com)

- **MSN Search** (http://search.msn.com)

- **Ask** (http://search.ask.com)

- **Lycos** (www.lycos.com)

Examples of general search engines:

- **Metacrawler**
  (http://www.metacrawler.com)

- **Dogpile** (http://www.dogpile.com)

- **43 Marks** (www.43marks.com)

- **Mamma** (www.mamma.com)

3.3.1 Google

At present, the most comprehensive Internet search engine is Google. For that reason, the following section describes information searching techniques using Google. It is assumed that most search techniques applied in Google are applicable to many other search engines.

Fig. 13: Google Home page
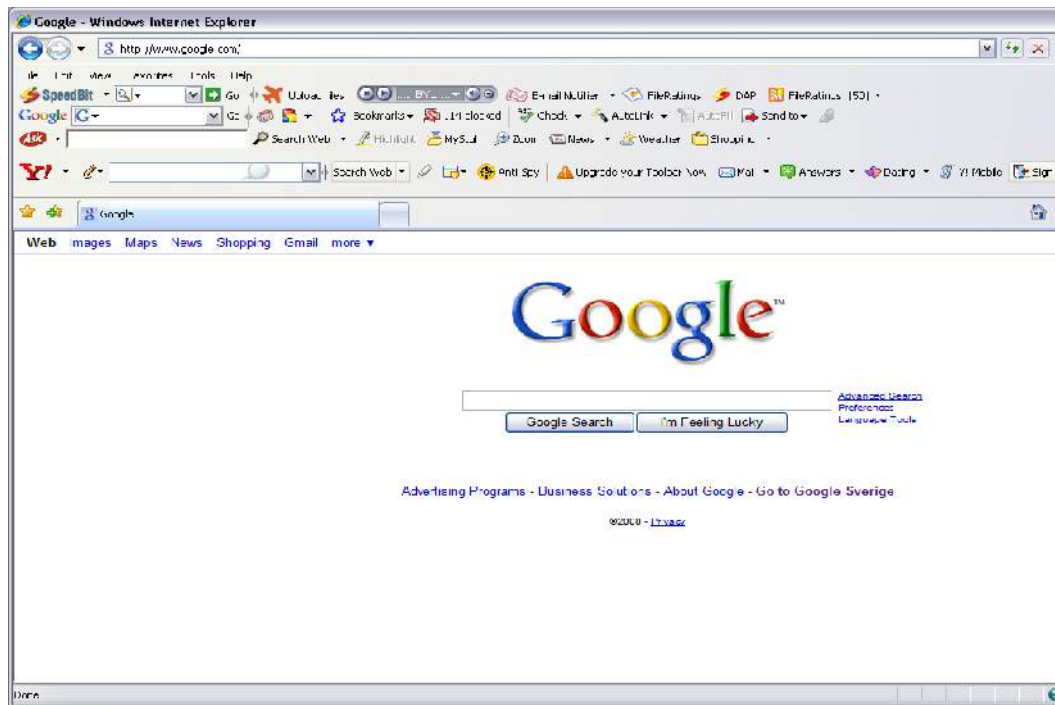
(a)   Crafting a Query

Google uses special characters and operators to craft search queries. Such characters and operators include: **+, -, ~, *, OR,** and **" ".** In addition, Google has special rules that one should observe when searching for information.

- Google is NOT case sensitive; it assumes all search terms are lowercase. The search results will be identical no matter how they are typed; they can even be MiXeD cases.

- Google limits queries to 32 words. It won't search more than 32 keywords at a time as any keyword past 32 is simply ignored.

- Google routinely and automatically ignores common words and characters known as "stop words" including a, about, an, and, are, as, at, be, by, from, how, in, is, it, of, on, or, that, the, this, to, we, what, when, where, which, with. However, Google can be forced to search for a stop word by putting a plus sign (+) in front of it or using quotation marks.

- Search for a phrase or a string of words by using quotation marks " ".

- Google ignores some punctuation and special characters e.g. **! ? , . ; [ ] @ / # < > .**

- A term which is not required for retrieval should be preceded with minus sign (-). e.g. gender –children

- Synonyms can be retrieved by proceeding the term with a ~, which is known as the tilde or synonym operator. e.g. ~inexpensive matches "inexpensive," "cheap," "affordable," and "low cost".

- In case the synonyms that Google suggests using ~ operator do not satisfy the searcher, one should specify own synonyms with the OR or |.

- A star sign (*) is used to match any word in a phrase e.g. "poverty * strategies"

(b)   Beyond +, -, ~, |, *, OR, and " "

A number of search operators (query modifiers) can be used to enhance information searching through Google. This section describes available query modifiers. Each entry typically includes the syntax, the capabilities, and an example.

allintitle: The query **allintitle:*term*s** restricts results to those containing all the query terms you specify in *the title*. The entry **allintitle:detect plagiarism** will return only documents that contain the words "detect" and "plagiarism" in the title. The query **intitle:*term*** restricts results to documents containing *term* in the title. For instance, **dryland agroforestry**

intitle:problems will return documents that mention the word "problems" in their titles, and mention the words "dryland" and "agroforestry" anywhere in the document.

 allintext: The query **allintext:***term***s** restricts results to those containing all the query terms you specified in the *text of the page*. The entry **allintext:rural development** will return only pages in which the words "rural" and "development" appear in the text. The query **intext:***term* restricts results to documents containing *term* in the text. For instance, **sisal export intext:Tanzania** will return documents that mention the word "Tanzania" in the text, and mention the names "sisal" and "export" anywhere in the document.

allinurl: The query **allinurl:***term***s** restricts results to those containing all the query terms you specify in the URL. The entry **allinurl:tourism Tanzania** will return only documents that contain the words "tourism" and "Tanzania" in the URL, such as www.**tanzania**.go.tz/**tourism**.html. The query **inurl:healthy eating** will return documents that mention the words "healthy" in their url, and the word "eating" anywhere in the document.

define: The query **define:***term* shows definitions for the term from pages on the web. This search operator is useful for finding definitions of words and phrases. The entry **define:globalization** will show definitions for the term "globalization".

filetype: The query *term***s filetype:ext.** restricts the results to documents in that file format. The entry **cotton disease tropics filetype:pdf** will return pdf files that match the terms "cotton," "disease," and "tropics". One can also restrict the results to documents with different file formats e.g. [cotton disease tropics filetype:pdf OR filetype:doc].

site: If you include **site:** in your query, Google will restrict your search results to the site or domain you specify. For example **admissions site:www.suanet.ac.tz** will show admissions information from SUA's site.

