This article describes the process of tracking down an obscure yet politically and culturally significant Russian theater review, “Сеанс черной магии” на Таганке” by N. Potapov, published in Pravda, the official organ of the Soviet Communist Party from 1918-1991. What at first appeared to be a routine reference question turned into a journey full of twists and turns, a “quest” for the “missing review.” The purpose of this article is threefold: to highlight some of the challenges that librarians and researchers face when working with Russian, East European, and Eurasian materials; to inform researchers about the capabilities and limitations of the Pravda Digital Archive and similar databases (i.e., Izvestiia Digital Archive, Literaturnaia Gazeta Digital Archive) produced by East View Information Services; and to encourage discussion about the preservation of and long-term access to research materials from the region.

KEYWORDS Bulgakov, East View, Liubimov, Margarita, Master, microfilm, optical character recognition (OCR), play, Potapov, Pravda Digital Archive, preservation, reference question, Russia, Soviet, Taganka, theater, theatre, Булгаков, любимов, мастер, маргарита, Правда, Таганка, театр.

1 Address correspondence to Jon Giullian, Librarian for Slavic and Eurasian Studies, Watson Library, University of Kansas, 1425 Jayhawk Blvd., Rm. 519, Lawrence, KS 66045, USA. E-mail: giullian@gmail.com
Every now and then librarians uncover hidden treasures from the past. These discoveries that make us feel a little like archeologists, detectives, or explorers from bygone years. Sometimes the treasure is large in scope, as with the cache of pre-revolutionary books and magazines recently found in the Russian State Polytechnical Museum Library in Moscow, Russia. Deep within the bowels of the library’s archive, librarians stumbled upon approximately 30,000 volumes hidden behind a series of false walls as the library was preparing to renovate of the 100-year-old building.1 Other treasures are small in size but rich in tradition. For example, I recently tracked down a signed first edition of Bunin’s novel, Zhizn’ Arsen’eva: Istoki dnei, based on a tip from a friend and former faculty member who years before had come across the item inconspicuously sitting on the shelves of Watson Library at the University of Kansas (KU). In my own small way I felt like the librarians who had uncovered the hidden collection in the Russian State Polytechnical Museum Library. Added to the excitement of finding something old that was lost is the anticipation of the story that will unfold as new details are uncovered. This copy of Bunin’s novel includes a dedication from Bunin to his dear friend (Дорогому И. М. Тол... от автора 7.VIII.1950, Grasse, a. m.), and while additional details about the journey of this book are still being researched, the discovery of Bunin’s signature reveals why working with collections is so interesting and rewarding.

Another rewarding aspect of library work is the opportunity to help others make discoveries. Librarians enjoy helping transform a student’s frustration into eager anticipation as useful sources are found together. When a student walks into a librarian’s office, discouraged at being unable to find that elusive source and later walks out with a smile, document in hand or on digital device, it is a rewarding experience. One such experience took place in the spring semester of 2012 when a graduate student2 from the Center for Russian East European and Eurasian Studies (CREES) asked for help in tracking down a review of Iurii Liubimov’s 1977 stage adaptation (postanovka in Russian) of Mikhail Bulgakov’s novel Master and Margarita. She was researching various adaptations of Bulgakov’s novel for her REES capstone seminar paper and stumbled across a reference to a review of Liubimov’s stage adaptation that was allegedly published in Pravda (Moscow) sometime after the premier of the play. Her request for help in tracking down this review seemed an ideal opportunity to use the Pravda Digital Archive (hereafter Archive), a full-image digital version of Pravda (1912-2009) that was made searchable through “unmitigated” Optical Character Recognition (OCR).3 The Archive was produced by East View Information Services (hereafter East View or EVIS)4 and KU Libraries had purchased access to the Archive in 2011.

