

Russian Periodical Index Digital Project (*Letopis' zhurnal'nykh statei*, 1956-1975)

Kristine R. Brancolini
Michelle Dalmau
John A. Walsh

ABSTRACT. The Indiana University Digital Library Program¹ has digitized and published on the World Wide Web a twenty-year portion of *Letopis' zhurnal'nykh statei* (1956-1975),² a serial publication that indexes Russian periodicals from 1926 to the present. It covers more than 1,700 journals, series, and continuing publications of academies, universities, and research institutes in the humanities, sciences, and social sciences, as well as some popular literature. The work was partially funded by a three-year (1999-2002) U.S. Department of Education grant. This article describes the challenges the Indiana University Digital Library Program encountered in digitizing and encoding this resource and creat-

Kristine R. Brancolini is Director of the Indiana University Digital Library Program (E-mail: brancoli@indiana.edu).

Michelle Dalmau is the Interface and Usability Specialist of the Digital Library Program (E-mail: mdalmau@indiana.edu).

John A. Walsh is the Manager of Electronic Text Technologies in the Library Electronic Text Resource Service of the Digital Library Program (E-mail: jawalsh@indiana.edu).

Address correspondence to the authors at: 1320 E. Tenth Street, Main Library E170, Indiana University, Bloomington, IN 47405.

Brancolini and Walsh served as Co-Directors on this project. Dalmau was responsible for interface design and usability activities on the project.

The *Russian Periodical Digital Project: Letopis' zhurnal'nykh statei, 1956-1975* was partially funded by a grant from the U.S. Department of Education (Technological Innovation and Cooperation for Foreign Information Access—TICFIA—Program).

The Serials Librarian, Vol. 47(3) 2004
<http://www.haworthpress.com/web/SER>

© 2004 by The Haworth Press, Inc. All rights reserved.
Digital Object Identifier: 10.1300/J123v47n03_05

ing a bilingual Web site to deliver a usable, searchable reference work. The grant has ended but work continues to improve the online index. [Article copies available for a fee from The Haworth Document Delivery Service: 1-800-HAWORTH. E-mail address: <docdelivery@haworthpress.com> Website: <<http://www.HaworthPress.com>> © 2004 by The Haworth Press, Inc. All rights reserved.]

KEYWORDS. Russian periodicals, indexes, digitization

One of the beauties of the Russian Periodical Index Digital Project is its simplicity: Convert a 20-year run (1956-1975) of the major Russian-language periodical index, *Letopis' zhurnal'nykh statei*,³ to electronic form and publish it free of charge on the Web. It is easy to describe, and for people who are familiar with *Letopis' zhurnal'nykh statei*, the value of the project is undeniable. This project involved the digitization and XML-encoding of 300,000 pages—more than 3,000,000 citations—and the creation of a bilingual Web site to provide access to the bibliographic data. Begun in 1999 and substantially completed in 2003, the Russian Periodical Index Digital Project will end in summer 2004.

The Indiana University Digital Library Program received one of eight Technological Innovation and Cooperation for Foreign Information Access (TICFIA) Program⁴ grants awarded by the U.S. Department of Education for the time period 1999-2002, beginning on October 1, 1999. The TICFIA Web site provides the following information about what was then a brand-new program:

The purpose of Technological Innovation and Cooperation for Foreign Information Access (TICFIA) Program is to support projects that will develop innovative techniques or programs using new electronic technologies to collect information from foreign sources. Grants will be made to access, collect, organize, preserve, and widely disseminate information on world regions and countries other than the United States that address our Nation's teaching and research needs in international education and foreign languages.⁵

TICFIA is administered by the Office of Postsecondary Education, International Education and Graduate Programs.⁶

Indiana University (IU) assembled a multidisciplinary project planning team, which submitted a proposal in mid-March 1999, and was tentatively awarded \$80,000 per year for a period of three years. However, funding was actually granted on a yearly basis and the final amount of total funding was \$277,000. At the end of Year 1, we submitted a project update and requested

funding for Year 2 in the amount of \$80,000. By the end of Year 2, we had substantially altered our Plan of Work. We realized that we had under-budgeted the project, and requested \$117,000 for Year 3. The university also increased its financial contribution to the project.

The grant ended on September 30, 2002, and we released *Letopis' zhurnal'nykh statei*, version 1.0, in March 2003. This version provided online access to the bibliographic citations for 1956-72, searchable by keyword across the entire index or limited by subject and date. Users could enter Cyrillic text directly or use a virtual, clickable Cyrillic keyboard. Users also had the option selecting citations and e-mailing their selections. Usability testing revealed some problems with the keyboard and the layout of the subject headings. As a result of this testing, the design team made a number of changes and a new version of *Letopis' was launched November 2003. The revision improved search functionality, the results display and overall help documentation. The current version of Letopis' includes citations from 1956-1972. The subject headings for 1973-75 changed significantly and are being reconciled with the headings for 1956-72. The final version of the index is scheduled to be released in summer 2004.*