(c)   Google Scholar

Google Scholar is a search engine that indexes the full text of scholarly literature across publishing formats and scholarly fields. It is a part of the Google search engines that enables people to search specifically for scholarly literature, including peer-reviewed papers, theses,

books, preprints, abstracts technical reports and selected Web pages that are deemed to be "scholarly" from all broad areas of research. However, since most of Google Scholar's search results link directly to commercial journal articles, often users will only be able to access a brief summary of information regarding the article, and possibly will be asked to pay a fee to access the entire article.

Fig. 14: Google scholar page
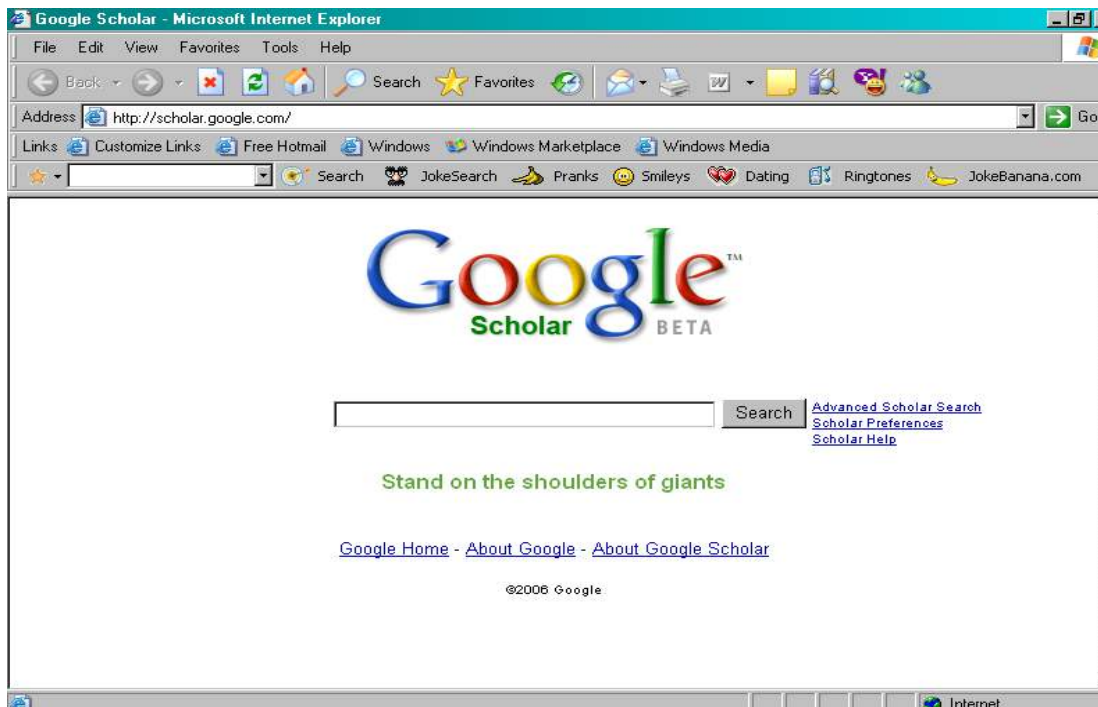
(d)   Google Advanced Search

One of the key features of the Google advanced search is the use of Boolean operators to combine the words or phrases that one is searching for. Google uses boxes explaining the type of search that one can carry out. With Advanced Search, one can search for documents:

- That contains **all** of the words, which is the equivalent of the Boolean AND operator.
- With the **exact phrase.** This is the same as using quotation marks.
- With **at least one** of the words, which is the equivalent of the Boolean OR operator
- Without the words. This is the equivalent of the Boolean AND NOT operator.

In addition to constructing an effective search string using Boolean operators (in the provided boxes), one can use Google's advanced search features to **limit** the search.

- **Language:** To limit your search to pages written in a language of your choice.
- **File Format:** To search the web for specific type of files such as Portable Document File (.pdf), Word (.doc) or Excel (.xls).
- **Date:** Limit by when the page was last updated. This enables one to limit a search to the most recent pages or a certain period of interest.
- **Occurances:** This option allows you to specify that the keywords appear only in the URL, title, or the text of the page. Limiting a search to page titles can be a powerful tool for focusing your search.
- **Domain:** This limits the search by the information that is found in the URL (Uniform Resource Locator) or Internet address.