I was delighted to get a reference question that offered an opportunity to use the newly acquired database. Like a colleague of mine who had recently described her own adventure of using the Archive to track down an article for a patron, the student and I would soon embark on an adventure of our own. Because I had previously provided training to faculty and students on how to search the Archive and had featured the database on our departmental website and research guides, I was eager to demonstrate the functionality and value of the product. It would be the proof that justified the purchase. I presumed that the student’s request would require little more
than a quick search of the Archive using carefully selected keywords. But as I soon found out, like my colleague, the routine reference question turned into a mystery as the student and I attempted to track down an article that had seemingly vanished into cyberspace, as if by some trick of Woland and his retinue.\(^5\)

**The Quest Begins: Searching the Archive**

The student and I searched the Archive using keywords: любимов, таганка, мастер, and маргарита in a variety of combinations. Because we did not know the exact date of the performance, we limited the search by date, surmising that the performance had likely taken place between 1 March and 30 June 1977. Initial queries retrieved two types of results, all of them “false hits”: \(^6\) (1) matches unrelated to the topic; and (2) matches tangentially related to the topic. In the second case, the terms were found in performance schedules of the Taganka Theater in Moscow, including performances of Liubimov’s production of *Master and Margarita*. The queries and their number of corresponding results are given below:

<table>
<thead>
<tr>
<th>Query</th>
<th>Matches</th>
</tr>
</thead>
<tbody>
<tr>
<td>&quot;мастер и маргарита&quot;</td>
<td>0 matches</td>
</tr>
<tr>
<td>любимов AND мастер</td>
<td>0 matches</td>
</tr>
<tr>
<td>таганка AND любимов</td>
<td>0 matches</td>
</tr>
<tr>
<td>мастер AND маргарита</td>
<td>1 match(es)</td>
</tr>
<tr>
<td>таганка AND мастер</td>
<td>7 matches</td>
</tr>
</tbody>
</table>

After the first three queries found nothing, the query for мастер AND маргарита retrieved one match on Page 6 of the 1977 April 30 issue of *Pravda* (Figure 2).

Figure 2

![Search results](image)

The occurrence of the term мастер in the PDF image of the newspaper (Figure 3 below), refers to the “ветеран труда, мастер Бобошко Василий Георгиевич” (lit. veteran of labor, master Boboshko Vasilii Georgievich) and to the “мастер спорта, Бобошко Ирина” (lit. master of sport, Boboshko Irina) which follows a few lines later.
Oddly, the results in the left column Search window (to the left of the article image in Figure 3) indicate no occurrence of the term маргарита in the document, even though we had used the Boolean search operator “AND,” which should have retrieved documents with occurrences of both terms. Margarita had either take flight or was somehow hiding “incognito.” Fortunately, the Adobe Acrobat Reader includes a text-search function to “Find a word in the current document,” which is located at the bottom of the left column Search window. Using the Reader’s text search function, we searched for маргарит (a truncated form of маргарита) and found the term маргаритой (an inflected form of маргарита) in the PDF image (Figure 4).
Although both terms, мастер AND маргаритой, occur on page 6 of the issue of Pravda (Figures 3 and 4 above), the two terms appear in different columns, neither of which relates to the review of Liubimov’s staging of Bulgakov’s novel.

The next strategy was to search for each term separately. As one might expect, the more general term мастер (master) appears much more frequently (199 matches) than the term маргариты (2 matches) for the specified date range. The results of the query for маргариты are shown in Figure 5.
The PDF image of Page 3 of *Pravda* for 1977-03-14 (Figure 6) shows another oddity. The left column Search window shows “0 documents with 0 instances” for маргарита. Once again using the function to “Find a word in the current document,” a search for the truncated form маргарит found the term маргаритовский, the masculine adjectival form of маргарита.

The results above triggered the question as to why the left Search column shows instances for мастер but not for маргарита. A close examination of the differences between the search terms and the results suggest that only exact matches appear as results in the left column Search window. Exact matches for мастер appear in the left Search column, whereas inflected forms of маргарита (e.g., маргаритой, маргаритовский) do not. Separate queries for each of the terms маргаритой and маргаритовский confirm this hypothesis. Instances of the term маргаритой (margaritoi) appear in the Results box of the left column Search window (Figure 7).
A query for маргаритовский yields a similar result, which demonstrates that, although the Archive’s search function accounts for inflected forms of a search term, the instance of the term does not appear in the Results box of the left column Search window of the Adobe Acrobat Reader unless that term is an exact match. The Adobe Acrobat Reader’s text-search function (“Find a word in the current document”), however, does not account for inflected forms of a search term.7