LETOPIS' ZHURNAL'NYKH STATEI

We selected *Letopis' zhurnal'nykh statei*, 1956-1975, for digitization based upon the unanimous advice of Slavic librarians who were queried via a listserv for Slavic specialists. One librarian wrote that this resource would be particularly valuable for the years we proposed digitizing because "there [are] no remotely comprehensive bibliographies that can compare to it." We supplied many letters of support to attest further to the significance of providing online access to *Letopis' zhurnal'nykh statei*. One supporter noted, "Senior Slavic scholars as well as students tend to ignore this resource. . . . Making the resource available over the Internet, with the enhancements that electronic formats offer, will ensure that this resource will be heavily used." We believe that the potential audience is enormous. Our goal and desire is to provide *Letopis' zhurnal'nykh statei* to students and scholars at home and abroad. As another supporter notes, "The Web site would provide access to a vast hidden treasure of information, not only for those at institutions with the largest Russian and Soviet collections, but for the many hundreds of researchers with medium-to-small collections and no holdings at all of *Letopis' Zhurnal'nykh statei*. In fact, I would predict that this database would also be heavily used by hundreds of libraries throughout Eastern Europe, Russia, and the countries of the former Soviet Union."

We chose to digitize *Letopis' zhurnal'nykh statei* for the years covering 1956-1975 for two reasons. First, Soviet-era paper is extremely acidic and many of the pages were already breaking along the edges and the gutter. It had

been reprinted on acid-free paper from 1926-1955, so the volumes beginning with 1956 were still at risk. Second, those two decades encompass one of the most interesting periods in Soviet history. It is the time when Soviets were leading in space research and were introduced to television. These are the years of "Thaw," during which Nikita Khrushchev revealed the horrors of the Stalinist regime to the Soviet people. The 1960s are remembered by many Russians as the time of "physicists" and "lyricists," when science and art were flourishing. The 1970s cover Brezhnev's "Period of Stagnation," when, some people say, not much was happening in the Soviet Union. However, this was time when many people began to realize the need for social and political reforms. *Letopis' zhurnal'nykh statei* covers a wide range of topics: from works of Communist leaders to technical innovations, agricultural advancements, medical techniques, communication technologies, and art and literature.

Copyright

As the project team began its planning, questions arose about the copyright status of *Letopis' zhurnal'nykh statei*: Is the work covered by copyright? Is it covered by copyright for all years of its publication or are some volumes in the public domain? Some of the librarians we queried about the project expressed concern that we might need permission from a copyright holder in order to legally undertake the project. In order to answer these questions, we spoke with Fred Cate, professor and noted copyright expert on the faculty of the Indiana University School of Law; and Michael Newcity, professor and expert on Soviet/Russian copyright law on the faculty of the Duke University School of Law. Independently, both concluded that this work is in the public domain for all years of its publication. This conclusion was based upon copyright law in both the Russian Federation and in the United States, as there are questions concerning which copyright law would apply in this case. However, in either case the conclusion is the same: Legally, we may digitize any and all volumes of *Letopis' zhurnal'nykh statei* and deliver the electronic resource to users via the World Wide Web, without permission from the publisher. There is no indication that the index was ever covered by copyright; as a government publication it may well have always been intended to be in the public domain. The index has been accessible since March 2003 and Indiana University has received no inquiries or complaints from Russia regarding the legality of the project.

TICFIA FUNDING PRIORITIES AND AUTHORIZED ACTIVITIES

The purpose of this project was to provide widespread access to the only comprehensive index to Soviet journals published during the years 1956-

1975. The index contained invaluable information available from no other information source during this time period. The entries themselves, however, were poorly indexed, making them so difficult to search that *Letopis'* was virtually unused, except by the most persistent scholars. Furthermore, as noted above, the issues were printed on highly acidic paper and in danger of disintegration. By digitizing these resources and offering them on the World Wide Web, we have accomplished three goals: we have preserved the information; made it available to users worldwide; and provided keyword searching, which helps overcome the lack of cumulative indexing. The project encompassed three authorized activities under Title VI, Section 606, numbers (1), (3), and (5) of the authorizing legislation:⁷

- to facilitate access to or preserve foreign information resources in print or electronic forms;
- to develop new means of shared electronic access to international data; and
- to develop methods for the wide dissemination of resources written in non-Roman language alphabet.

Facilitating Access and Preserving Foreign Information Resources

The project has preserved and provided improved access to a major foreign information source. Our first consideration was preservation. The paper stock used for the index, as already stated, was extremely poor. Although the volumes from 1926 through 1955 had been reprinted by Kraus on quality paper, the originals (and the volumes of the index we digitized) were, by contrast, generally deteriorating. These volumes could not sustain normal use without suffering damage and the potential loss of the information they contained. Volumes from 1957, for example, were so brittle that the pages would break when used or copied. Some volumes as late as the early 1970s were in the same condition.

At Indiana University the issues were library-bound using oversewing to consolidate the text block. Oversewing, which was the standard method at the time, reduces the inner margin of the text and perforates the pages at 1/4" intervals. As oversewn volumes deteriorate, the fragile paper detaches along the perforation lines. Oversewn volumes are also very difficult to use without causing damage, since the tight sewing resists opening or photocopying. The disadvantages of oversewing are in this case compounded by the 3-inch and greater width of many of the volumes, which causes even more difficulty in handling and copying.

