

What are the main differences between human indexing and automatic indexing?

There has been a debate for a number of years about which method is "better" – human indexing or automatic (machine) indexing. To weigh in, it would be helpful to know what each type of indexing involves and their major differences, as well as what is meant by indexing itself.

Anderson and Perez say their definition of indexing simply means pointing to or indicating the content, meaning, purpose, and features of messages, texts, and documents (2001a, p. 233). So why should we care? If you think about the amount of information that we have available through electronic resources alone – databases and the Internet – then you realize the importance of finding access points to the needed information. As Tenopir succinctly writes, "[I]ndexing improves search results" (p. 34).

Human indexing is done by –big surprise– humans. Human indexers use their knowledge to find the "aboutness" of the writing they are analyzing. They can find concepts within the writing and then use terms to help the searcher connect to that writing. Human indexing tends to focus on larger documentary units, such as complete periodical articles, complete chapters in collections, or even complete monographs (Anderson & Perez, 2001a, p. 236). Human indexers, however, can have their personal biases or experiences influence their work. Fidel also says that human indexers are inconsistent.

Automatic indexing often refers to indexing done by computer algorithms. Obviously, humans are involved with creating the programs for the computers, and in setting the parameters, but the work is done by computers. Fidel highlights the contradiction of automatic indexing. On the one hand, she says, it's the most user-centered approach because of its dynamic, helpful, and flexible nature. On the other hand, indexing is based solely on the text stored and is completely immune to the particular group of users and their queries (p. 575). Browne states that there is a "continuum of use of computers, from no computer at all to fully automatic indexing." The importance of this is to confirm that human indexing often uses the computer as a tool (clerical support/data entry, quality control, intellectual assistance), but that these uses of the computer is not the same as automatic indexing when the computer extracts the terms to index. Cleveland and Cleveland break automatic indexing into four approaches: statistical; syntactical; semantic systems; and knowledge-based. Each approach looks at the material being indexed in a different way.

What are the advantages and disadvantages of automatic indexing over human indexing?

Using information from many of the articles referenced, and especially from Anderson and Perez, the following table attempts to show the strengths and weaknesses between human indexing and automatic indexing when looking at certain variables. Anderson and Perez (2001a) say that in many of these categories, because humans *could* do what automatic indexing does, they are not necessarily an attribute unique to automatic indexing.

Variables	Advantages or Disadvantages of Types of Indexing
Cost	<p>Human Indexing expensive per unit indexed because it is labor-intensive Automatic Indexing inexpensive per unit indexed</p> <p><u>Who wins ?</u> Depends on what you are seeking</p>
Time	<p>Human Indexing involves more time per unit indexed Automatic Indexing can index large amounts of material in short amount of time</p> <p><u>Who wins ?</u> Once again, it depends on what the needs are</p>
Extent of indexable matter	<p>Human Indexing may be limited to abstract or summarization of text Automatic Indexing routinely based on complete text</p> <p><u>Who wins ?</u> Automatic indexing can index more of the indexable matter</p>
Exhaustivity	<p>Human Indexing tends to be more selective Automatic Indexing considers most of the words in indexable material</p> <p><u>Who wins ?</u> Automatic indexing is by nature more exhaustive</p>
Specificity	<p>Human Indexing uses more generic terminology, smaller vocabulary Automatic Indexing uses very specific terminology, larger vocabulary</p> <p><u>Who wins ?</u> Automatic indexing has a higher specificity</p>
Browsable displayed indexes	<p>Human Indexing use multi-term context-providing headings Automatic Indexing limited use of term combinations</p> <p><u>Who wins ?</u> Depends on what you are seeking</p>
Searching syntax, display syntax	<p>Human Indexing use wide-range of syntactic patterns and can adapt quickly to include newer terminology, as well as older subject headings Automatic Indexing becoming more sophisticated, and is selecting, combining, manipulating, and weighing terms. Usually limited to key-words in, out of, or along-side context.</p> <p><u>Who wins ?</u> Human indexing has an advantage here</p>
Vocabulary management	<p>Human Indexing can cross-reference, link synonyms or like terms, point to related terms easily Automatic Indexing being experimented with</p> <p><u>Who wins ?</u> Human indexing currently has better vocabulary management</p>
Surrogation	<p>Human Indexing not often used by human indexing Automatic Indexing being used frequently often as visual displays, such as icons or graphs</p> <p><u>Who wins ?</u> Automatic Indexing often uses surrogation, while human indexing does not</p>

If one was going merely from the variables in the chart, the assumption would be that automatic indexing should be the logical choice in the debate. However, there is other information to consider.

