

University of Nebraska - Lincoln

DigitalCommons@University of Nebraska - Lincoln

---

Library Philosophy and Practice (e-journal)

Libraries at University of Nebraska-Lincoln

---

March 2010

## Automated Indexing: The Key to Information Retrieval in the 21st Century

Tony I. Obaseki

Ambrose Alli University, Ekpoma, Nigeria, aligaji4@yahoo.com

Follow this and additional works at: <https://digitalcommons.unl.edu/libphilprac>



Part of the [Library and Information Science Commons](#)

---

Obaseki, Tony I., "Automated Indexing: The Key to Information Retrieval in the 21st Century" (2010).

*Library Philosophy and Practice (e-journal)*. 338.

<https://digitalcommons.unl.edu/libphilprac/338>

## **Automated Indexing: The Key to Information Retrieval in the 21st Century**

Tony I. Obaseki  
Osiwa Library Internet Center  
Library  
Ambrose Alli University  
Ekpoma, Nigeria

### **Introduction**

Global changes in physical infrastructure, population, technological development, and climate have contributed to an information explosion. This is a major challenge to information managers, who are faced, not only with the challenge of selecting, acquiring, and storing the information, with the perennial problem of how to make it available to potential users quickly and easily.

The world is shifting from manual to automated practices. Information centers are following suit, paving way for automated acquisition, processing, and dissemination of information to clientele. Indexing services may be the solution to providing current and reliable information to information seekers.

Indexing connotes the processes of creating an index. It is derived from the Latin root “indicare,” to point or indicate (Chakraborty and Chakrabarti 1983). Its current meaning has hardly changed from the initial meaning embedded in the root. An index is a means to an end and not the end itself.

An index could be said to be as old as human existence. It has been used to encompass not only an index at the back of a book, but also calendars, tables catalogues, inventories, registers, summaries, and syllabi. Witty (1978) studied the evolution of indexing practices through the examination of indexes in two fourteenth century manuscripts, an incunabulum and a book printed in the sixteenth century. He discovered that early indexing practice created entries entered roughly alphabetically with key words in the index arranged alphabetically by the three initial letters.

### **The Case for Automated Indexing**

Seth (2004) argues for the use of automated indexing because it is faster and cheaper. Seth asserts that this is one way of achieving the goals of information centers. This view is welcomed by numerous scholars, because automated indexing can deal with the increasing amount of new material being produced that has made manual indexing slow and expensive.

Automated indexing is the process of assigning and arranging index terms for natural language without human intervention (Tulic 2005). The index is produced using algorithms. These algorithms works on database containing document representations, including full text or bibliographic records, but also on non-text data bases such as images or music.

Automated indexing is not new. Its earliest known forms include:

- KWIC (key word in context)
- KWAC (key word at context)
- KWOC (key word on context)

These are falling out of use and being replaced by new methods such as:

- Pragmatic
- Historicist
- Rationalistic
- Classical IR
- Neutral network
- Probalistic
- Semantic text
- Vector space models

Some of these techniques or approaches are fully automated, while others are semi-automated or machine-aided. Problems with automated indexing include:

- Low precision, which arises from the fact that automated indexing does not check for rare expressions but indexes common words or phrases.
- Low recall due to the presence of synonyms.
- Generic results, which arises from authors' use of too broad or too narrow terms (Hjorland 2007).

Tulic (2005) lends support to human indexing, saying that, "to date, no one has found a way to provide computer programs with the judgment, expertise, intelligence, or audience awareness that is needed to create useable indexes. Until they do automated indexing will remain a pipe dream".

This paper supports the use of automated indexing. Indexing is a pivotal tool in information management and it must be automated like other library processes. Anderson and Perez-Carballo (2001) state that, "the bottom line is clear; automated indexing works! And it appears to work just as well as human indexing but differently." The automation of libraries has progressed so far, in Africa and elsewhere, that no library with an array of electronic resources could still rely fully on manual indexing.