Our second consideration was to improve access to the information. Students and scholars needed access to *Letopis'* and to the citation entries, and yet access was extremely limited. Few universities in the United States had back files or current subscriptions to it. It was a reference work, and so students and

scholars did not have access via interlibrary loan; the volumes themselves could not be loaned, and photocopying was not a viable option for an index.

Access to the information recorded in *Letopis'* was also limited due to its frequency of publication and a lack of detailed subject analysis. Refined subject searches were almost impossible principally because *Letopis'* lacked a cumulative index. Over the years, it included periodic indexes of varying quality appearing at different intervals (bimonthly, quarterly, annual). IU Slavic Studies Area Bibliographer Murlin Croucher noted, "For the years that do not have quarterly indexes, or if one wishes to search by subject or title, it is necessary to use the individual weekly issues with their indexes. Thus a subject or title search of 20 years of articles would require perusing 1,040 issues of this bibliography."⁸ For the years we digitized for this project, there was a quarterly name and geographic index; it began at a length of about 250 pages, increasing to over 500 pages per year by the mid-1970s, making a complete search nearly impossible. There were no title indexes, and the subject categories were prohibitive ("so broad as to prove almost useless," notes the bibliographer for Slavic and East European Studies at the University of Chicago). Digital access has overcome the barriers to use created by the lack of indexes and adequate subject analysis; users now have access by any keyword in the citation or may search within assigned subject headings.

Keyword searching provides significant advantages over the arrangement of the paper index. However, with more than 3,000,000 citations, it is also imperative that we offer users a way to limit their searches. At the beginning of the project, we planned to encode the text heavily, allowing users to limit their searches by author, title, journal title, year, and subject. Unfortunately, this proved to be excessively time-consuming and labor-intensive. We discussed the situation with subject specialists and scholars, and then reexamined the index to see how we might combine the need for refined searching with citation elements that would lend themselves to automated tagging. Despite the limitations of the subject headings, they do provide a level of content analysis to supplement citation information. Consequently, we settled upon subject headings and date as the important and feasible limiters.

New Means of Shared Electronic Access to International Data

The World Wide Web has proven to be an effective and reliable distribution medium for electronic publications. Users from all parts of the world have access, often free of charge, to this unique resource. Since research libraries in the former Soviet Union have complete runs of the journals indexed in *Letopis'*, we suspect that it will be more heavily used overseas than in the United States. Digital technology and the World Wide Web allow us to provide vastly improved access to a relatively inaccessible information resource, including access to the index itself and access to the bibliographic data contained in the index, through keyword searching.

Wide Dissemination of Resources Written in Non-Roman Alphabets

Letopis' zhurnal'nykh statei is an information resource itself; the citations offer an informative view of scholarly publishing during the Soviet era. However, the ultimate goal is to provide users with the actual articles once they have identified citations of interest. Most of these journals are too old and too specialized to be available electronically, so we turned our attention to providing users with information about the availability of paper copies. Most users of the index have reported to us that they know how to obtain copies of journal articles. However, we anticipate that online access to the index will attract new users who may be less familiar with the various document delivery options. We plan to provide information about document delivery options on the Web site.

PRIMARY AND SECONDARY PROJECT GOALS

Our primary goals for the project were to preserve the content, improve searching of the index, and increase use of the index. We have successfully met the first two goals and we are convinced that once we have more widely publicized availability of the index, the third goal will be realized. We believe we have already brought some new users to the index. We also had secondary goals related to the technical aspects of the project. We hoped to learn more about optical character recognition and Cyrillic text; presenting and searching Cyrillic text on the Web; and effective methods of sharing our findings with other digital library programs. Although we believed initially that the technical aspects of the project were fairly straightforward, it turned out that they were vastly more challenging than we had anticipated. In retrospect, we realize that our goals were extremely ambitious for a three-year project. However, through creative project management and the considerable skills of our technical staff, we found cost-effective ways to realize all project goals.

The goals of the project remained constant for the duration of our work, but we significantly modified our Plan of Work as the project proceeded; flexibility was crucial to successful completion of the project. Because the index was in Russian, we hired a full-time Russian-speaking librarian, Andrew Spencer, to manage the project. Mr. Spencer believed that we had greatly underestimated the technical challenges that the project would present, so near the start of the project, we shifted production from Indiana University to outside vendors, leaving our project team to deal with the technical issues rather than production. We had initially planned to digitize the pages in-house; we eventually outsourced this work to Northern Micrographics, which took nearly one year to complete the digitization of page images. We performed the optical character recognition (OCR) processing in-house and began to proofread and encode the text files with local staff. However, this proved to be too slow, due to the constraints of working with student employees with difficult or inconvenient schedules. Had we continued with this method of proofreading and encoding,

the project would have taken ten more years instead of the two remaining years of the grant.

