Preface

The value of a body of knowledge depends very much on whether it is accessible; more precisely, on how well it is accessible. One potentially valuable tool are indices. Here is a (subject and author) index for volumes 1(1970)–89(1979) of the journal *Artificial Intelligence*.

The idea of trying to put together such an index originated with the publishers of the journal and its editorial board. Actually, two indices were made. The current one, for which the powerful facilities of the Macintosh wordprocessing program Nisus (version 4.06) were used, and a second one using the dedicated index making program TExtract™. Both came, surprisingly, to about the same size, some 18000 entries and about twice that number of references; otherwise they are rather different. Both will be available through the Web, and a merged version of the two will also be made available to the community in that way.

The numbers behind the key phrases in the index itself refer to the 1221 articles that appeared in these eighty-nine volumes. They are numbered more or less in historical order. Just about any user will notice that there are gaps in the number sequence; that does not reflect missing articles or other incompleteness aspects; it is simply a reflection of a change in the way numbers were and are assigned. This issue also contains the thus numbered list of these articles giving author(s), title, volume number, and page numbers.

The present effort should be seen, I feel, as no more than a first step in the direction of creating an information finding tool for the knowledge sciences; a first offering to the community that hopefully will lead to further efforts. It is absolutely clear that a comprehensive index (or, better, an enriched thesaurus) for a given field is a most valuable thing to have available. "Comprehensive" (up to a given level of detail) is a key property here.

Some insight into the comprehensiveness of the present effort can be obtained as follows. Imagine that the field in question consists at a given time of a collection, A, of accepted key phrases and key words. In addition there is an infinite universe, B, of other potential key phrases and words which authors and scientists can use to describe their new ideas, algorithms, concepts, etc. Indexing starts at a given point in time; for convenience think of it as being done in historical order. At each stage a random number a of terms from A is selected as well as a random number b from the universe B. These latter are added to the collection A, which thus grows in time. Let the random variables

a and b have means $\bar{a}$ and $\bar{b}$ (and, say, a binomial distribution). Then, as indexing proceeds, one expects in the beginning just about linear average growth (of slope $\bar{a}+\bar{b}$) in the number of different index terms selected, followed by a saturation period during which the terms from the original collection gradually all get selected, followed by linear growth of slope $\bar{b}$. This is, of course, a quite simple model but it can maybe serve as a first indication. I kept count of the number of terms selected. The graph of these numbers of terms in the indexing process for the present index is just about a perfect straight line, indicating that the saturation regime has not yet been reached. What this means I leave to the experts to decide. Perhaps the field is growing at such a furious rate that even a leading journal like the present one cannot expect to pick up most of the significant terms in almost thirty years.

Using the index is straightforward. Little effort has been made to do any systematic "stemming". Thus, for instance, singular and plural forms of a noun can both occur (in various combinations). If this happened right next to each other nothing has been done. If the two (groups) did not occur next to one another a "See also" remark has been inserted. Thus, to cite another type of occurrence, at the end of the list of terms headed by "non-monotonic" there is a note '[See also "nonmonotonic"]'.

There are other efforts under way of making indices and related tools in the knowledge and computer sciences. Two years ago a subject index and cumulative index was put together for volumes 101–150 of the journal *Theoretical Computer Science* (Volume 150:2 (1995), pp. 195–376). A complete one for volumes 1–150 is in preparation. Still more plans exist and in such a way a kind of atlas covering these fields can be created giving rise to the very nontrivial mathematical problem of describing the overlaps and correspondences of the charts of the atlas (which are the individual specialized indexes which can be of varying levels of detail and of differing kinds, as is indeed the case in the atlas metaphore).

As already said, the present index is but a beginning, but perhaps it can be hoped that it will stimulate a variety of efforts on the part of the AI community. And, of course there are AI and automation aspects to the indexing problem itself introducing just that whiff of self-reference that could make things even more interesting.

Michiel Hazewinkel
Bussum, 1 September 1997