A close examination of the PDF images of Pravda in the Archive also reveals a discrepancy between the date appearing on the list of search results and the date printed on the newspaper. For example, the search results in Figures 2 and 5 display the date of 1977-04-30 (30 April 1977), but the PDF image of the printed issue in Figure 8 shows the date of 1 мая 1977 года (1 May 1977).
Although not a serious problem, the discrepancy should be addressed, not only to ensure accurate searches but also to facilitate accurate source citations. The query for та́ганка AND ма́стер yielded seven matches (Figure 9).
All seven matches (Figure 9) refer to performance times at the Taganka Theater as they appear in Pravda’s daily theater schedule (Figure 10). The date discrepancy between the search results and the printed newspaper persists, such that the dates listed in the results above precede the dates on the printed newspaper by one day. For example result #3 in the list above (Figure 9) shows 1977-06-04 (1977 June 4); whereas the date on the performance schedule (Figure 10) and the date at the top of the issue (Figure 11) both correspond to the 5th of June 1977 (5 июня 1977).

Figure 10

Figure 11
A close examination of the image in Figure 10 signals another apparent discrepancy. This image, the result of Search 5: таганка AND мастер, shows the phrase “мастер и маргарита’ in the left column Search window and on the Taganka Theater performance schedule. When the text from the Taganka Theater’s performance schedule is copied into a word processor (Figure 12), the Optical Character Recognition (OCR) software correctly recognized the phrase “Мастер и Маргарита.” For some reason, the OCR software did not recognize маргарита in Searches 1 and 4 (Figure 1) but did recognize маргарита when copied and pasted from the PDF image. This apparent discrepancy is puzzling.

Figure 12

Original text in Pravda (from Figure 9) | Text produced by OCR software
---|---
ТЕАТР ДРАМЫ и КОМЕДИИ на Таганке — в 12 ч. — Антимир; Вечер — Мастер и Маргарита. | ТКЛТР ДРАМЫ и КОМЕДИИ на Таганке -- и 11 ч — Антимир; вечер ~ Мастер и Маргарита.

In the example above, the OCR software demonstrates a 57% success rate, correctly recognizing 8 words (highlighted in bold) out of the 14 words and numbers in the theater schedule. It is unknown whether 57% is above or below the average rate of success for the Archive’s OCR, but from the researcher’s perspective, 57% is a mixed blessing when one considers that nearly half (43%) of the potential results are not captured. A statement from East View describes, in general terms, the range of accuracy of the Archive’s OCR:

“…the archive is full-image only, which includes PDF images that have been OCR’d to allow copying and pasting text. The accuracy of OCR ranges widely, from perfect in recent issues (which are 'born digital') to scattered (difficult-to-scan sporadic early years with poor 'best available copy' for imaging).”

The search capability of the Archive depends upon the ability of the OCR software to recognize the electronic text underlying the Archive, and the accuracy of the OCR depends upon the quality of the source from which the PDF image was created. In some cases the source was of poor quality. The result is that “what you search may not always be what you get” but rather what the system is able to recognize.” It is understandable that the OCR of the Archive was not intended to be perfect because it would have been prohibitively expensive to manually correct the thousands of pages of text of “unmitigated OCR.” In the absence of a full-text version, East View’s Pravda Digital Archive is currently the best available option for searching the contents of Pravda.

The process of searching for Pravda’s review of Liubimov’s stage adaptation, revealed several nuances vis-à-vis use of the Archive, which are highlighted below:

- The search engine does, indeed, support Boolean searching.
• The search engine accounts for inflected forms of words.
• Not all matches appear in the Results box of the full image PDF display.
• The full-image PDF display includes a link to the internal text-search function that is built into the Adobe Acrobat Reader software.
• Because the internal text-search function within Adobe Acrobat Reader does not have the ability to find multiple inflected forms of Russian words, truncated forms may need to be used in the internal text-search function.
• The date displayed in the search results is one day earlier than the date printed on the newspaper.
• Because the Archive is not a “full-text” database but rather a “full-image archive enhanced by unmitigated OCR,” other options or “work-arounds” may be necessary to find desired content.