We had hoped all along to automate as much of the text encoding as possible. That required correcting the OCR first, then running a locally developed program to insert the XML tags. Even paying local staff to correct text proved to be prohibitively expensive. We searched for another alternative. Due to changes in the global economy, it became cost-effective to outsource this work to a vendor in the United States with a work force in Russia. In the fall of 2001, we outsourced proofreading to Varda Graphics, located in Skokie, Illinois, and wrote a program to perform most of the XML encoding. Our own staff added tagging that could not be automated.

PROJECT ACCOMPLISHMENTS

We are still working on providing access to the citations from 1973-1975, but our project Web site demonstrates our accomplishments to date:

- A bilingual Web site that offers a Russian-language and English-language interface and all documents and project reports translated into Russian.
- Access to nearly 3,000,000 citations from all Russian-language periodicals published in the Soviet Union between 1956 and 1972, plus selected articles in translation from other journals.
- A search engine that allows keyword searching of the index with filtering by subject heading and/or date.
- A clickable Cyrillic keyboard for users who do not have an appropriate input device or operating system configuration.
- The ability to control the number of citations displaying on the page.
- Extensive "Help with *Letopis*" documentation designed to guide both novice and advanced users in effective use of the index.
- The option to e-mail results.
- An outline of the subject headings used in the index, so users can browse for the ones that best match their topic.

Many of these features go far beyond what we had envisioned when we planned the project five years ago. One of the first features we added was the clickable Cyrillic keyboard. We have received valuable feedback from our users during the design and development phases of the project, and this has shaped the services we offer. One example of this is the use of subject and date filtering as refinements to keyword search. We also learned during testing that most users expected to search *Letopis*' phonetically. We are exploring ways to provide this capability, such as an automatic mapping from phonetic transliteration to Cyrillic.

USE OF TECHNOLOGY

There were two major technical phases to the project, digitization and delivery. Digitization covers the process of moving from the original print source, to digital image, to text, to XML-encoded text. Delivery covers the client-side and server-side developments required to deliver the data online. A number of technologies have been employed in each phase.

Digitization

The digitization process involved the following steps and technologies:

1. The original print pages were scanned and saved in the TIFF image file format.
2. The TIFF images were processed by ABBYY FineReader⁹ optical character recognition (OCR) software and saved as editable Unicode (UTF-8) text files.
3. The text files were processed by a markup program (LMU, or Letopis' MarkUp) that parsed the unencoded text files and produced pseudo-XML output. LMU was developed by the *Letopis*' project team and is written in the Java programming language and uses the open source gnu.regexp Regular Expressions for Java library.¹⁰ The pseudo-XML output produced by LMU is neither valid nor well-formed XML and requires further human editing to produce valid XML.
4. The pseudo-XML files output by the Letopis' MarkUp program were edited, checked, and validated by *Letopis*' project team staff using various XML editing tools, including Corel XMetaL¹¹ and Corel WordPerfect.¹²

Delivery

The *Letopis*' Web application is written in Java Server Pages (JSPs) and runs under Apache Tomcat servlet container.¹³ The application connects to Software AG's Tamino XML Server,¹⁴ where XML data are stored.

Problems and Solutions

The issues presented by the multilingual nature of the *Letopis*' data were by far the biggest challenge throughout all stages of development of the *Letopis*' project. *Letopis*' contains text in a number of different scripts, including Cyrillic, Roman, and Greek. As a result, we made the seemingly obvious decision to use the Unicode UTF-8 character encoding, which accommodates all the

languages and scripts found in the *Letopis'* data. UTF-8 is also the default character encoding for XML. Unfortunately—in the early stages of the project, especially—many products, even those that purported to be XML products, were lacking in support for Unicode. We faced unwelcome Unicode-related issues with XML search engines, XML editors, programming languages, Java servlet containers, fonts, and Web browsers. In general, the solution was to patiently search for and test a number of different products and technologies until finding those that suited our needs. We found that Software AG's Tamino XML Server provides excellent support for storing and searching multilingual Unicode and XML-based data. Java, and the `gnu.regex` Java library, had to be used for regular expression matching, text processing, and automated markup. Typically, one would use Perl for such text-processing, but earlier in the project, Perl had very incomplete support for Unicode; that situation has since improved.

Query input was a troublesome issue. We anticipate that a diverse, world-wide community of scholars and researchers will be using this important resource. Not all these users will have a Russian-localized operating system that is set up for simple input of Russian/Cyrillic characters. As a result we must provide users with more extensive help that will instruct users how to configure their systems to input different languages and scripts. In most cases, we simply point users to the relevant documentation provided by the various operating systems. Additionally, using JavaScript, we created on-screen "virtual keyboards" in the browser that allow users to use their mouse to point-and-click on different characters to enter those characters into search input fields.

USABILITY EVALUATION

The usability expert on the project team conducted a thorough evaluation of *Letopis'* in September 2002 to ensure that the Web site's functionality, content and organization met the representative user's expectations. Areas under investigation included: (1) the ability to intuitively interact with, including input into, the Web site; (2) the ability to promote successful searching; (3) the provision of useful results display and options; and (4) the creation of clear, concise, context-sensitive help to aid searching a bilingual Web site. The overarching focus of this usability study was to gauge general interaction with the search interface, based on representative tasks, and help documentation, as they are intended to work together to help users overcome some of the technological challenges presented by the unique characteristics of the resource.