Fig. 15: Google Advanced search
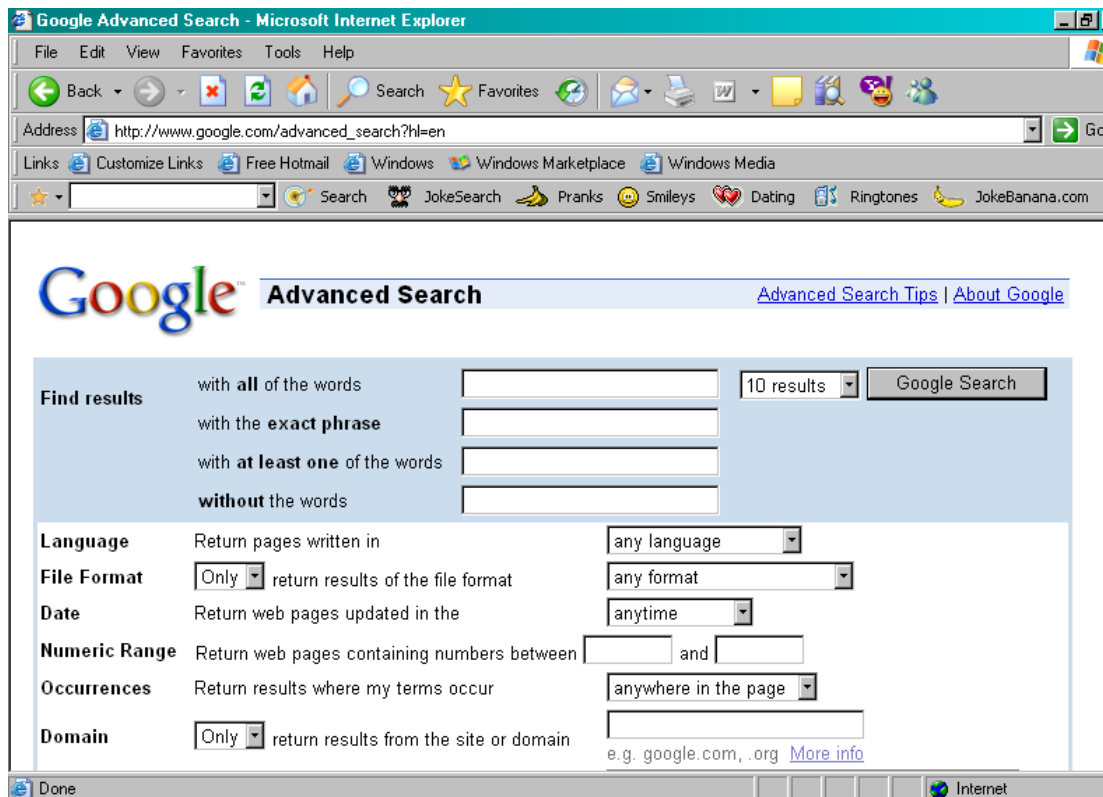
3.4   Deep Web

Searching information on the web has been compared to dragging a net across the surface of the ocean. While a great deal may be caught in the net, there is still a wealth of information that is deep, and therefore, missed. The reason is that most of the Web's information is on

dynamically generated sites, and standard search engines never find it. Search engines create their indices by crawling surface Web pages. Deep web pages cannot be found by search engines like Google because they are within specialized databases. To be discovered, the page must be static and linked to other pages. Traditional search engines cannot retrieve content in the Deep Web because such pages do not exist until they are created dynamically as the result of a specific search. It is believed that the Deep Web is made up of valuable material from specialized databases and often includes sites that require registration. Therefore, the Deep Web (also called Deepnet, the invisible Web, dark Web or the hidden Web) refers to the pages on the World Wide Web that are not indexed by conventional search engines. It is important to know where to go to search for this information in the deep web.

Sites that collect content from the deep Web include:

- **http://aip.completeplanet.com**
- **http://www.invisibleweb.com/**
- **http://www.invisible-web.net/**
- **http://www.search.com/**

3.5   Web 2.0 technology

The term "Web 2.0" describes a variety of web services and applications that allow interactive information sharing, user-centered design, and collaboration on the World Wide Web. A key element of Web 2.0 is that it does not require any web design skills to participate. This suggests that the Web 2.0 technology allow users to do more than just retrieve information. They can upload information on a Web 2.0 site and exercise control over that information. While Web 2.0 technologies are mostly used for social networking, they are increasingly being used for educational purposes. These new technologies make sharing of information among users much easier. The technologies change the way documents are created, used, and shared. It should be noted however there is no quality control for much of the information available through these technologies. Such information should be used with a caution, especially if used for academic and research purposes!

A list of Web 2.0 includes:

- **Blogs:** A blog is a short form of the term "Web log". It is a website, usually maintained by a group or an individual with regular entries of commentary, descriptions of events, or other material such as graphics or video. The posting and

commenting process contributes to the exchange of views between a primary author and a group of secondary comment contributors. Examples of blogs include:

- **Intute: social sciences blog** (**www.intute.ac.uk/blog**) provides a weekly round-up of news, updates and links.
- **Social Science Statistics Blog** (**www.iq.harvard.edu/blog/sss**) supports conversations about social science statistical methods and analysis from the Institute for Quantitative Social Science and related research groups.

- **Wikis:** A wiki is a page or collection of web pages designed to enable anyone who has access to contribute or modify content. Wikis are often used to create collaborative websites or community spaces for discussion. Examples include:

  - **Wikipedia (www.wikipedia.org***)* serves as a free multimedia encyclopedia whereby anyone can edit it, with no formal academic peer review.
  - **Scholarpedia (www.scholarpedia.org):** Written exclusively by professionals focusing on their field of expertise; subject to peer review. Its features more commonly associated with open-access online academic journals.
  - **Wikiversityen (wikiversity.org/wik/):** Supports free learning communities, projects, materials, and learners. It offers a series of tutorials and courses to foster learning.

- **Social networks:** A social network is a web-based service that provides ways for users to interact, such as file sharing, blogging, and discussion groups, to build communities of people who have common interests. A good example is:

  - **Facebook:** A free access social networking website. Users can join networks organized by location to connect and interact with other people. There are many dedicated groups to join, and some may be of academic interest.

- **Online chat** refers to any kind of communication over the Internet, but generally means direct one-on-one chat or text-based, using tools such as instant messengers.

- **YouTube** is a video sharing website where users can upload, view and share video clips. Increasingly educational videos are being posted by lecturers on YouTube.
- **Flickr** is an image and video hosting website and online community platform. While they are mostly used for social networking, they can also be used for educational purposes.