The discovery of these nuances was useful, but they did not bring us closer to finding the review article of Liubimov’s production. Given the probability of more accurate OCR from later issues of Pravda, it seemed strange that the aforementioned queries of the Archive had not retrieved the review article about Liubimov’s production, unless the review was not published in Pravda between 1 March and 30 June 1977. A technical problem with OCR did not seem to be the culprit because our queries for мастер retrieved results. In fact, a query for мастер alone during the three month period (March 1 – June 30 1977) found 295 instances of the term, and one would expect the words мастер and маргарита to appear multiple times in a review of the play entitled “Мастер и Маргарита.” If the review was indeed published in Pravda, then it should have appeared in the results. Before expanding the date range or moving on to search a different newspaper, we decided to use the pre-digital era strategy of browsing issues of the newspaper, page by page.

A Change of Direction: Browsing the Archive

The main advantage of the Pravda Digital Archive is the ability to search the entire contents of Pravda, but in terms of browsing the pages of the newspaper, there is little difference between the microfilm copy and the digital copy, other than the medium for viewing. Whereas the digital copy is more convenient to use because it can be viewed on a desktop computer, a microfilm copy in some instances can be more readable than a digital version even though use of the microfilm necessitates coming to the library to use an expensive microfilm reading machine. Most of the problems of readability that librarians have encountered in the Archive deal with earlier issues of Pravda (i.e., 1930s, 1940s), a consequence of the print quality of the newspaper when it was microfilmed. For example, not long after the Archive became available, a library colleague related how one faculty researcher had to use their library’s microfilm copy of Pravda because the issues from the 1920s and 1930s in the Archive were illegible. Since the majority of the Archive was built by creating PDF images directly from microfilm, the quality of OCR depends upon the quality of East View’s microfilm copy, which itself depends upon the quality of
the newsprint that was used to create the microfilm. Better condition of the newsprint and better print quality in later issues of Pravda, such as 1977, produced sharper images on microfilm, which in turn generated sharper PDF images in the Archive. Given the improved print quality of 1977 issues, the OCR should have recognized keywords in a review of Liubimov’s production. In addition to readability problems of earlier issues in the Archive, there have been a few instances of missing articles or pages, which were probably missing from the paper copies of the newspaper at the time they were microfilmed. Although unlikely, it is possible that the issue containing the review of Liubimov’s Master and Margarita had gone missing before the microfilm was created.

Prior to browsing issues of the Archive we needed to know the date of the premier of Master and Margarita at the Taganka Theater. A brief search of the World Wide Web placed the premier on 6 April 1977 at the Taganka Theater of Drama and Comedy (Театр драмы и комедии на Таганке). It seemed logical to begin browsing issues of the Archive from 7 April 1977. Unfortunately, the review did not emerge in the first few issues following the premier, and the likelihood of finding the review seemed to diminish with each additional issue, since one would expect a review to appear immediately following or soon after a play’s first performance. The absence of the review in the issues of Pravda that followed the premier seemed to cast a shadow of doubt on the effectiveness of browsing and even called into question the existence of the review. The student and I, therefore, put the brakes on our browsing session, shifted gears, and headed in a different direction. Searching the World Wide Web for information about a Pravda attack on Liubimov’s play, we stumbled across an article in the 3 August 1977 issue of the Montreal Gazette (hereafter Gazette), entitled “Daring Soviet Play survives criticism to become hit” (Figure 13).

