ELSEVIER

Discrete Mathematics 227/228 (2001) 1–4

DISCRETE 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 two hunderd volumes of the journal Discrete Mathematics (DISC), representing over six thousand articles, is practically zero if it is just standing on your shelves in the library, and 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 1–200 of the journal DISC, in which precisely 6032 articles appeared.

A companion index for the first 91 volumes of the companion journal Discrete Applied Mathematics has already been published [1]. There will also be a combined version of these two, but only electronically. Other similar indexes for Information Processing Letters, Journal of Logic Programming and for Linear Algebra and its Applications are in progress or being planned. There already exist indexes for the first 200 volumes of Theoretical Computer Science (some 42000 terms) [2] and the first 98 volumes of Artificial Intelligence (some 18000 terms) [3]. These are the first steps in an attempt to built 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 6034 articles that have appeared in these 200 volumes (there are two empty numbers). They are numbered in more or less historical order. This double volume also contains the thus numbered list of these articles giving author(s), titles, volume numbers, and page numbers. There are two 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 some 30000 different 'index phrases' and close to 60000 'citations' (= reference numbers), for an average of about 2 citations average per index phrase. That is

[1] Discrete Applied Mathematics **106** (2000) 1–261.
[2] Theoretical Computer Science **213/214** (2000) 1–660.
[3] Artificial Intelligence **96** (1997) 1–302.

Elsevier Science B.V.

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 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' accounts for the fact that the index phrase (index *word* in this case) 'graph' has 1793 citations on its own while it occurs some 3513 times in total (as a 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.

I also believe that this index is pretty much complete as regards meaningful phrases from Discrete 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 [4]. 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 [5].

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 alphabeth 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. This means, by way of example, that 'Fan' comes before 'fan' but that 'fan' comes before 'Fan condition'. E.g.

Fan, 2374, 3626, 3693, 4943, 5042, 5294
fan, 3747, 4453
fan argument, 5158
Fan condition, 5906
Fan condition for Hamiltonicity, 3693
fan sequence, 1759
Fan type condition, 3838

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 ∞ and a few Greek letters. The glyph ∞ falls somewhere between 9 and A, and the Greek letters are treated as if they were written out: thus α is treated like alpha, ϕ like phi, etc.

When looking for something like '*P-complete', look just after 'P-complete'.

[4] Michiel Hazewinkel, Topologies and metrics on information spaces, CWI Quarterly **12**:2 (1999) 93–100.

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

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.

Things are different, however, with quantified nouns and adjectives like 'n-set' (= a set of size n). Consider for instance the two key phrases

'family of n subsets' = (a family of size n) of subsets of a set
'family of n-subsets' = a family of subsets such that each of those subsets is of size n.

Quite generally quantified nouns and adjectives occur very frequently in discrete mathematics and combinatorics. Moreover for generic (symbol based) quantifiers there is very little in the way of standards. For instance 'd-dimensional', 'n-dimensional', 'k-dimensional' all occur. For these a standardization has been introduced: n, k, m, in that order (if more than one occurs in the same key phrase). For example:

n-cycle free
n-connected
n-regular
n-edge connected
(n, k)-colouring

There are exceptions where a more or less standard notation exists or where the meaning is not that of some kind of numerical quantification, such as

(v, k, λ)-design
t-design
f-domination number (see 5655)
k-algebra (here k is a field)
τ-design
p-connected graph (see 5801)
P-domination (see 5805)
A-optimal (see 5860 a.o.)
D-optimal
L-coloring (see 5877)
h-vector
p-algebra
p-group
Y-dominating

However, if other variants, then these standardized ones, actually occurred in the material at hand a *see* reference has been inserted; e.g. k-connected, *see* n-connected. For numerical quantifiers like '2-node connected', respectively 'two-node connected', this has been done both ways with a '*see also*' reference.

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

Michiel Hazewinkel
<mich@cwi.nl>
Bussum, 1 October 2000