- **Podcast** is a series of audio or video digital media files which is distributed over the Internet by syndicated download, through web feeds, to portable media players and personal computers. The same content may also be made available by direct download or streaming. Examples of podcast sites for social researchers include:

    o **LearnOutLoud.com** is a one-stop destination for audio and video learning where one can browse educational audio books, MP3 downloads, podcasts, and videos. Many cover interesting research findings.

    o **OUPodcasts** provide podcasts in areas of criminology, economics, geography, media studies, politics, psychology, social policy and sociology.

- **Mailing lists** provide forums for the exchange of views for particular topics. Examples include the following:

    o **qual-software@jiscmail.ac.uk** is a discussion list to increase awareness and debate about Computer Assisted Qualitative Data Analysis Software. The list provides an instant forum for users and developers to air problems, offer opinions, argument and advice on the variety of packages in use.

    o **qualitative-research@jiscmail.ac.uk** is a list devoted to all aspects of qualitative research: methodological; theoretical and practical and is also intended to facilitate discussion of diverse qualitative research: interviewing; ethnography; participation observation; focus groups; biographical and life history studies.

- **Social bookmarking** sites allow users to store, describe, and share web addresses with others. Users can explore bookmark collections of others by subscribing to their bookmark pages. Students could also collaborate on group projects using book-marking sites, sharing links, and uploading resources discovered, while educators could follow their students bookmark pages to gain insight on their research process and progress.

CHAPTER FOUR

OPEN ACCESS SCHOLARLY CONTENT

4.1 Overview

Whenever searching for information from the web as demonstrated in the previous chapter, the search output results into both free materials and those that can be accessed by subscription only. In the former instance, one is able to access full text publication freely without restrictions. In the latter scenario, only bibliographic information of the documents such as author, title, source of publication and sometimes the abstract can be accessed. Access to full text is made available through the Internet on the emerging mode of scholarly publishing known as Open Access.

The Berlin Declaration (2003) defines open access as a mode of scholarly communication through which the author(s) and right holder(s) of scholarly work grant(s) to all users, a free access to, and a license to copy, use, transmit, and distribute derivative works, in any digital medium for any responsible purpose, subject to proper attribution of authorship. According to this definition, a complete version of the work and all supplemental materials, including a copy of the permission to use should be deposited in at least one online repository using suitable technical standards to enable open access, unrestricted distribution, and long-term archiving of such works. This new form of scholarly communication is achieved through two main channels: Open Access Journals (OAJ) for electronic refereed journals and Self-archiving (Chan and Costa, 2005; Bailey, 2006). It should be noted that by searching specific open access sources on Internet, the subscribed content is eliminated. The following sections briefly describe some of the open access outlets on the Internet.

4.2 Open Access Journals

Open access journals also referred to as "Gold Road" to open access, are peer reviewed journals made available free of charge to the public through the Internet. Unlike the business publishing model, in open access publishing the end user is not charged to access journal articles. Various funding strategies such as direct author fees, institutional membership to sponsor, all or part of author fees, funding agency payment of author fees, grants to open access publishers and institutional subsidies are used to cover the costs for publication and distribution of OA content for free access by the end user (Hirwade and Rajyalakshmi, 2006). The Directory of Open Access Journals (DOAJ) (**www.doaj.org**) provides a list of registered

freely accessible journals on the Internet. By the end of April 2010, DOAJ had 4905 journals with a total of 383, 919 freely accessible articles.

Similar techniques of online information search as elaborated in the previous sections apply for one to discover scholarly materials from open access journals. As indicated in Figure 10, one can find articles from DOAJ by clicking a link on "find journals". A search template will appear and the user may opt for a simple or advanced search to discover articles of interest. The user may also choose to search specific journals through "find journals" option or discover journals of interest by browsing journals by title or subject. Using the last option as an example, clicking "agriculture and food sciences" link results to journals in agriculture (general), animal science, aquaculture and fisheries, forestry, nutrition and food science, and plant sciences sub-disciplines. By clicking on "agriculture (general)" link, a list of individual journals for this sub-discipline will be displayed. At this stage, one can select a journal to view individual articles.

Fig.16: DOAJ search template
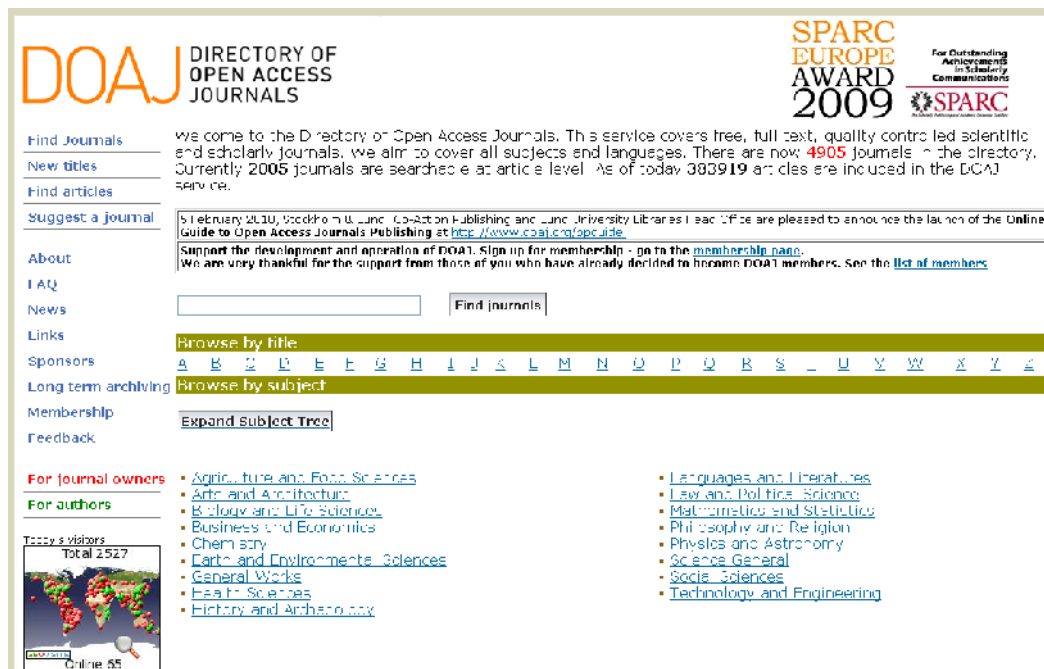
4.3   Self-archiving

Self-archiving also referred to as "Green Road" to open access is making articles freely available in digital form on the Internet by authors (Budapest Open Access Initiative, 2002).

