

University Information System RUSSIA: Scientific and Social Challenge

An appropriate information base is the main challenge for research and education in social and human sciences in Russia, especially in distant areas. Due to an information shortage, the general level of teaching and applied investigations is decreasing - university professors are unable to recommend as obligatory for study recently published books and periodicals: funding for book purchases is poor. As the Ministry of Science of RF (the Russian Federation) recently reported, only 67 scientific journals are available for 10,000 investigators in Russia (408 - in Great Britain, 186 - in the USA). Official government documents and reports, state statistics are also not available for educators and investigators. Due to the lack of public domain state statistics, the new research methods based on processing of large sets of numerical data are not developed in Russia. In the current situation, Internet-based collective resource is not only the most rational but the only possible approach to arrange the information supply and build the research base for investigations and advanced education in Russia.

The Moscow State University (MSU) Research Computing Center and non-commercial organization Center for Information Research since 1994 have been working to meet the challenge and develop the University Information System RUSSIA (UIS RUSSIA). In January 2000 the UIS RUSSIA (www.cir.ru) started operating on regular basis as a collective information base providing free access to all Russian universities. During 2000-2002 the UIS RUSSIA will compose an appropriate resource base for full-scale investigations in main human and social sciences. The Internet access ensures equal opportunities to researchers and educators in all regions of Russia. The universities of the former Soviet Union (FSU) countries are also granted free access.

In 2001 the UIS RUSSIA is planned to be available to local public libraries in Russia and FSU countries.

The current version of the UIS RUSSIA includes the data and documents' sources recommended as the first priority collections by the Center for Sociological Research and Economic Faculty of Moscow State University:

- official data and documents (laws, presidential

by Tatyana Yudina*

decrees and directives, governmental enactments, acts and regulations) since 1991;

- stenogramms (daily records) of State Duma of Federal Assembly of RF from 1996;

- Goscomstat of RF data (all available in electronic format);

- election statistics of both federal

and local levels since 1993, provided by Central Election Commission of RF;

- mass media sources (8 newspapers and 2 information agencies);

- databases, publications and reports of leading analytical centers;

- reference data on the Russian political system (brief history, prerogatives, structure and personnel of federal institutions, political parties, churches, etc.);

- extended reference information on the components of the Russian Federation.

All data collections are obtained for free from official holders/producers under legal agreements with Research Computing Center of MSU. The provision to process the information, integrate the results into the UIS RUSSIA and provide access to all universities of RF makes the UIS RUSSIA a valuable resource for full-scale socially relevant investigations.

Information update

Full text documents - official data and documents, stenogramms, mass media electronic versions are received electronically on daily basis, bulletins and analytical reports - on weekly or monthly basis (upon publication).

New full text collections will be added in 2000 :

- international agreements, signed by RF since 1991, international agreements signed by USSR,
- Constitutional Court of RF, Supreme Court of RF, Arbitrary Court of FR, decisions,
- Commonwealth of Independent States countries multilateral and bilateral agreements.
- local mass media sources.

All the documents are automatically processed – metadata created, classified, indexed, annotated and integrated into the UIS RUSSIA. The NLP technology provides for 20 Mb (equivalent of up to 10,000 pages) processed daily.

Appropriate retrospective coverage of each source will be realized during 2000.

Statistical data - Numeric data are the mostly used resource for social research. State statistics are the basis for socially-relevant investigations and sound recommendations for decision-makers. The UIS RUSSIA legally obtains, stores and updates collections from State Statistical Committee of RF. Under discussion are agreements with other main statistics-producing government institutions.

The Goscomstat of RF collections are received upon publication - on monthly, quarterly or annually basis. Statistical data are received in .doc format (digital versions of publication). To make the data available for Internet search, the UIS RUSSIA specialists convert the data into HTML format; and as a next step - into Excel spreadsheet format to make the data usable for secondary analysis. Currently available are the following data collections :

- Russian Annual Statistical Report, 1999,
- Industry of Russia in 1999,
- Regions of Russia in 1999,
- National Accounts of RF, 1991-1999,
- Environment Protection in Russia in 1999,
- Finance in Russia in 1999,
- Prices in Russia in 1999,
- Russia and Commonwealth of Independent States Countries in 1999,
- Russian Annual Demographic Report, 1999.

The data collections are also indexed to make them searchable using the UIS RUSSIA Thesaurus.

In 2000, all other Goskomstat of RF collections will be added. For 2000-2001, there are also plans to obtain and integrate the data maintained by the Centrobank of RF, the Ministry of Finance of RF, Goscomimyushestvo of RF, the Ministry of Labor of RF, other ministries, committees and agencies of RF, regional statistics of components of RF, international organizations measurements, and the databases created under the foreign grants. The numeric data collections are complemented by methodological notes.

Part of the Statistics of RF bloc are analytical reports prepared by leading “think tanks” in Russia - Russian-European Center for Economic Policy, Bureau of Economic Analysis, Fond for Population Sentiments Index research, etc. Reports of main Russian and foreign foundations’ grantees are included.

The documents are also indexed to make the analytical materials retrievable by Thesaurus-based cross-search.

Electoral statistics are received shortly after elections. The current version stores all general elections results since 1993, and local election results. The data are converted into Excel spreadsheet format and may be analyzed using standard software packages like Statistika, SPSS, SAS, etc. Electoral statistics is region-tailored and displayed in a map format.

NLP technology

To process and integrate large scope of electronic documents the technology of Automatic Linguistic Text Processing (ALTP) is realized under the project.

The ALTP performs:

- processing of electronic text corpora in main formats (ASCII, HTML, MS Word) in Windows and operating as DLL;
- morphological analysis of Russian texts;
- terms’ recognition/disambiguation;
- thematic analysis - event categorization, indexing, annotation/summarization;
- download of results on Oracle database server.