Figure 13

The Gazette article about Liubimov’s success mentions a review of Liubimov’s adaptation that appeared in Pravda toward the end of May: “The Soviet press was silent about the play for seven weeks. Then at the end of May, ‘Pravda’ attacked, saying the production distorted the realities of today’s Soviet life, and panning the play with such words as “unjustified, subjective, vulgar, grotesque.”’ One can only speculate that Pravda’s editors needed extra time to craft a response and to get it approved before sending it to press, especially given the extraordinary popularity of Liubimov’s adaptation.
Even without advertising or newspaper review, the Taganka Theatre production of ‘The Master and Margarita’ was the hit of the just-concluded theatre season. The 650-seat hall was sold out so consistently that even the play’s director, Yuri Lyubimov, joked that he could not get a ticket…. But ‘Pravda’ stopped short of demanding that the play be removed, and the production continued through the remainder of the season. The only change some playgoers note was that the nude, who sits centre stage with her back to the audience, turned slightly to offer a slightly more risqué view…. Lyubimov told Western reporters he fought for three years to win permission from Soviet authorities to stage the play…. The fact that it was finally approved, then attacked, but then allowed to remain on stage is seen as an indication of disagreement among Soviet officials. Clearly there is high-level interest in the production, three members of the Soviet Politburo are reported to have attended the play, but this could not be confirmed. 14

One can imagine how the Taganka Theater and the public anxiously awaited the “official” response to the performance and the potential implications of their response. While a negative review was probably to be expected, a positive review might open the door to greater freedom in the theater and in other forms of art, and in society, etc. Indeed, “Soviet theatregoers predict this success may lead to more experimental theatre in the coming season.” 15

On the same page of the Gazette immediately below the article describing the success of Liubimov’s production, one cannot but notice two photographs and their accompanying articles (Figure 14). On the left one sees a large close-up photo of a young Barbara Walters, who is not faring well as anchor for ABC News.16 To the right of Walters is a large photograph of a young Sylvester Stallone as a tough labor leader in a new film.17 The juxtaposition of the article about a potential “revolution” in Russian theater atop two articles about American media and entertainment is intriguing if not insightful. The juxtaposition displays an extraordinary contrast of Russian and American culture on the page of a popular Quebec newspaper. American viewers dine on images of the working class success story of an American cinematic icon (Sylvester Stallone) and the rise and fall of the first woman (Barbara Walters) to ascend the throne of American network news. Meanwhile, on the other side of the world, Russian theater lovers, watching Liubimov’s stage production of Bulgakov’s Master and Margarita, dine on Bulgakov’s once forbidden fruit and are filled with visions of devilish mayhem over Moscow. While Russians taste theatrical freedom in the form of high-culture art, Americans consume popular media sensations. Although not of particular relevance to the task at hand, the contrast is striking even if it speaks more to the perception created by the juxtaposed Gazette articles than to an accurate reflection of cultural tastes.
The Gazette article about Liubimov’s success confirmed the fact that a review of Liubimov’s play was published in Pravda and appeared toward the end of May 1977. However, the reason why the review did not appear in our search results remained a mystery. Since the quality of the Archive was good enough for the OCR to find terms in the theater schedules, then why did these queries not find key terms that would invariably appear in a theater review of Master and...
Margarita? This paradox, befitting the story of Master and Margarita, left us dumbfounded—perhaps a trick of Woland and his entourage.

The only remaining option was to return to browsing issues of Pravda, this time beginning with 31 May 1977 and working backwards. It seemed logical that if the review did indeed appear in Pravda, it would appear towards the back of the issue, near the theater schedules, rather than on the first few pages. Scanning issues from back (usually page 6) to front, we discovered something odd about page 6 of the 29 May 1977 issue. Some of the text and columns did not line up correctly, as if the page had been folded or smashed together with the previous page. The two columns on the left exhibited a folded tear, and the photographs in the center of the page seemed to be torn along the bottom edge. The columns in the center-right were also distorted, and the distortion was uniform and parallel to the column (Figure 15).

Figure 15

It finally dawned upon us that a portion of page 4 was visible underneath page 6 where a large section of page 6 had been torn out of the newspaper (Figure 15). We clicked back to page 5, which revealed a large black square in the bottom right quadrant of page 5 (Figure 16)—presumably the missing review. The hole (пробел in Russian) had been preserved on the microfilm copy that was used to create the PDF images for the Archive.
Figure 16

According to East View, the Archive was created directly from their own microfilm holdings of Pravda. Dima Frangulov, Vice President of East View, describes how East View obtained copies of Pravda for scanning onto microfilm:

Starting circa 1992 we purchased copies of Pravda and filmed them in our Moscow office. Most of the back issues were filmed from the paper copies borrowed from Moscow libraries, such as Istorichka and Leninka.18 The black hole in the Archive (Figure 16) from East View’s microfilm copy of Pravda probably originated from paper copies at either the State Historical Library of Russia (Istorichka) or the Russian State Library (Leninka) in Moscow, as if Woland or one of his minions had slipped out from the Taganka long enough to extract the unflattering review from Truth (Pravda). Yet, it was inevitable that one day, far into the bright future the truth would prevail, a beam of light shining into a dark little corner of library and literary history that previously had been known to only a few and would now be visible to the world.