There are three most common ways of self-archiving on the Internet: authors' personal websites, disciplinary (research-specific), and institutional repositories of individual universities/institutions (Bjork, 2004, Bailey, 2006). The Registry of Open Access Repositories (ROAR) and the Directory of Open Access Repositories (DOAR) provide the list of open access compliant archives from disciplinary and institutional archives Worldwide. As the case with open access journals, articles from ROAR or DOAR may be accessed through direct search of respective repositories/directories using the following links:

- The Registry of Open Access Repositories (ROAR) (**http://archives.eprints.org**)
- The Directory of Open Access Repositories (DOAR) (**http://www.opendoar.org**).

By end of April 2010, DOAR had over 1500 listings of disciplinary/institutional repositories Worldwide. It should be noted that in most cases a single repository is registered in both of the two directories. In these directories, two main types of repositories can be identified - disciplinary or institutional repositories.

4.3.1 Disciplinary repositories

Disciplinary repositories include e-prints from one or multiple scholarly disciplines and are usually hosted at a central server. Some of the existing disciplinary archives are:

- Arxiv (**http://arxiv.org**) - the oldest disciplinary archive which covers physics, mathematics, non-linear- science, computer science and quantitative biology;
- E-LIS (**http://eprints.rclis.org**) for computing and library/information science disciplines;
- AgEcon (**http://agecenresearch.umn.edu**) for full-text scholarly material in agricultural and applied economics; and
- RePEc (**http://www.repec.org**) covering the economics disciplines.

4.3.2 Institutional repositories

Institutional repositories (IRs) include e-prints or other types of digital works by authors in a single academic department or school or the whole institution. It is possible to identify institutional repositories by browsing the list of either DOAR or ROAR. It should be noted that the list for disciplinary and institutional repositories is note exhaustive. Depending on individuals' specific research interest one can go through each of these directories to identify suitable repositories for their specific searches. The digital works in such repositories [disciplinary or institutional repositories] may comprise of electronic theses and dissertations

(ETDs), working papers, journal articles, learning objects, presentations, technical reports and software components (Bailey, 2006).

As is the case for DOAJ, one may search DOAR or ROAR directly to discover articles or documents from open access archives which are registered in these directories. Each of these directories provides search templates to guide users in their information search process. It is also possible to obtain documents or articles from DOAR or ROAR through a single point search from OAIster site (**http://www.oclc.org/oaister**/). The OAIster database is a union catalogue of millions of records harvested from open access archive collections throughout the world. By end of April 2010, OAIster had over 23 million records representing digital resources that were online freely accessible.

CHAPTER FIVE

QUALITY CONTROL AND CITATIONS

4.1 Evaluating information sources

With so much information available, in different formats, from a variety of sources, information collected must be carefully reviewed to ensure its quality. The ability and popularity of self-publishing on the Web make it possible for *anyone* to post *anything* and make it available worldwide. Unfortunately, there is neither established publishing process nor umbrella organization to ensure some kind of quality control of the Web content. In addition, the ephemeral nature of much Web content means that what is found on the Web today may vanish or being changed tomorrow. Thus, it is important to ensure that the information obtained is of high quality and relevant to the one's information needs. The following criteria can be used to evaluate information and its sources:

Authority

- Who is the author/creator of the information?
- Evidence to show that the author is knowledgeable, reliable, and truthful.
- The reputation of the organization behind that information.
- There should be author's credentials such as:
  - Education, training, and/or experience in a field
  - Author's title or position of employment.
  - Author's reputation or standing among peers.

Currency

- Is the publication date clearly stated?
- How current is the information?
- Is the web page maintained and updated regularly?

Accuracy

- Is the information accurate? How does it compare with other sources on the subject? Is it consistent with what you know about the subject?
- Is the information complete? Does it provide enough evidence to support its claim?
- How reliable and error free is the information?
- Is there evidence of potential bias?

Objectivity

- Does the source present a balanced view on the topic?
- Is the bias of the author/creator obvious?
- Is the source trying to convince the user on a particular point of view?
- Is the publication in which the item appears published, sponsored, or endorsed by a political or other special interest group?

Intended audience / Level of information

- Who is the intended audience… the general public? the educated? layperson? professionals? practitioners? scholars?
- Is it written at a level that is understandable? Does it makes sense? Consider the vocabulary used.
- Does it build on what you already know? Does it include a bibliography or links to additional sources to consult?

4.2 Plagiarism

There is no widely applicable definition of plagiarism as different disciplines and institutions may have different perceptions on what might be considered "common knowledge". However, a widely shared understanding is that plagiarism occurs when someone tries to present someone else's work, thoughts or ideas as his/her own, whether deliberately or unintentionally, without crediting the source. It is the false assumption of authorship. In the academic world, plagiarism is a problem because it is a form of academic misconduct or dishonesty. It undermines the whole basis of scholarly academic values, and undermines academic standards and the credibility of awards. In addition, plagiarism demotivates those students who see their efforts as being undermined by the unfair advantage gained by others.

Forms of plagiarism

Intentional plagiarism occurs when writers know well that they are passing off someone else's words or ideas as their own. Examples of intentional plagiarism include:

- Passing off one's own pre-written papers.
- Cutting and pasting from more than one source to create a paper without giving credit.
- Borrowing words or ideas from other authors without giving credit.

