ELSEVIER

Discrete Applied Mathematics 106 (2000) 1–2

DISCRETE
APPLIED
MATHEMATICS

# Preface

The value of a body of knowledge depends very much on whether it is accessible; more precisely, on how well it is accessible. Very roughly the value of about a hunderd volumes of the journal Discrete Applied Mathematics (DAM), representing some two thousand articles, is practically zero if one has no way to selectively find information in it. One potentially valuable tool is a comprehensive index. Here is a (subject and author) index for volumes 1–92 of the journal DAM, in which precisely 2009 articles appeared.

A companion index for the first 200 or so volumes of the companion journal Discrete Mathematics is in progress. Another one for Linear Algebra and its Applications is planned. There already exist similar indices for the first 200 volumes of Theoretical Computer Science (some 42000 terms) [1] and the first 89 volumes of Artificial Intelligence (some 18000 terms) [2]. These are the first steps in an attempt to build up an adequate (standardized) *phrase vocabulary* for mathematics and computer science. A dictionary of words is of but very limited use for information retrieval purposes in science (as can already be guessed from the numbers just quoted).

The numbers behind the key phrases in the index itself refer to the 2009 articles that have appeared in these 92 volumes. They are numbered in more or less historical order. This volume also contains the thus numbered list of these articles giving author(s), title, volume number, and page numbers. There are gaps in the numbering sequence; that is just the result of the procedure that assigns numbers to articles and does not mean that there are any articles missing; the list is complete.

The index was generated from titles and abstracts only (and key words, in the low percentage of cases that these were available; these author supplied key phrases were usually far from adequate to describe the article in question).

There are 11734 different 'index phrases' and 20429 'citations' (= reference numbers), for an average of about 2 citations average per index phrase. That is reasonably high for this type of index which is in first instance for human use (as opposed to computer use). Thus meaningful single words, like 'algorithm' and 'bound' which occur hopelessly frequently (resp. 1884 and 851 times) in the material at hand, have been left out; but have been retained of course in those compound phrases of which they form a part.

The index is 'reference-complete' in the following sense: if a phrase occurs in the index than all occurences are listed, taking account of such obvious linguistic variations as occurred, unless the phrase was part of a larger index phrase. This last 'unless clause'

[1] Theoretical Computer Science **213/214** (1999) 1–659.

[2] Artificial Intelligence **96** (1) (1997) 1–302.

Elsevier Science B.V.

accounts for the fact that the index phrase (index *word* in this case) 'graph' has 385 citations on its own while it occurs some 3100 times in total (as noun, in singular or plural) in a compound phrase. These all occur in the index. The word graph is the record holder as regards number of citations in this index; the next one is 'edge' (273); there are only rather few more which have largish numbers of citations.

I also believe that this index is pretty much complete as regards meaningful phrases from Discrete Applied Mathematics in so far as they occurred in the material available. One way to try to judge completeness of an index for a given field is on the basis of a simple stochastic model that is briefly explained in [3]. The saturation phenomenon that that model predicts (for a currently complete index; most scientific fields continue to grow) are clearly starting to show in the case of the present index, suggesting that this one goes some way towards completeness. How far is hard to say. Precise statistical tools for judging such matters are being developed [4].

Let me make a few remarks on the lexicographic ordering that has been used. Basically, that one is: numerals first, then funny symbols like %, &, #, \, etc., (space) and (tab), punctuation signs and quotation marks, various kinds of brackets and slashes, and then the alphabet itself; (space) comes before (hyphen). Upper case and lower case letters and letters with diacriticals are basically folded together; more precisely, by way of an illustrative example, part of the ordering sequence is: c ç D d E É e é è ê ë F f and another part is A Á À Ä Ã Å a á à â ä ã å Æ æ B b. Mathematical symbols are usually largely ignored in an index like this (partially because the various fonts that are available order very differently). In the present case practically none occur except $\infty$ and a few Greek letters. The glyph $\infty$ falls somewhere between 9 and A, and the Greek letters are treated as if they were written out: thus $\alpha$ is treated like alpha, $\varphi$ like phi, etc.

A reasonable amount of effort has gone in identifying and lumping together linguistic variations and spelling differences. Thus of singular and plural versions just one occurs in the index with the singular preferred if that made sense. If both American and English spellings occur that also has been homogenised with a preference for the English orthography (if present) and a *see* reference for the key words involved; thus, for example: 'colorable, *see* colourable'. Hyphenated versions and 'written together versions' have been treated similarly. For instance 'fixpoint', 'fixed point', 'fixed-point', 'fix-point', with, in this case, 'fixed point' preferred. However, there certainly remain quite a few instances of phrases that occur separately, are further apart linguistically, but really have the same mathematical meaning.

Like the various other indexes already mentioned, this one is a first attempt, a gesture to the community, to be used, criticized, improved, and enlarged.

Michiel Hazewinkel
<mich@cwi.nl>
Bussum, 19 November 1999

[3] Michiel Hazewinkel, Topologies and metrics on information spaces, CWI Quarterly **12**:2 (1999). Preliminary version: http://www.mathematik.uni-osnabrueck.de/projects/workshop97/proc.html.

[4] Institute of Mathematics and Informatics, Lithuanian Academy of Sciences, Vilnius, Lithuania.