Although the student and I had discovered why the article did not appear in the search results, the text of the review remained elusive. Presuming (incorrectly) that other microfilm copies would have the same “пробел,” we did not check the microfilm copy of Pravda available in the KU Libraries. Rather, as any Gen-X-er would do, the student and I turned to the World Wide Web and after a short time found the text of the review by N. Potapov, entitled “Сеанс черной магии” на Таганке” (“Séance chernoï magii’ на Taganke). The review was reproduced...
on the Taganka Theater’s website at taganka.theatre.ru, the same website where we had discovered previously the date of the premier. 20 A link to Potapov’s review appears at the bottom of the list of links (сылки) on the Taganka’s webpage about Master and Margarita. Only later did I discover that KU Libraries’ microfilm copy of the newspaper, which was filmed at the University of Chicago by Regenstein Library’s Department of Photoduplication, preserves Potapov’s review intact in the 29 May 1977 issue of Pravda (see images in the Appendix). Given this fact, it seems likely that the microfilm copies of Pravda produced by other institutions (i.e., the British Library, CRL, Harvard, the Library of Congress, University Microfilms International (UMI), etc.) would also contain Potapov’s review intact.

Finally, our quest had come to an end. Review in hand, the student departed with a smile on her face, eager to continue work on her research project with the new piece of information we had found together. The journey went far beyond the simple database query that both the student and I had anticipated. Each new obstacle prompted new questions that demanded a new approach to the problem, a different search strategy. The crooked path to the review article was a journey of discovery, a valuable experience in asking questions and using the research tools at our disposal to answer those questions. The only question that remains is the question of why Potapov’s review of Liubimov’s staging of Master and Margarita was excised from the Soviet Library’s holdings of Pravda, remains a mystery, a hidden secret of the past, or at least the topic of another study.

The Moral of the Story

Research in every discipline comes with unique challenges, and every researcher can tell stories about the challenges they have faced. Our search for the “missing” review is an example of the type of challenges that researchers and librarians in Slavic and Eurasian Studies encounter. This story is an astonishing example of how “Soviet censors continue to impact our access to information today” and “proof that a fallen regime can have an unintended impact on access to information today, even where an archive appears to be ‘open.’”21

The moral of this story is not limited to just one. Several proverbs or sayings in English apply: patience is a virtue; good things come to those who wait, perseverance over defeat; never assume; all is not what it seems; one person’s trash is another person’s treasure, and, of course, if at first you don’t succeed.... Regardless of one’s preference, no measure of effort, patience, or perseverance, is sufficient without the preservation, acquisition, and curation of research materials. Future generations of researchers depend upon librarians to recognize the value of preserving research materials, sometime in multiple copies and multiple formats. Even libraries that currently have access to the Pravda Digital Archive should not discard the microfilm copy. Both types of media hold value as important tools of discovery and preservation. Both types can and will be used, each for its specific advantages. The search capability of East View’s digital archives of Pravda, Izvestiia and Literaturnaia gazeta, (hereafter Archives) enables researchers to analyze content in these sources in new and different ways that are impossible to do with the microfilm or paper copies. Although the unmitigated OCR of East View’s digital archives
mentioned above would preclude specific linguistic or textual studies that depend upon the ability to find all occurrences of words, these digital archives will probably be sufficient to find most articles needed.