Certain technical implementations were taken into consideration when the evaluation plan for this usability study was prepared. Of significance is the use of Unicode to render multilingual characters on the *Letopis'* Web site. While a standard for Unicode has been defined, it is not uniformly interpreted by all the Web browsers. It was important to learn from representative users their

preferred browsers, typical or expected mode of input with Russian online resources, and e-mail programs of choice, as well as how successfully they were able to complete tasks using *Letopis'*.

The following goals and objectives guided the evaluation of *Letopis'*:

1. Determine if participants are successful at searching, including:
 - a. interaction with the linked Cyrillic alphabet or keyboard settings
 - b. utilizing search filters to limit search results
 - c. understanding the sorting of the results display
 - d. understanding the e-mailing of results
 - e. referencing getting started tips/troubleshooting to enable usage.
2. Determine the advantages or disadvantages of both context-sensitive and overall help.
3. Determine if screen instruction and labeling are sufficient or if more is required rather than having to launch the help.

Methodology

Because of the various ways in which a user could interact with *Letopis'*, especially in terms of input, it was decided that a usability lab test would be the appropriate usability approach. The lab provided the audio and video technologies necessary to capture keyboard and mouse movements, commentary as well as screen shots during the evaluation.

Participants

Participants were recruited from Indiana University's Russian and East European Institute (REEI). Participant criteria for recruitment included faculty, students and subject specialists/librarians with: (1) a Russian studies background; (2) reading knowledge of the Russian language; and (3) no prior exposure to the *Letopis'* search and help pages.

Four graduate students and two professors volunteered to evaluate *Letopis'*. Their areas of study or expertise ranged from Russian and Ukrainian Studies to Central Asia and East European History to Library Science. Russian was not the native language of any of the participants. With one exception, all participants were fluent in Russian, with language study ranging from two to over ten years; the other participant reported a reading knowledge of Russian. Four of the six participants used an actual Cyrillic keyboard to input Russian text. Of these, three preferred to input in a transliterated format, either as defined by the Library of Congress or the International Phonetic Library.

Testing Protocol

Participants were asked to complete a background questionnaire, tasks, satisfaction questionnaire, and a brief interview. The tasks presented simulated

research questions pertaining to Russian history during the Soviet era. Participants completed six tasks, which are listed below.

1. Find article citations about ленинизм (Leninism). Review the results.
2. Find citations authored by хрущев (Khrushchev) during 1956 and 1957. Select a few citations relevant to your interests from each of the pages of results.
 - a. Check to make sure that the citations to be e-mailed are the ones you expected.
 - b. E-mail those citations.
3. Can you explain how you think the “Limit by Subject Headings” feature works?
4. Search for рабочий класс (working class) under the topic Литературоведение. Художественная литература. Фольклор (Folklore/Literature). How many results do you see?
 - a. Conduct the same search using the more specific topic Русская литература (Russian Literature). How many results do you see?
5. Find information about великий октябрь (Great October) published during 1957 under the topic, история (History). Review the results.
6. You want to use the online *Letopis' zhurnal'nykh statei* on your computer at home. Can you find information regarding browser settings and language input options?

Findings

The usability test uncovered two severe obstacles to successful use of the index: participants were unable to input queries in Cyrillic and unable to use the subject headings as search filters.

Query Input. Initially, the participants were unable to input their queries successfully. All six participants opted for the default Latin keyboard with the expectation that the Web site would support transliteration. Although a linked Cyrillic alphabet was provided as a form of input in the version of the Web site tested, the participants did not understand its purpose:

- “When I first saw it, I didn’t know what to do. I was expecting a piece of paper reference for keyboard input [transliteration].”
- “I didn’t understand the purpose of the linked alphabet. I thought it would be in here [keyboard].”
- “I have no idea why it’s there. Are we supposed to put it in with that? Never clicked on Cyrillic letters always used the keyboard. Very strange.”
- “This keyboard is just a regular English keyboard.”

One participant requested an actual Cyrillic keyboard, which is his preferred form of input. Once the participants realized the utility of the linked Cy-

rillic alphabet or switched to an actual Cyrillic keyboard, queries were successfully entered and their impressions mostly changed:

- “Now it [linked alphabet] is fine. It’s cute. I like it. Now I get it. I like it. It’s cool.”
- “I think this [linked alphabet] is very good, I’m glad you have it.”
- “Not a big hardship.”
- “I can get used to this but it [linked alphabet] is slow.”
- “Becoming much more comfortable [using the linked alphabet].”
- “Can we switch keyboard to Cyrillic input?”
- “I don’t like it [linked alphabet]. I’d rather use the keyboard.”

Subject Headings. *Letopis’* utilizes a hierarchical, three-level subject schema, which was implemented in the online version as a search filter for narrowing results. Two tasks were designed to explore interaction with the subject headings: task 3, which was a predominantly talk-aloud assessment, and task 4, which required searching using the headings—including a third-level heading. Not one participant could successfully complete task 3 and only two out of the five were able to complete part of task 4.