What are the main issues of automatic indexing?

When looking at the table, it seems as if automatic indexing has a lot of advantages over human indexing. Diakoff, in his article about database indexing, had this to say:

The sophistication with which the categorization of text can be automated can only increase. With a reliability that no human indexer can match, automatic techniques will eventually rival the best that humans can provide. But this remains well in the future. At present automatic techniques can only provide either a conveniently rapid but very rough categorization of heterogeneous materials, or a precise categorization of thoroughly predictable and homogeneous materials. This will change, but not tomorrow. (p. 92)

Automatic indexing, therefore, is:

- predictable
- becoming more sophisticated
- good for material that is similar (homogeneous)
- less expensive
- able to extract terms, as well as use clustering techniques
- helps searchers find information

However, automatic indexing is not:

- flexible
- precise when looking at unique material
- able to adapt quickly to new terminology
- able to do more than rough conceptualization
- helping searchers find the same information they would have found with human indexing

Improvements are being continually made, and experiments with automatic indexing continue. The importance of searching electronically and finding consistent results is being studied by experts, and the research on searching will impact automatic indexing.

Indexing Revolution or Evolution?

Does this mean that we are in an indexing revolution or simply is it indexing evolution? I believe it is both. By *indexing revolution*, I draw parallels with the industrial revolution and to the way humans have changed the way items are created. With the industrial revolution, items that were previously created by hand were suddenly being created more quickly and cheaply using humans aided by machines. Now in most industries, machines do a predominant amount of the work, and are guided or monitored by humans. The same is happening to indexing.

In a technologically-developed society like ours, we need the inexpensive automatic indexing to complete the vast and growing amount of indexing for electronic resources. In terms of clothing, this is our trip to Wal-Mart and the mall where we find shirts, pants, sweaters, and the like at

very affordable prices. The cost and amount of time to make those clothes has been reduced, and the savings are passed on to the consumer. However, just as people will pay more money for an item that is hand-crafted and unique, human indexing is still available and appreciated. When items (whether clothing, furniture, or indexing) aren't mass-produced or done as quickly, there is a craftsmanship that is superior to those done by machines.

Perhaps the most interesting aspect of the human versus automatic indexing debate is what the research is showing. Anderson and Perez say the research tends to show that the two approaches produce different results but that users find them, on balance, more or less equally effective (2001a, p. 233). This is amazing. Users are getting different results from their searches where human indexing has been used or where automatic indexing has been used. However, the users *don't seem to be bothered by those differences*.

We will find ways to improve automatic indexing, because it helps resolve the problem of the increasing quantity of indexing to be done. Human indexing will remain as well, although the numbers of people doing it for a living will decrease. I believe the two types of indexing will co-exist since each meets a unique need.

References

- Anderson, J.D., & Perez-Carballo, J. (2001a). The nature of indexing: How humans and machines analyze messages and texts for retrieval. Part I: Research, and the nature of human indexing. *Information Processing and Management*, 37(2), p. 231-254.
- Anderson, J.D., & Perez-Carballo, J. (2001b). The nature of indexing: How humans and machines analyze messages and texts for retrieval. Part II: Machine indexing, and the allocation of human versus machine effort. *Information Processing and Management*, 37(2), p. 255-277.
- Browne, G. (1996). Automatic indexing. *Australian and New Zealand Society of Indexers*. Retrieved April 11, 2005, from <http://www.aussi.org/conferences/papers/browneg.htm>
- Cleveland, D.B., & Cleveland, A.D. (2001). *Introduction to Indexing and Abstracting*, 3rd ed. Greenwood Village, Colorado: Libraries Unlimited.
- Diakoff, H. (2004, October). Database indexing: Yesterday and today. *The Indexer*, 24(2), p.85-96.
- Fidel, R. (1994). User-centered indexing. *Journal of the American Society for Information Science*, 45(8), p. 572-576.
- Tenopir, C. (1999, November 1). Human or automated, indexing is important. *Library Journal*, 124(18), p. 34-38.