- Failing to put quotation marks around the words of others (if the words are taken directly without paraphrasing).

Unintentional plagiarism or **accidental plagiarism** is often the result of poor citation or referencing, perhaps because the author do not know how to.

Self-plagiarism is the act of copying own published or submitted writing without mentioning the previous use or publication. For example, in academic assignments, the submission of the same work in more than one course is considered self-plagiarism.

Punishments of plagiarism

Plagiarism is a serious academic offence which can result in punishment such as failing a grade on the particular assignment. Being found guilty of plagiarism can ruin an academic career; it may result in revocation of one's degree, or the loss of one's job, and will result in the loss of academic credibility. Many institutions have own policies, regulations and procedures of dealing with plagiarism.

Avoiding plagiarism
- In academic writing, one has to read and understand relevant material but not copying everything. The ideas are owned by the writers.
- Plagiarism is avoided by learning and using the agreed citation style. Students, lecturers and researchers are expected to learn and adopt the style that has been adopted by the university or institution they are attending.
- Use quotation marks when using someone's language and provide pages from the source.
- Avoid plagiarizing someone's grammatical structure when you paraphrase by setting the original source aside and rewriting the concepts in your own words.

5.3   Citations

Any academic or research work should make the reader agree that what is said is true. Writers should not simply make assertions and generalizations of facts, unless that fact is something about which there can be no doubt and which the reader can be expected to know. For example, it is unlikely for people to urge if one asserts that *information technology is having a major impact on organizations and society*. However, if one states that *the*

introduction of IT has led to the loss of jobs in the Tanzanian banking industry, evidence must be cited to support that claim. So, a citation is a reference to a primary or secondary source with sufficient details to identify the cited work. Generally, a reference should be given for:

- Assertions of facts that cannot be regarded a common knowledge
- Opinions and generalizations derived from other writers
- Tables and diagrams
- Direct quotations

There are several citation methods and the Harvard Citation Style is the most commonly used method throughout the world. This method is used at Sokoine University of Agriculture as well. The following sections describe the Harvard citation method:

5.3.1 Citing within the text

When referring to someone's work, the author's surname and the year of publication are placed in the text in brackets.

Agroforestry and agricultural interventions need to take into account the gender division of labour (Njuki, 2001).

Direct quotations

If directly quoting a piece of work one should enclose the quotation in quotation marks and then identify the page numbers the quotation is taken from.

…"There have been no defined national forest research priority areas although the nine prioritized TAFORI research programmes are always mistakenly taken to be the national priority research areas" (Abeli,2000, p. 108).

Multiple authors

Three or fewer authors

Cutler, Williams and Schmitt (1986) argued that the fermentation process is prolonged by such low temperatures.

Four or more authors

In the case of four or more authors, only the first author is written followed by '*et al.*' (this is an abbreviation of the Latin term for 'and others').

Ishengoma *et al.* (2000) argue that *Albizia amara* is among the lesser utilized species in Tanzania.

If the author's name occurs naturally in the text:

The year follows in brackets.

A study by Njuki (2001) investigated the gender roles in the agroforestry system and how these affect productivity.

If the source is cited within another book:

For sources that one has not actually seen but which are referred to in another work, both the original and the secondary sources must be cited.

Adamu (2001) as quoted in Geleja (2004, p.36) states that…

If there is more than one reference at the same point in the text:

When more than one reference is given at the same point in the text, they should be listed in order of publication date (earliest first).

Musa (1997), Brown (1999) and Sife (2003) agree that information seeking behaviour of forestry researchers …

This can also be written using semicolons to separate the authors:

A number of studies (Musa, 1997; Brown, 1999; Sife, 2003) have investigated the information seeking behaviour of forestry researchers in the electronic environment.

If there is more than one author with the same surname and the same year:

In the main text, author's initials are only used, in brackets, when two or more authors have the same surname and have published in the same year, in which case they should be identified by initials in order to avoid confusion.

> Cover crops have strong influence on nitrogen availability and soil moisture content (Hassani, S., 1994). This, in turn, has influence on maize yield (Hassani, B.L., 2002)

If the author has more than one item published in the same year:

Use lower case letters after the date if referring to more than one item published in the same year by the same author.

> (Musa, 2006a), Musa (2006b)

If there is no date on the publication:

If the date of publication is not known, use "n.d." ("not dated") in place of the year.

> It has been reported that there is no healthy relationship between farmers and pastoralists in Morogoro (Makiya, n.d.).

If there is no author:

If the author(s) of a work can't be established, include the title within the brackets.

> …This disease occurs more frequently in men (Preventing coronary heart disease, 2003).

For visual information, photographs, illustrations and diagrams:

Acknowledge the source from which the image has been taken in the same way as would be in quotations; with the author and date given alongside in brackets and full details in the bibliography or list of references.

For personal communications:

In rare cases, personal communications can be cited in the text, but as the information is not recoverable they do not need to be included in the bibliography. Give initials as well as the surname of the communicator, and provide as exact a date as possible.

> More and more buildings are now providing access for disabled people (R. J. Thompson, personal communication, 22 March 2000).

Web site with an author:

Cite the author and date as usual.

> Turner (2001) has created many useful graphs showing the transition between different age groups.

Web site without an author:

• Where the web page has no author whatsoever, use the title as the point of reference.

> Cancer experts estimate that changes to our diet could prevent about one in three cancer deaths in the UK (CancerHelp UK, 2002).

Electronic journal citation

• Cite the author and date as you would for printed sources.

5.3.2 Citation in the list of references/bibliography

A list of references usually comes at the end of the text (essay, research report or paper) and should contain only those works cited within the text. The term **reference** is normally used to cover works cited in the text. On the other hand, the term **bibliography** refers to works read, whether cited or not. It includes documents one made use of, additional background reading, and any other articles the author thinks the reader might need as background reading.

