

UNITED NATIONS GROUP OF EXPERTS
ON GEOGRAPHICAL NAMES
EIGHTH SESSION
New York, February 26 - March 9, 1979
Agenda item No. 10 (h)

Working paper No. 24

AN AUTOMATED INFORMATION RETRIEVAL SYSTEM
FOR GEOGRAPHICAL NAMES (TOPONYMS)^x

The demands for dictionaries, indexes and other kinds of information concerning geographical names are steadily increasing and a great deal of time and labour is required to meet it. The main way to speed up progress in this field and that of map production is the automation of manual work.

For this purpose an automated information retrieval system for geographical names or toponyms (AIPST^{xx}) has been worked out and put into operation in the Soviet Union. The system allows for automating a number of labour-consuming processes of accumulation, processing, storage and output of information on the names of geographical entities. The information on geographical names comprises the Russian standard (officially approved) form of a name, its original lettering (for non-Russian names), the designation of the kind of entity named, its geographic co-ordinates and/or its belonging to a certain political-administrative or natural unit.

^x This report prepared by A.G.Ivanov and A.M.Komkov was submitted by the Soviet delegation to the IXth International Cartographic Conference (Washington, 1978). The members of the UN WG on the automated data processing which attended the conference also acquainted with it. But the most participants of this UN GEGN session were not present at the conference. It seems reasonable therefore to present here the main points of the report.

^{xx} The abbreviation is based on the Russian initials.

/...

The system provides the following main forms of information service:

- preparation of reference documents of various kinds for the mass user: geographical name indexes to maps and atlases, reference and normative dictionaries of geographic names as to countries or regions, current information on changes of geographic names and the like;

- selective dissemination of information in accordance with the requests of permanent information users;

- retrieval and output of information upon requests of individual information users.

The AIPST is based on the principle of a single input of the results of analytic-synthetic processing of each place-name into the computer and a reiterative usage of the results of machine processing of information to meet various information needs.

The structure of the operating system in conformity with this principle is as follows (Fig. 1).

Initial data on the names of geographic entities collected from different sources undergoes analytic-synthetic processing which includes: the systematization of data by territorial and linguistic factors; the establishment of a correct standard form of each name in conformity with the strict rules for spelling Russian names and Russian rendering of non-Russian names; the cataloguing of established names in a continuously updating card file form. These operations are performed by a staff of specialists in toponymy, geography, cartography and linguistics.

A catalogue containing the full information on the established place-names serves as a basis for the AIPST. The information retrieval on inquiries concerning individual names is realized manually in the card files of the basic catalogue, as the employment of a computer in such cases is not profitable. But when an inquiry concerns numerous place-names the employment of an automatic system is rather advantageous in obtaining necessary information according to a given program.

To be introduced into the computer the initial data properly processed undergoes a secondary processing which consists in (a) indexing and coding the information in terms of specially developed information retrieval language and (b) transferring coded data to a punched tape. The devices employed for the purpose transform alphanumerical information into a digital form suitable for input into the computer.

The computer data processing is executed employing a 13 programme software that allows for an output of various kinds of documents differing in contents, volume of information and form, e.g. a geographical name index to a map, a dictionary of the names of all the geographic entities within the limits of a given region, a selected list of names as to kinds of features: cultural, orographic, hydrographic and the like.

Two types of units can be used to receive data output from the computer. One is the punch tape unit used for the data input. Its advantages: simplicity of exploitation and a relatively rapid output of express information for a small number of users in tabulation form. Its disadvantages: uncompactness of the documents received, a limited choice of types and obligatory photoreproduction if publication of documents is necessary.

The other output unit is the 2-NFA, an electronic photocomposing machine made in the USSR. Though more complicated this equipment is very effective as it provides for the reproduction of alphanumeric information in more than 100 languages since it has types of various sizes and forms plus all the necessary diacritics. A sorted name file can be directly introduced into the phototypesetter 2-NFA with the help of composing and programming apparatus "EPF" and photo error-correcting unit "FK".

Unsorted name files undergo preliminary computer sorting.

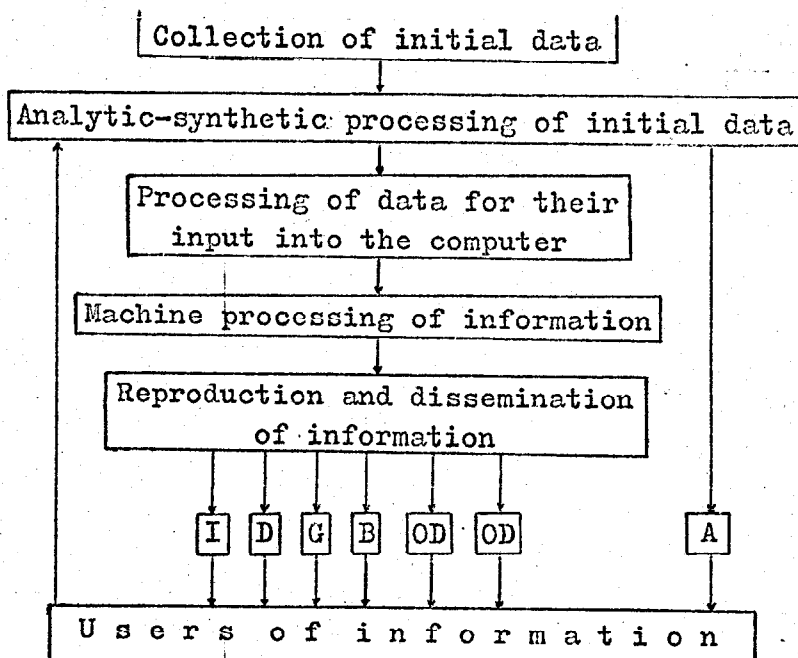
The employment of the 2NFA photocomposing machine as an output unit for the AIPST system enables to obtain high-quality original copies of information documents on a positive film.

/...

At present the AIPST system is rather effectively employed in preparing place-name dictionaries with Russian and non-Russian name-forms and indexes to some maps and atlases. The employment of the system allows for the compilation of a complete index of place-names to the World Map produced by the cartographic and geodetic services of Bulgaria, Czechoslovakia, GDR, Hungary, Poland, Romania and the USSR.

In conclusion it should be noted that the AIPST system allows for accumulating a data bank on machine carriers along with the preparation of information documents. In combination with the data bank of objects of mapping it will provide for the functioning of a universal automated cartographic system the elaboration of which is now in progress.

Fig. 1
Structure of the AIPST system



Abbreviations: I - indexes, D - dictionaries,
G - gazetteers, B - bulletins of current
information,
OD - other documents,
A - answers to individual inquiries