

Lecture no. 11

Department: Information and Library Science

Subject: Information Storage & Retrieval

Name of the lecture: Indexing and Searching Languages

Dr. Arwa Z. Nasser

Indexing and Searching Languages:

An **indexing language** can be defined as: the terms or codes that might be used as access points in an index.

A **searching language** can be defined as: the terms that are used by a searcher when specifying a search requirement. If the terms or codes are assigned by an indexer when a database is created, then the indexing language is used in index. The same terms or codes may also be used as access points to records during searching.

Indexing languages may be of three different types:

1. **Controlled Indexing Language**: may be used for names and other labels. But much emphasis is placed upon languages with terms that describe subjects.

There are **two types** of subject based controlled indexing languages: 1. **alphabetical indexing languages** 2. **classification schemes**.

In alphabetical indexing languages, such as descriptors are recorded in thesauri and subject heading lists. the subject terms are the alphabetical names of subjects, controls exercised over which terms are used, and relationships between terms are indicated but the terms themselves are ordinary words. **In classification schemes each subject is represented by a code or notation.**

2. **Natural Indexing Languages** means: **that you can search the familiar terms that the authors originally used.** Thus author indexes, title indexes and citation indexes, As well as natural language subject indexes, are derived- term systems. Natural indexing can be executed by a human indexer, or automatically by the computer.

Natural language indexing and controlled language indexing are used extensively in many information-retrieval applications. Both are used in retrieval on CD-ROM, via the online search services, in document management systems, and in online public access catalogues (OPAC).

3. **Free Indexing Term**: is by no means new. Its fundamental methods is to pick out key words (or sometimes) key phrases from the textual material by human intuition.

The choice of the word (intuition) is purposeful. No one really knows exactly what goes on in this selection process. The human indexers scan the material and underline or (otherwise indicate) those terms felt to be important or indicative of the subject matter contained in the textual material. It should be obvious, but apparently is not so that computers can not accomplish this task in the strict sense. One can not program computers intuitively to determine importance or relevance of terms or phrases in textual material, however while in theory human indexers can accomplish such indexing. In practice there is too much inconsistency among the human indexers to produce indexing useful for retrospective searchers.

Comparing Uncontrolled and Controlled Indexing Language

1. **Uncontrolled indexing languages** (advantages and disadvantages):

A) Advantages:

1. Low cost 2. Simplified searching. 3. Full database contents searchable. 4. Every word has equal retrieval value. 5. No human indexing errors. 6. No delay in an incorporating new terms.

B) Disadvantages:

1. Greater burden on searcher. 2. Information implicitly but not overtly included in text may be missed. 3. Absence of specific to generic linkage. 4. Vocabulary of discipline must be known.

2. **Controlled vocabulary or controlled indexing languages**

A) Advantages:

1. Solves many semantic problems. 2. Permits generic relationships to be identified. 3. Maps areas of knowledge.

B) Disadvantages:

1. Height cost.
2. Possible inadequacies of coverage.
3. Human errors.
4. Possibility of out-date vocabulary.
5. Difficulty of systematically incorporating all relevant relationships between terms.