The version of the Web site used for the testing displayed headings in text area boxes, one box for each subject heading level (see Figure 3). The search screen initially loaded with only the first level headings populated in the first text area box. The user would have to click on a top-level heading to reveal level 2 subheadings, and on level 2 headings to reveal level 3 subheadings. Users were confused because not all subject levels were initially visible and because the expected functionality of a text area, particularly a blank text area, is to support data entry, not data selection.

Letopis’ Redesigned

In light of the usability findings, the *Letopis’* search page was significantly redesigned. We needed to retain the virtual Cyrillic keyboard to facilitate input for those who do not have an actual Cyrillic keyboard. However, the virtual keyboard was redesigned to look more like a keyboard (see Figures 1-2).

The subject heading display was converted to a hierarchical tree-structure, which can be expanded or collapsed to more easily reveal the various subject levels (see Figure 3, left). In addition to redesigning the virtual Cyrillic keyboard and the subject headings filter display, we improved the results display, labeling, instructions and help text.


Some difficulties identified during usability testing remain unresolved due to limited time and resources, including the support of transliterated data input and session-based ability to e-mail citations. Further investigation will determine the difficulty of mapping the Cyrillic character to a transliterated version of Russian. The ability to e-mail citations is currently supported but only from

FIGURE 1. Revised Linked Cyrillic Alphabet

Search Letopis'

Conduct a keyword search by entering **Cyrillic characters**. Select a letter from the Cyrillic alphabet provided below to compose your search terms. You can also use a Cyrillic keyboard or change the input of your current keyboard to type in Cyrillic. Visit help on [Language Input Options for Searching](#) or [Additional Alphabets](#) for more information.

Limit your search results by using the year and/or subject search filters in the "Limit To" section below.



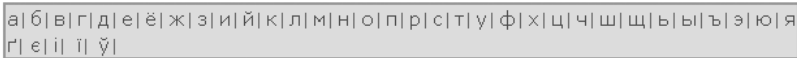
Search for: All words ▾

Tips: Limit your search by using "Exact Phrase" option: useful for known names and common phrases
Broaden your search by using wildcard *: **Гагарин*** will result in Гагарина, Гагарину, Гагарине, Гагариным

FIGURE 2. Original Linked Cyrillic Alphabet

Search Letopis'

Conduct a keyword search by typing term(s) or using the linked Cyrillic alphabet provided. Limit your search by using the year and/or subject search filters. [Show Cyrillic Alphabet](#) | [Hide Cyrillic Alphabet](#) ?



Search for: All words ▾ ?

Limit search by using "Exact Phrase" option: **Иванов** will not also find Иванова
Expand search by using wildcard *: **Гагарин*** will result in Гагарина, Гагарину, Гагарине, Гагариным

one results page at a time. The next version of *Letopis'* will include session-based e-mailing, which will support citation selection across results pages.

USING LETOPIS'

The current version (1.1) of *Letopis' zhurnal'nykh statei* was launched in November 2003. According to our usage statistics the average number of hits

per day has increased since December 2003 (see Figure 4). The number of unique visitors has also increased per month from 202 in December 2003 to 512 in March 2004. In comparison to other Indiana University Digital Library Program resources, these numbers are low.

However, *Letopis'* is a unique resource and not comparable to online color photographs or full-text collections of nineteenth-century fiction. We recognize that *Letopis'* is a specialized, historical bibliographic resource with a limited audience, but we are developing a publicity plan that should

FIGURE 3. Revised Subject Headings Filter (Left); Original Subject Filter Display (Right)

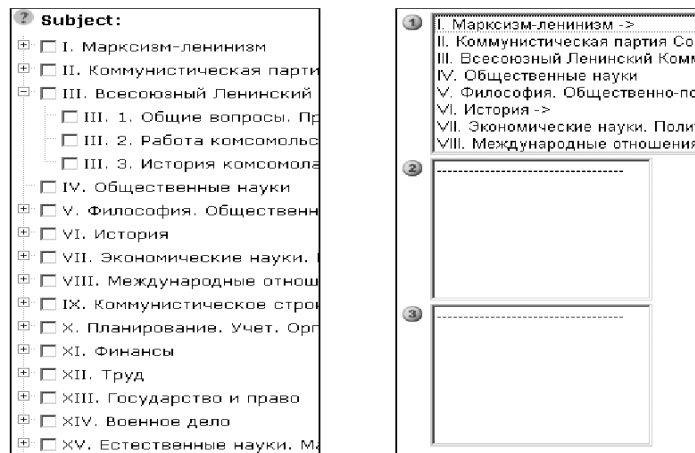


FIGURE 4. *Letopis'* Usage Statistics, Webalizer

Summary by Month										
Month	Daily Avg				Monthly Totals					
	Hits	Files	Pages	Visits	Sites	KBytes	Visits	Pages	Files	Hits
Apr 2004	152	76	71	22	76	11710	331	1076	1143	2285
Mar 2004	112	48	60	16	110	15934	512	1868	1504	3499
Feb 2004	186	86	64	15	131	20780	437	1872	2510	5414
Jan 2004	88	50	35	14	132	13490	444	1092	1559	2741
Dec 2003	89	44	21	12	76	5581	202	344	717	1438
Totals						67494	1926	6252	7433	15377

produce more visibility and, thus, promote its usage. Even without extensive publicity, it is clear according to the Summary by Month chart (see Figure 4) that information about the availability of *Letopis'* online is circulating among users.