### **Attributes of Automated Indexing**

**Faster and easier to produce.** This claim is relative to the process of producing an index manually. Automatically indexing an information source will save time at same time be faster, since most of the rigorous human effort would now be performed by a machine. There is software such as Microsoft Word and Adobe Framemaker that do automatic indexing of documents. These packages simplify and speed up the process, alphabetizing and assigning page numbers. Repetition of index terms is minimal. Software used for this function are programmed to locate and in delete terms that are repeated.

**Easily modified.** Automated indexing is easily retrievable, revisited, and modified when errors are noticed or due to future developments. This is an obvious advantage over manual indexing.

**Transferability.** ICTs have turned the universe into an information global village. Automated indexing permits information centers to share the information resources available at a particular center. This includes indexes and indexing.

### **Hindrances**

While automated indexing provides a means to cope efficiently with a flood of information, there are things that hinder its progress. The lack of technological know-how of the librarian indexer affects the process of automated indexing. Some LIS professionals still shy away from the use of ICTs. Some are reluctant to get the necessary training, and even when trained they still feel apprehensive about using ICTs. Most information centers train their staffs in the use of ICTs while awaiting the actual acquisition of such ICTs, especially at the planning stage of automation and computerization. The skills acquired become stagnant and rusty.

Some LIS professionals are not interested in indexing or aware of its importance, both manual and automated. The indexing process is rigorous and intellectual. The demands of indexing inhibit some information workers from learning and doing it, even with the importance attached to the index as a primary information retrieval tool. Oyinloye (2000) asserts that "most LIS professionals see indexing as a part-time occupation rather than as a profession in its own right."

Some librarians have questioned the quality of automated indexes, feeling that no software could match the intellectual needs of indexing. This view can be seen from my as that of the LIS professional lacks the patience to work with ICTs, and at same time tries to transfer the annoyance from the lack of ICTs and software to that very same unavailable and unacquired software. Anderson and Perez-Carballo (2001) assert that, "the bottom line is clear: automated indexing works! And it appears to work just as well as human indexing, just differently."

### **Automated Indexing: A Way Forward**

For automated indexing to be fully implemented, there is need for LIS professionals to be constantly trained and re-trained in technology and its application to information resource handling, both before and after computerization has taken place. Almost two decades into this century, many LIS professionals have not come into contact with important new technology. LIS professional education programs must be overhauled to accommodate innovations in line with technological developments. Teaching on the concepts of and approaches to indexing must be re-defined. The notion that an index is what one sees at the back of a published book still holds sway. A new introduction to indexing will not only will bring out another dimension to the profession, but also give a place of pride to indexing as a process to be carefully and quickly accomplished with the aid of computer technology.

### **Conclusion**

In the present era, there is a need to use technology to quickly control and give access to the influx of information sources. This makes automated indexing a necessity. Patrons who need information need it quickly. The world is moving with speed, and information centers and LIS professionals must move along with it, in the shift from manual to fully automated services.

### **References**

Anderson, J.D., & Perez-Carballo, J. (2001). The nature of indexing: How humans and machines analyze messages and texts for retrieval. Part 1&2. (Research and the nature of human indexing) *Information processing and management* 37 (2): 231-277.

Chakraborty, A.R., & Chakrabarti, B. (1983). *INDEXING: Principles, process, and products* . Calcutta: World Press Private Limited.

Hjorland, B. (2007). Semantics and knowledge organization. *Annual Review of Information Science and Technology* 41: 367-405.

Oyinloye, A.M. (2000). The Nigerian experience: Indexes, indexers, and indexing. *Indexer* 22: 78-8

Seth, M.A. (2004). Notes on automatic indexing. Available:  
<http://taxonomist.tripod.com/indexing/autoindex.html>.

Tulic, M. (2005). Automatic indexing. Available: [www.anindexer.com](http://www.anindexer.com)

Witty, F.J. (1965). Early indexing techniques: A study of several book indexes of the 14 th, 15 th, and early 16th centuries. *Library Quarterly* 35 (3): 141-148.