In the Harvard citation method:

- Arrange references in alphabetical order of authors and in order of the year of publication.
- References by one author take precedence over references by same plus additional authors irrespective of year of publication.
- Where no name of an author is given, use the name of the sponsoring or issuing organization. The title of the document should be used in case author's name and the name of the sponsoring or issuing organization are not shown in the document.

Books

The details required, in order, are:

- **Name** of author(s) or editor(s) (surname, and initials), or the institution responsible.
- **Year** of publication in brackets.

- **Title** and **subtitle** (if any) – <u>underlined</u>, in **bold** or in *italics* and followed by a full stop.

- **Series title** and individual **volume** if any – followed by a full stop.

- **Edition**, if any.

- **Place of publication** if known – followed by a comma**.**

- **Publisher** – followed by a full stop.

> Basford, L. and Slevin, O. (eds.) (1995) **Theory and practice of nursing: an integrated approach to patient care.** Edinburgh, Campion.
>
> Dutta, A.C. (1979) **Botany for degree students**. 5th ed. Calcutta, IBH Publishing Co.

'In' references

Individual chapter in multi-authored book

The details required, in order, are:

- Name of author(s) of a chapter/section

- Year of publication - brackets.

- Title of chapter/section.

- The word **In:**

- Author/editor of collected work

- Title of a collected work- <u>underline</u>, in **bold**, or in *italics* and followed by a full stop.

- Place of publication – followed by a comma.

- Publisher – followed by a comma.

- Range of chapter or article pages - followed by a full stop.

> Gardner, B.R. (1993) Ensuring successful IT utilization in developing countries **In:** Ojo, S ed. **Information technology utilization in developing countries.** Gaborone, Printing and Publishing Co, pp. 11 - 20.

Secondary reference (book cited within another book)

- A slightly different example of an 'in' reference would be for a piece of work cited within another book, known as a secondary reference as you have not actually seen the original source yourself.

- Here you would include the date and publication details of the piece of work, and then '**Quoted in:**' followed by the reference for the book you have seen.

> Mbwambo, L. (2004) Status of agroforestry in arid and semiarid areas of Tanzania. **Quoted in**: Jama, B. and Zeila, A. (2005) **Agroforestry in the drylands of eastern Africa: a call for action**. ICRAF Working Paper – no. 1. Nairobi.

Article in a journal

The details required, in order, are:

- Name(s) of author(s) of the article (surname, and initials).
- Year of publication – in brackets.
- Title of article – followed by a full stop.
- Title of journal – underlined, in **bold**, or in *italics* and followed by a full stop.
- Volume and issue number.
- Range of article page numbers – followed by a full stop.

> Lwehabura, M.J.F. and Matovelo, D.S. (2000) Effective library management: issues for managers and subordinates. **New library world**. 101 (1158), pp. 263 - 268.

Newspaper articles

> Mayallah, P (2006) Government to prioritize technological innovations. Daily News, 5th September, p. 3.

Theses and dissertations

• You need to include the name of the awarding institution

> Chilimo, W. L. (2002) **Accessibility and use of nutrition information in addressing underfive child malnutrition in Morogoro Urban, Tanzania**. M.A. thesis, University of Dar es Salaam.

Research reports

- No one example can explain the type and range of reports. However it is important to include the subtitle and series information.

- The correct content and order is the same as for books.

Nchimbi-Msolla, S. (2005) **The impact: assessment and perspectives**. Research Report. Morogoro, TARP II – SUA Project.

Proceedings of workshop/conference
- Contributing author
- Year of publication – in brackets
- Title of conference paper – followed by a full stop.
- The word **In:**
- Author/editor of conference proceeding (if any)
- Title of conference proceeding including date and place of conference – underline, in **bold**, or *italics* and followed by a full stop.
- Place of publication – followed by a comma
- Publisher – followed by a comma
- Range of article page numbers – followed by a full stop

Lema, N.M and Kapange, B. (2004) The potential of participatory approaches in rural development. In: Batamuzi, E.K., Tarimo, A.G.P. and Kinabo L.D.B. **Proceedings of the 3rd collaborative research workshop, Morogoro,** TARP II – SUA Project, pp. 3 – 11.

Encyclopaedia entries
- Author(s) or editor(s)
- Year of publication – in brackets
- Title of article – followed by a comma
- The word **In:**
- Title of encyclopaedia – underlined, in **bold**, or in *italics* and followed by a comma
- Volume number if applicable – followed by a comma
- Edition number – followed by a full stop.
- Place of publication – followed by a comma
- Publisher – followed by a full stop

Hibbard, J.D., Kotler, P. and Hitchens, K.A. (1997) Marketing and merchandising. In: **The new Encyclopedia Britannica**, vol. 23, 15th revised ed. London, Encyclopaedia Britannica.

Dictionaries

Dictionaries normally do not have an author as such, so references are based on their titles.

- Title of dictionary – underlined, in **bold**, or in *italics* and followed by a full stop

- Year of publication – in brackets

- Volume number (if applicable) or, if you are referencing the complete work, define the volume set (for example vols. 1-32) – followed by a comma

- Edition number – followed by a full stop

- Place of publication – followed by a comma

- Publisher – followed by a full stop

The Oxford English dictionary. (1989) vol. 5, 2nd ed. Oxford, Clarendon.

Publications written by a corporate body

United Republic of Tanzania (2005) **National Strategy for Growth and Reduction of Poverty**. Dar es Salaam, Government printer.

