Data Bank on the U.S.A. and Soviet-American Relations

by Tatyana Yudina [1]
Senior Researcher, SPASIBO
Institute of the USA and Canada Studies USSR,
Academy of Sciences Moscow, USSR

INTRODUCTION

Our Institute - Institute of USA and Canada Studies of the USSR Academy of Sciences - was established in 1969. There are several departments in it; USA domestic policy, foreign policy, military policy, agriculture, management system, and others. Some 200 researchers and post graduates work in our institute.

Two years ago several dozen personal computers (IBM - AT and XT clones) were purchased and a new section was established - that of applied research and informatics. Invited to join the section were researchers interested in new technologies and new sources of information. I was one of 12 memebers to join this new group. We also have two professional full-time programmers and several part-time programmers working with us; not enough to meet our needs, however. In this we are not alone as I learned from my attendance of the 1990 IASSIST conference. I discovered that this is a common problem among data libraries throughout the world - understaffed with the personnel that are needed most of all.

Our section has several major functions:
- teaching researchers computer skills
- studying the software market for emerging applicationsthat may be rquired by our department an providing traing in software use
- designing the computer-based information-retrieval system for our institute and coordinating the efforts of different departments of our institute taking into account the strategic aim of our data base - to have as much data as possible,
- studying and evaluating new sources of information; CD-ROMs, on-line data-banks, archives of machine-readable data etc.

These are our main tasks to say nothing about educating our chiefs and encouraging them to give money for ongoing operations and development of our programs.

We were among the first of the humanitarian institutes to use personal computers, with limited access to consultants, and with no prior computing experience, we have made do through trial and error.

COMPUTER-BASED INFORMATION RETRIEVAL SYSTEM "USA AND SOVIET-AMERICAN RELATIONS

Our Institute is a section of the USA studies in the USSR, and our mandate is to serve as a national data bank on the USA and Soviet-American relations.

We are fortunate in that there is a great quantity and variaty of high quality data that the USA produces about itself. Our problem is one of finding funds for the purchase of the data we choose. Happily we have other responsibilties in addition to building data collections.

The system is multifunctional and multilingual, it contains data of various genres, original full texts or created especially for the system in Russian, in English and some in French. We hope that in future it will be supported by the automatic translation subsystem.

There are two blocks of information in our system - one about the USA and the other about the Soviet-American relations. Both blocks contain several types of data sets :
- the "USA reference module"; an electronic version of the USA Encyclopedia recently published by our institute,
- a statistical module consisting of long range time-series data on economic indicators in the USA,
- full-text documents, such as Soviet-American agreements from 1933 to present (updated on a fortnightly basis),
- a chronology of USA-USSR relations,
- articles and statistics about economic and trade relations between the two countries,
- mutual Soviet-American projects and joint ventures,
- biographies of the American political leaders, with a special module: "speeches" - short references of main ideas processed by our linguistic means (further they will be mentioned),
- archive collections,
- bibliography of publications about the USA and Soviet-American relations done by our library, which has started to create an on-line catalogue of its holdings.

In our data collection we have data sets on such subjects as: the Congress of the USA, President Bush's cabinet, White House personnel and the structure of the American administration, the constitution of USA, etc. These

subject data sets are being created by our researchers - we encourage such projects, since we simply can't afford to purchase such collections from the USA. In collectiong information our researchers make useof American online services such as Dialogue, Lexis-Nexis, among others, and which is funded by the Academy of Sciences.

WORK ORGANIZATION
In developing our information system we have introduced some new elements of work organization - our researchers are invited to build data collections in their field of expertice. We did this for several reasons, the first one is that we hope that a researcher - expert in some particular sphere knows the information market better, can evaluate different sources, and choose the most informative ones. He is also supposed to process the data, to administer and update the collection. We encourage our researchers to start such data collections, though there are some social and psychological problems to this in that researchers become possive of their data. We hold competitions amoung the researchers, with the winners going to attend the summer workshops at the University of Michigan or Essex University , where they learn about new methods of analysis and modern software for personal computers. The idea of data collections built by experts is that they use the data as a base for their analysis while others mostly use it for information.

We especially encourage such data sets that may be used for modelling.

NEW SOURCES OF INFORMATION
As for the on-line sources of data we have on-line communication with Soviet Press Agency which widely reports international news, and speeches of American political leaders. All the speeches of American political leaders which touche upon the problems of Soviet-American relations are supposed to be processed and added to our political "portraits" databases.