Our usage statistics are provided by Webalizer, a free Web server log file analysis program, which presents high-level summary information about usage. We plan to further customize Webalizer so that we can more easily track certain types of usage patterns. For instance, some of the users who visit *Letopis'* are unable to search the resource due to older browser versions. Because of the various technologies *Letopis'* requires in order to search and display Cyrillic content, only current browser versions (5 and greater) are supported. Those who visit *Letopis'* with earlier browsers are redirected to a help page with links to download current browsers. Current Webalizer customizations make it difficult to determine if these same users come back after downloading a supported browser. We would also like to display the top 25 countries of origin for users rather than the top 9, which is the default. With the current settings, we are able to see that Russian Federation users are recurring, but we would like to verify if users from Eastern Europe and other former eastern bloc nations have discovered *Letopis'*. With further customizations of the log statistics, we hope to gain a better picture of those using *Letopis'*: their successes and their challenges.

The IU Digital Library Program policy is to provide access to all users, so most of our collections and resources are designed to support older browsers. However, *Letopis'* is a special case and it required that we use some more recently supported technologies, such as Unicode and JavaScript to provide the necessary functionality for the index. We knew this would cause problems for some users, including many users accessing *Letopis'* from abroad. We expect this situation will improve steadily as users worldwide upgrade to more current technology; however, we will continue to investigate ways to deliver the required functionality to older browsers.

NEW AREAS/NEW KNOWLEDGE

The *Letopis'* project employs a number of relatively new technologies, including XML, Native XML Databases/Servers, and Unicode. We believe that it is one of the first large-scale digital library projects to rely so extensively on these technologies to deliver multilingual content over the Web. We have disseminated information about the technical challenges, problems and solutions posed by this combination of multilingual content and new technologies in forums such as the 2001 XML Conference and Exposition and the fall 2002 Digital Library Federation Forum. Please see Appendix 1 for a complete list of papers and presentations related to the project; these resources are also linked

from the project Web site, under "About *Letopis*": <http://www.dlib.indiana.edu/reference/letopis/>.

PLANS FOR THE FUTURE

Discussions with librarians and other users of *Letopis' zhurnal'nykh statei* helped the project team identify possible enhancements to the online resource. These enhancements fall into one of two categories: improved functionality for the index and additions to the Web site; or improved access to the journals themselves. In the first category, we have considered the possibility of offering input of transliterated Russian. Usability testing revealed that most users in the United States do not input Cyrillic characters, so we have been exploring methods of mapping Cyrillic to transliteration as a further input option. The second improvement to searchability would be more extensive encoding of the citation to allow searching by author, article title, journal title, and more. Now that all of the citations have been corrected, it might be possible to write a program to add these tags.

In the category of providing improved access to the journals, we are working on one improvement, providing information on the site about document delivery options; and considering another, the addition of full-text journals to the resource. Before planning the project to convert *Letopis' zhurnal'nykh statei* to electronic form, we investigated the availability of these journals in U.S. research libraries. We were satisfied that many of the most important journals covered by the index are available in U.S. research libraries; Appendix 2 provides a list a selected list of the most important Russian-language journals of the period held by the Indiana University Libraries. However, due to the high acid content of Soviet paper, the condition of many Russian-language publications is poor and deteriorating rapidly. Many other important journals may not be so readily available in the U.S. and other countries. We are considering digitizing some of the journals that are indexed in *Letopis' zhurnal'nykh statei*, 1956-1975, and linking articles to entries in the index. This project would be planned with input from Slavic librarians and scholars, to insure that we created a prioritized list that corresponds to user needs.

CONCLUSIONS

At times during the three-year grant-funded Russian Periodical Index Digital Project: *Letopis' zhurnal'nykh statei*, 1956-1975, the project team feared that it would be impossible to meet our goals with the money available and within the project time frame. In truth, the project was seriously under funded; we had underestimated the technical challenges presented by converting a

multilanguage index to a Web-based resource. We could have spent much more than the \$277,000 we received from the U.S. Department of Education plus the considerable cost-share contributed by Indiana University over more than four years. However, we are extremely pleased with the results. All 3,000,000 bibliographic citations have been corrected and encoded and we are delivering the index free of charge to users worldwide. The content of the index has been preserved for generations of future users, sustained by the technical infrastructure of the Digital Library Program. Because of our extensive user testing, we know that we have created a usable Web site, designed around the needs of actual users of the index, and we have valuable input for making improvements to the site and to the functionality of the index. In the process, we learned much about the experimental nature of this type of work. Without significant financial support from University Information Technology Services and the University Libraries, the two major partners in the Digital Library Program, it would have been difficult, if not impossible, to complete this project to our specifications.

