Valentina M.Efremenkova
Nikolai Milovidov
VINITI

The Data Structure of the VINITI "Computer Networks" Information File

Dr.Valentina M.Efremenkova (b.1939) graduated from the Faculty of Physics of Moscow University in 1965. Her doctorate is in solid-state-physics. She heads the VINITI group on automation technology, including the introduction of the Rubricator.

Efremenkova, V.M., Milovidon, N.: The data structure of the VINITI "Computer Networks" information file.
Knowl.Org. 20(1993)No.4, p. 208-209
Presents an attempt of scientometric analysis of information, which is contained in databases produced by the All-Russian Institute of Scientific and Technical Information (VINITI). The "computer networks" area is analyzed and a new classification structure is described. (Authors)

Dr.Nikolai Milovidov (b.1946) graduated from the Moscow Physical-Technological Institute in 1969. Doctoral degree in systems analysis, control and computer science 1978. Heads the VINITI Dept.of Computer Science and is editor of two VINITI abstract journals (on "Computer Science and Computer Systems" and on "Software").

The main aim of scientometry is the acquiring of qualitative information on different fields of science and technology on the basis of analysis of the flow of scientific publications without reading their contents. Such specific features of documents as: classification system numbers, terminology, authors of articles, names of organizations or firms, factographical information, nature of the work (theoretical or experimental), date of issue, etc., can be used for assessing the state of the art and the prospects of development in specific subject fields.

The annual information file in the multi-theme databases on the subject "Computer Networks" contains 1.500-2.000 documents. The main flow of publications, which represents the information model of the examined field and adequately reflects it, is concentrated in the following VINITI databases: "Computers and systems", and "Software" about 50%, "Communication networks. Telephone network. Transfer of data" - about 30%, "Automatics and teleautomatics" - about 5%. Publications selected from other VINITI databases characterize the areas of application of computer networks and make up about 15%. In the INSPEC database in Great Britain the documentation flow is divided between two databases: the INSPEC-C "Computer and Control" and INSPEC-B "Electricity and Electronics". The division of the flow of publications into computer technology and communication formed historically and has found reflection in different classification systems of the databases. In the VINITI database, the themes examined have the following classification numbers:

In computer technology:
502.05.15   Computer networks, theory and problems
502.39   Computer networks

502.41.23   Software for computer networks

In communication:
493.03.03   Theory of transmission of information through communication channels
494.03.07   Theory of structures of information networks
493.37.29.31 Computer applications in data transmission networks

In the INSPEC database the following codes are used:

C 56.00   Data communication equipment and techniques
C 56.20 L   Local area networks
C 56.20 M   Metropolitan area networks
C 56.20 W   Other networks
C 56.30   Networking equipment
C 56.40   Protocols
C 56.70   Network performance
C 56.90   Other data communication equipment and techniques.

INSPEC-B
62.10 L   Computer communications.

The growth in the number of publications in separate classification codes and the appearance of new headings in the rubricators speak of active work going on in the field under review. The number of periodical and serial publications in the last seven years (1985-1992) has increased 1.7 times. In this period, five new classification codes have appeared in both databases.

A statistical analysis according to form of publication revealed that articles from periodicals and serials made up approx. 60%, articles from transactions and proceedings 20%, patent documents approx. 15%. Original sources come from 25 countries and in 13 languages. The main

producers are the USA (approx. 35%), Germany (approx. 20%), Japan (15%), and Russia (15%). Approx. 60% of the documents are published in English, in Russian approx. 15%, in German approx. 10%, and in Japanese approx. 8%.

The distribution of publications according to subjects is as follows:
- Territorial WAN (wide-area networks)
  (numbers are all approx.)
  global networks 10%
  regional 6%
  local 60%
  institutional (incl.private and public) 40%
- Networks of data transmission
  computer networks with commutation of packages,
  messages, channels, integral service networks 50%
(The share of publications dealing with methods of physical transmission of data - analog or digital - is less than 2%, and with the logical aspect of interconnection is about 1%)
- Specialized, universal, experimental computer networks, uni
  ted in accordance with the functional characteristic - 1%
- homogenous and heterogenous hardware - 1%
- Computer networks: ring, bus, tree-like, distinguished accor-
  ding to topoloy - 1%.

The qualitative analysis of the trends in document accumulation of the different classes reflects the development of research dealing with computer networks. At first networks were created by independent firms and were intended for the processing and storing of information. Subsequent (beginning with the mid 80s) organizations, which ensured the transmission of information, began to actively study and create computer networks. This development is reflected by the stormy increase of the flow of publications dealing with the transmission of data and local networks.

Inna Gendlina
VINITI

Computerized Formation of Alphabetical Subject Indexes to Classification Systems

In 1991-1992 VINITI's experts worked on the compilation of algorithms and programs of automated formation of an alphabetical subject index (ASI) to classification tables. The ASI to the State Rubricator of STI (the former GASN-TI Rubricator) was prepared with the help of the EX-1045 computer, and in its structure there is a KWOC index (permuted-title index of keywords out-of-context).

The main element of this structure, i.e. the entry, consists of the title word denominating the concept, and of the denominations following it in alphabetical order (i.e. full texts) of the classification divisions in which this word apprears, with the appropriate codes. On the basis of the machine file of the ASI, built up in this way, proof-sheets for the 4th edition of the State Rubricator of STI, put out in 1992, were prepared.

The package of programs developed by the compilers makes it possible to pick out the terms from the context of formulas of classification divisions, normalize the selected terms according to their number and case, permute the terms of word combinations and draw up an alphabetic list of records and solutions in order to reduce the volume of the ASI without loosing its informative capacity. This software can, after some modifications, be used for the compilation of an ASI to any classification systems.