One of the most important source of data for our research is the data archive of the Inter-University Consortium for Political and Social Research at Univeristy of Michigan. We believe that soon we'll be honoured with the opportunity to join it and to access its data collections which will help us to raise the level of our research. For a decade we were unable to join ICPSR because of not fulfilling one of its major conditions - the USSR had no machine-readable information about itself in the market. By now the situation has changed - our State Statistics Agency (Goscomstat) announced that a special section has been organized with the aim to produce machine-readable collections of data about the development of the USSR.

Our other interest in joining the Inter-University Consortium for Political and Social research is its Summer School program This year we send two of our researchers to ICPSR's Summer Program and hope that it will help us to move to new levels of research using modern software and methodology in analyzing social and political proccesses.

SOFTWARE
The software that is now being designed by our main programmer, Valentine Ponomarenko, is supposed to allow to use direct entry, scanners, on-line communication, and CD-ROMs to add data to our information systems. It will also provide us with an interface transition from data set to data set. Having information in several languages we come accross the problem of using English and cyrrilic alphabets which we manage to solve by having software designed by our prgrammers which is specific to our needs.

LINGUISTIC MEANS OF ANALYSES
Most of the documents we are dealing with are full-text and to create an effective information system is impossible without the built in and complex linguistic means of processing and analysing the full-text databases; thus a sub-system of advanced automatic analyses of the full text documents becomes an mandatory component.

In full-text documents special knowledge and information are accumulated in natural language. The idea of "restricted language", an approach adopted by most artificial intelligence systems, is irrelevant for the field of social and political sciences.

Moreover as for political text it's insignificant parts may be conveyed by so called key words. The more exact expression of the content calls for taking into account the relations between key words and the use of logical inference.

So we've started research in a field of structural linguistics which is absolutely untraditional for our institute. We have had a group of professional linguists assigned to our section, without costs to us, and are enthusiastically assisting our researchers in their tasks. The description of this project may be best addressed in a seperate article by Nina Leontieva, who is in charge of this work.

FINANCING
Our work is financed by the Academy of Sciences of the USSR. For the past two years we applied and received additional funds to work on our linguistic means of analysis project, to invite professional linguists and designers, and additional salaries for researchers working on this project. I realize how difficult it must have been for our academy chiefs to make the decision to finance

our idea of linguistic analysis and I am very pleased that they decided to finance the project. Our results to date are gratifying, we have developed several functional modules of linguistic analysis and other modules are in process - most notably the Russian Semantic Dictionary where political lexics are being fully described and the creation of the Thesaurus of Political Terms. Both of them may become commercial products and that is very important for us since it helps us to earn additional money to support our project.

Another funding source is the "USSR Congress of Peoples Deputies" database which was created by our institute. It is of interest to several universities in the USA and Europe. Version I contains the biographical information on deputies; education, profession, career; and information about the district the deputy represents, or public organization that elected him/her; the deputy's position in the Supreme Soviet and membership in committees and commissions. The database is in Russian. Names, geographic names, names of public organization have equivalents in English. We are planning to have Version 2 available soon which will have some additional fields - voting behavior of the deputies, voting results, etc.

So our section of applied research and informatics after two years of looking for our own way of developing new technologies and new thinking in our institute has some results, a lot of problems, and many ideas.

In conclusion I would like to express my great gratitude to all the IASSIST people who have made it possible for me to attend this conference - it was very helpful and informative. I hope that one day soon an IASSIST conference may be held in Moscow, USSR. IASSIST 90 has helped a lot to make this possible.

[1] A presentation to the IASSIST 90 Conference held in Poughkeepsie, N.Y. May 30 - June 2, 1990.

Career Achievement Award

On Febuary 21 the second IASSIST Career Achievment Award was awarded to Don Harrison. Judith Rowe presented the award to Don on behalf of our organization, at a retirement reception held in his office at the National Archives building in Washington, D.C. The award read, "IASSIST honors Donald Fisher Harrison for his career of dedicated service to the archival preservation of computerized information. February 1991." The authorization of the award followed the procedures approved last May by the General Assembly and finalized by the Admin Committee last October.

Tom Brown, IASSIST president and a colleague of Don, noted that "this marks a personal passing for me. For it was Don who introduced me to bits and bytes and everything else about machine readable data files sixteen years ago when I joined the staff of the National Archives. Many thanks Don!"

Iassist members join Tom Brown and the staff of the National Archives in extending the warmest of thanks to Don for his many contributions to IASSIST and his profession.

IASSIST