NOTES

1. *Digital Library Program Web site*. Last updated 19 April 2004. Copyright 1997-2004, Trustees of Indiana University. URL: www.dlib.indiana.edu
2. *Letopis' zhurnal'nykh statei Web site*. Version 1.1 (Released November 2003). Copyright 1999-2003, Trustees of Indiana University. URL: <http://www.dlib.indiana.edu/reference/letopis/>
3. Ibid.
4. *Technological Innovation and Cooperation for Foreign Information Web site*. The project Web site is maintained by the Center for Research Libraries. The site maintains links to all TICFIA-funded projects, as well as the Department of Education Web site. URL: <http://www.crl.edu/areastudies/ticfia/index.html>
5. *Technological Innovation and Cooperation for Foreign Information (TICFIA) Web site*. U.S. Department of Education. URL: <http://www.ed.gov/programs/iegpsticfia/index.html>
6. *Office of Post Secondary Education Web site*. U.S. Department of Education. URL: <http://www.ed.gov/about/offices/list/ope/dues/index.html>
7. *TICFIA Web site*. U.S. Department of Education. URL: <http://www.ed.gov/programs/iegpsticfia/index.html>
8. Murlin Croucher, "Digitizing and Making a Web Site for the Soviet *Letopis' Zhurnal'nykh Statei*, 1956-1975," *Fifth International Slavic Librarians' Conference* in Tallinn, Estonia, July 26-28, 2000. *Letopis' zhurnal'nykh statei*, 1956-1975, Web site. URL: <http://Webapp1.dlib.indiana.edu/letopis/about.jsp?lang=en&subject=croucher>
9. *ABBY Software House Web site*. URL: <http://www.finereader.com/>
10. *Regular Expressions for Java Web site*. Last updated on 21 October 2001. URL: <http://www.cacas.org/java/gnu/regexp/>
11. "XMetaL," *Corel Web site*. URL: <http://www.corel.com/xmetal/>
12. *Corel Web site*. URL: <http://www.corel.com>

13. *The Apache Jakarta Project Web site*. The Apache Software Foundation. URL: <http://apache.jakarta.org>

14. "Tamino," *Software AG Web site*. URL: <http://www2.softwareag.com/Corporate/products/tamino/default.asp>

APPENDIX 1

Project Presentations and Papers (Chronological Order)

- Spencer, George Andrew. "Digitization as a Means of Preservation of Russia's Literary Heritage—A Case Study: The Digitization of the *Letopis' zhurnal'nykh statei*." *Managing the Digital Future of Libraries Conference*. Moscow, Russia, April 18, 2000. URL: www.iis.ru/el-bib/2000/200004/spencer/spencer.en.html
- Croucher, Murlin. "Digitizing and Making a Web Site for the Soviet *Letopis' zhurnal'nykh statei*, 1956-1975." *Fifth International Slavic Librarians' Conference*. Tallinn, Estonia, July 26-28, 2000. URL: www.dlib.indiana.edu/collections/letopis/papers/croucherTallinn.html
- Spencer, George Andrew. "Digitization, Coded Character Sets, and Optical Character Recognition for Multi-script Information Resources: The Case of the *Letopis' zhurnal'nykh statei*." *5th European Conference of Research and Advanced Technology for Digital Libraries*. Darmstadt, Germany, September 4-8, 2001. URL for abstract: link.springer.de/link/service/series/0558/bibs/2163/21630429.htm
- Walsh, John. "Creating an XML-Based Scholarly Research Tool: Challenges and Methodology." *XML 2001 Conference and Exposition*. Orlando, Florida, December 9-14, 2001. URL: www.idealliance.org/papers/xml2001/papers/html/03-04-06.html
- Rome-Lindval, Natalia, and Brancolini, Kristine R. "Digitizing *Letopis' zhurnal'nykh statei* for the New Generation of Researchers." *The Impact of New Technologies on Teaching and Research in International Studies Conference*. Indiana Memorial Union, Bloomington, IN, October 13-14, 2002.
- Walsh, John. "'Unicode Ate My Brain,' or The Trials and Tribulations of Implementing a Unicode-Dependent Digital Library Project." *Digital Library Federation Fall Forum 2002*. Hilton Seattle, November 4-6, 2002. URL for slides: www.letrs.indiana.edu/jawalsh/presentations/uamb/

APPENDIX 2
Selected Russian-Language Journal Holdings, 1956-1975,
Indiana University Libraries

TITLE	SUBJECT	HOLDINGS
V mire knig	Publishing	1970-1979 1982-1988
Vestnik vysshei shkoly	Education	1946-1948 1955-1990
Voennyi vestnik	Military	1969-1994
Voprosy istorii	History	1945-2003
Voprosy istorii KPSS	Comparative communism	1958-1991
Den'gi i kredit	Economics	1960-2002
Zvezda	Literature & politics	1924-1926 1944-2003
Izvestiia Akademii nauk SSSR Institut etnografii. Trudy Institut iazykoznaviia. Institut istorii Institut literatury Institut narodov Azii Institut russkogo iazyka Institut russkoi literatury	Ethnography Linguistics History Literature Central Asian studies Russian language Russian literature	Various