5.3.3 Citing electronic information

Electronic journal articles

Include the following information. The order should be:

- Author/editor

- Year – in brackets

- Title of article – followed by a full stop

- Title of journal – underlined, in **bold**, or in *italics*

- The word **Internet** – in square brackets and followed by a comma

- Volume number

- Issue number – in brackets and followed by a comma

- Pagination or online equivalent – followed by a full stop

- The words **Available from** – followed by a colon

- The Internet address – in chevrons

- The word **Accessed** and the date you viewed the web page – in square brackets and followed by a full stop

Shepherd, B., and J. Whittington (2006) Response of wolves to corridor restoration and human use management. **Ecology and Society** 11(2): 1. [Internet] Available from: <http://www.ecologyandsociety.org/vol11/iss2/art1/> [Accessed 4 October 2006]

Electronic books

Include the following information and the order should be:

- Author/editor
- Print version year (use the electronic version year if there is no print version) – in brackets
- Title – underlined, in **bold**, or in *italics*
- The word **Internet** – in square brackets followed by a full stop
- Print version place of publication – followed by a comma
- Print version publisher – followed by a full stop
- The words **Available from** – followed by a colon
- The Internet location the electronic book was accessed from (e.g. netLibrary)
- The Internet address – in chevrons
- The word **Accessed** and the date you viewed the electronic book – in square brackets and followed by a full stop.

Ker, A. (1995) **Farming Systems of the African Savanna: A Continent in Crisis.** [Internet] Ottawa, IDRC. Available from: IDRC books <http://www.idrc.ca/en/ev-9355-201-1-DO_TOPIC.html > [Accessed 6 October 2006].

Web pages

Include the following information and the order should be:

- Author/editor
- Year – in brackets
- Title – underlined, in **bold**, or in *italics*
- The word **Internet** – in square brackets and followed by a comma
- Edition if available (for example version 4.1) – followed by a full stop
- Place of publication – followed by a comma

- Publisher (if ascertainable) – followed by a full stop

- The words **Available from** – followed by a colon

- The Internet address – in chevrons

- The word **Accessed** and the date you viewed the web page – in square brackets and followed by a full stop.

> Kaliba, A.R.M., Verkuijl, H., Mwangi, W. and Byamungu, D.A. (1998) **Adoption of Maize Production Technologies in Western Tanzania.** [Internet] Available from: http://www.cimmyt.org/Research/Economics/map/impact_studies/ImpStudEA_List/Western Tnz.pdf >, [Accessed 7 July 2006]

Whole web sites

Sometimes you may need to reference a whole web site, an example is shown below:

> International Hotels Environment Initiative. (2004) **International Hotels Environment Initiative (IHEI)** [Internet]. London, IHEI. Available from: <http://www.ihei.org/> [Accessed 04 March 2004].

CD-ROMs

- Citing of information from computer databases varies. If you have, for example been using a CD-ROM to obtain journal references you only need to cite the journal as your source of information not the CD-ROM as this is simply a bibliographic database.

- If you have used a full text journal article from a CD-ROM, however, the reference would contain the same information as for an electronic journal found on the Internet [except that the URL is not indicated in this case].

BIBLIOGRAPHY

American Library Association (1989) **Presidential Committee on Information Literacy: Final Report.** [Internet]. Available from: <*www.ala.org/ala/mgrps/divs/acrl/.../presidential.cfm*>(Accessed 20 March 2010)

Bailey, C.W. (2006). *What is open access?* [Internet]. Available from: <http:www.digital-scholarship.com/cwb/whatIsOA.htm> (Accessed 8 August 2007).

Berlin Declaration of Open Access. (2003). [Internet]. Available from: http:www.zim.mfg.de/openaccess-berlin/berlin_declaration.html (Accessed 24 July 2007).

Bjork, B.C. (2004). Open access to scientific publications-an analysis of the barriers to change. *Information Research, 9(2).* [Internet]. Available from: <http://InformationR.net/ir/9-2/paper170.html> (Accessed 2 September 2006).

Blachman, N and Peek, J. (2006) **GoogleGuide: Making Search even easier.** [Internet] Available from: <http://www.googleguide.com > [Accessed 2 August 2006]

Budapest Open Access Initiative. 2002. [Internet]. Available from: <http://www.soros.org/openaccess/read.shtml October 2006)>.

Chan, L. and Costa, S. (2005). Participation in the global knowledge commons: challenges and opportunities for research dissemination in developing countries, *New library world,* 106 (1210/1211): 141-163.

Cohen, L (2006) **Internet tutorials** [Internet] Available from: <http://www.internettutorials.net> [Accessed 30 September 2006].

Directory of Open Access Journals (DOAJ). 2010) [Internet]. Available from: <http://www.doaj.org> (Accessed 25 April 2020).

Directory of Open Access Repositories (DOAR) (2010) [Internet]. Available from: <http://www.opendoar.org> (Accessed 25 April 2010).

Hirwade, M. and Rajyalakshmi, D. (2006). **Open access: India is moving towards third world super power.** [Internet] Available from: <*http://eprints.rclis.org/archive/00006798/01/99107D29.pdf*> *(*Accessed 14 October 2006).

Intute: Virtual Training Suite (2010). Internet for Social Research Methods: A free online tutorial. [Internet] Available from: http://www.vts.intute.ac.uk/tutorial/socialresearchmethods [Accessed 5 May 2010].

Korfhage, R.R (1997) **Information storage and retrieval.** New York, Wiley Computer Publish.

Learning Support Services (2004) **The Harvard Style of referencing published material:** Including electronic information [Internet], Leeds Metropolitan University. Available from: <http://www.leedsmet.ac.uk/lskills> [Accessed 2 October 2006]

MSU (2006) **Teeal tutorial** [Internet] Available from: <http://www.msu.ac.zw/libraries/teeal.htm> [Accessed 29 Sep 2006]

OAIster (2010). **The OAIster database**. [Internet] Available from: <http://www.oaister.org> [Accessed 25 April 2010].

TEEAL/AGORA Africa Office (n.d) **How to Use AGORA** [Internet] Available from: <http://www.msu.ac.zw/libraries/tutorials/agora.ppt> [Accessed 2 October 2006]

Wikimedia Foundation (2006) **Plagiarism** Available from: <en.wikipedia.org/wiki/Plagiarism> [Accessed 2 October 2006].