The main instrument of the technology is the *Thesaurus on Contemporary Russia (Thesaurus)*, created under the UIS RUSSIA project. In its current version it incorporates 18,500 concepts/descriptors, includes 6,500 geographic names, 39,000 synonyms, 70,000 relations between concepts, 200,000 inherited relations. The tool assists in detecting of main and subordinate topics in a document as a result of analysis of macroconcepts and relations between them. Macroconcepts are modeled by groups of concepts semantically related in Thesaurus. Thematic representation provides for evaluation of weigh of each term in a text and performs event categorization and annotating/summarization of a document. The Thesaurus enables to determine up to 90 - 95 % of terms.

The technology provides for up to 20 Mb of electronic texts to be processed on each Pentium200 PC and integrated into the University Information System RUSSIA daily.

The technology was evaluated by experts from NIST and DARPA in 1996 under the TextRetrievalConference-6 program and SummarizationConference in 1997. The results are among the best in a group of 14 participants.

The ALPT ensures advanced search instruments.

Search engine

Being initially designed to serve scientific needs the UIS RUSSIA provides for advanced search instruments: in

addition to traditional tools it includes value-added elements - the System of Subject Headings and Thesaurus on Contemporary Life in Russia. The System of Subject Headings consists of 200 topics (rubriks). All full text sources are filtered and event categorized according to System of Subject Headings. The Congressional Research Service, LC Legislative Indexing Vocabulary-based search is also available.

The UIS RUSSIA provides for research assistance, user services, metadata and annotation browsing, and thesaurus based query refinement. User-tailored automatic information update is realized.

Technical base: Architecture

The UIS RUSSIA operates at the Research Computing Center of MSU server. Mirror sites will be maintained in Novosibirsk and St. Petersburg to ensure more reliable access for universities in the Northern part of RF, Siberia and Far East.

Analytical bloc

Educational activity in advanced research methods has been started. Main research institutions were contacted and several software programs, Workbench of Sociologist, Workbench of Economist, etc., are presented by the authors. Special training class is arranged by the Research Computing Center of MSU, where the analytical software is downloaded and made available for university faculty. Special course is scheduled, it includes lectures analyzing main approaches to computer-based investigations and demonstration of working models, teaching and training in basic and advanced technique of social quantitative analysis. The main idea is to make available for investigators and educators the sound projects, to store the research results costly in both financial and human terms and to preserve them for future use. Consultations of authors of the software will be available for the faculty ready to use the programs in educational courses and investigations. Socially-relevant projects using the UIS RUSSIA stuff will be initiated. Not only MSU faculty is informed and invited but other universities of Moscow and regional universities.

Bilingual search instruments

The UIS RUSSIA has been initially designed as part of the international information structure to serve not only Russian researchers but also foreign specialists on Russia and general public. To meet the challenge, a special complex of the bilingual searching tools is being developed. The prototype of the bilingual complex is ready for testing and evaluation by a team of Russian American specialists. Funding to evaluate the bilingual search tools is currently being sought.

The NLP technology and developed bilingual complex will produce an annotation in English on each Russian

document. This accomplishment widens the UIS RUSSIA audience, helping foreign public to open Russia and foreign specialists to investigate Russia.

Work is underway on the Global Information Service (GILS)-profile to provide for the UIS RUSSIA integration into the world information space assisting Russian specialists in international cooperation in economic, social, political, human research.

Russian universities network

The UIS RUSSIA has been designed as a base for inter-university cooperation in consorted and rational efforts to build a networked collective information infrastructure. The regional universities may actively participate. Currently up to 50 local universities are technically and technologically equipped to take part in the cooperation. The Open Society Institute (Soros Fund)-funded "Russian Universities' Internet Centers" program provided the regional universities with the hardware-software platform compatible with that of the UIS RUSSIA. The NLP technology developed under the UIS RUSSIA project may be passed on for free to the regional universities to enable them to develop information systems of their own on local resources, integrated into the UIS RUSSIA. The local stuff is important to make the social analysis relevant.

In this respect the UIS RUSSIA is close to the American universities' Internet2 initiative directed to rationally build Internet-based far-reaching educational and research network.

From the very beginning the UIS RUSSIA project was developed in cooperation with the Michigan Inter-university Consortium for Political and Social Research and European Consortium for Social Research. Specialists of both structures visited Russia and have been of help to the Russian researchers.

Program to become self-supportive

Maintenance of the system on self-supporting basis is a challenge of the project. The experiences of information structures in the USA, Europe and other countries' have been analyzed, and main elements of financial activity of those organizations will be realized - institutional membership with annual dues for foreign universities and other organizations. Preliminary discussions with American and European colleagues - university professors, think tank specialists, government analysts, journalists prove that this way is the most acceptable for them as well. The dues will create a relatively stable and predictable financial base and allow the program to engage in long-range policy to develop the information resource and provide access for free to the high education institutions in Russia.

The team

The UIS RUSSIA project began 1994. The key specialists have worked together since that time. The team includes 20 specialists from the Research Computing Center, other faculties of Moscow State University, academic institutions and other universities of Moscow. Several specialists are invited for half-time job and consultations. A group of American consultants provide their expertise of the project on the permanent basis.

Since 1994 the project has been supported by grants from Russian Fund for Basic Research, Russian Humanitarian Scientific Fund, Ministry of Science and Technologies of RF "Informatization of Russia" program, MacArthur Foundation, USA, Ford Foundation, USA.

* Tatyana Yudina, Ph.D., Leading researcher of Moscow State University Research Computing Center, Director of University Information System RUSSIA project.
yudina@mail.cir.ru