Unfortunately, access to the Archives is not universal. Researchers whose institutions cannot afford the digital version may be forced to depend upon microfilm copies that can be borrowed through interlibrary loan. Moreover, a few researchers may have to consult a microfilm copy depending upon the digital Archives’ print quality. As a long-term preservation copy, the microfilm might prove to be superior. Any provider of digital content is only one generation, one merger, or one financial crisis away from vanishing. Although one solution may be the “dark archive” for digital objects, even digital experts admit that for the present time the long-term viability of the dark archive is still unknown. Today’s “dark archive” could be tomorrow’s black hole, consuming whatever gets pulled into the vortex. Given the fast pace of IT development, in which digital devices and the software that run them become obsolete seemingly overnight, there is no guarantee that the digital text of the Archives will even be readable one-hundred years into the future or whether there will be devices around to read the text. A possible solution, at least for the present time, would seem to be the LOCKSS (Lots of Copies, Keep Stuff Safe) program based at Stanford University. It will be expensive, but essential, to continually migrate content to new digital technologies as they emerge, in order to ensure that content remains accessible indefinitely. Only time will reveal the best format for preservation.

The long-term preservation of digital content is an important question. It is a question that lies beyond the scope of this paper and far beyond the expertise of this author. The long-term preservation of digital content, nevertheless, is a dilemma that librarians grapple with as they struggle to decide how to allocate dwindling funds for the acquisition of research materials in a variety of formats. The significance of librarians’ acquisitions decisions will only become clear to future generations. The decisions that librarians make today will, to some degree, shape the research that can and will be done in the future. These decisions will determine how effectively librarians of tomorrow will be able to respond to requests for research assistance that makes the profession so much work and so much fun.

Response from East View Information Services

East View digitizes a variety of materials, using a variety of workflows. They are not all uniform in approach or output, and the methodology and selection for each project (as with, say, a binding, remote storage, or reformating decision in a library) is largely determined by the scope of material addressed and its likely patterns of use, as well as the size/nature of the audience who will be using the content. In the case of our recently-released Iskusstvo Kino Digital Archive, every word of every page is manually corrected, and so this complete archive 1931-2011 can truly be described as “full-image and full-text,” because the accompanying text is manually reviewed and corrected, and warranted to be ‘perfect.’ For IK, this is in part possible because of the smaller volume of output and proportional cost of that approach. By contrast, the nearly 100-year archive of Pravda is immense and dense. It was only recently published as a digital archive, because until
very recently the means of production were prohibitively expensive. (Earlier estimates suggested it would require a price tag of $50,000+ per customer institution to support, which is well out of the reach of many libraries, particularly for a resource in Russian, which directly limits the size of readership at any particular school). Eventually, costs came into a favorable juncture with a reasonable, hybrid approach: scan (mostly) microfilm to form the most complete possible full-image archive. Add “unmitigated” OCR to allow at least a relatively high degree of searchability to aid discovery of content. It is certainly “better than nothing” and in most cases, quite useful. We have been careful to never characterize DA-PRA (nor DA-IZV or DA-LG) as “full-image and full-text” because of these transparently emphasized limitations. We have declared them, appropriately, as “full image with available full text” and have clarified consistently that the process for the creation of these specific archives was to focus primarily on offering a maximally exhaustive, maximally legible full-image archive, and to enhance it with the best possible unmitigated OCR. This enabled us to produce and price the archive in such a way that more than 80 institutions were able to acquire it, in many countries. So, “pound for pound” (or, perhaps, “byte for byte”) it can be argued that the tradeoffs of an exhaustive qualitative approach must be weighed vs. the more temporal realities of size, effort, and cost. In essence, this approach has allowed more people to have more access to this corpus of content, than if the presentation was without these limitations (but then made inaccessible to most institutions and users because of prohibitive cost).

Your article identified a very interesting case, indeed. Although our QC process was multi-staged, and has involved the replacement of defective pages (including, it should be noted, a “report a problem” link, which allows libraries, like KU, to submit their own replacement pages in cases when their own local holdings or other known holdings are discovered to contain better copies of individual pages or issues to replace flaws or lacunae in the EVIS product), this case clearly went undetected. If you are gracious enough to lend us the use of your reel containing the Liubimov article, we’ll be more than happy to replace our own defective page, which will make all users “whole” in this matter.

The pagination error you identified is certainly a mistake. This happens at times (when processing over a quarter of a million pages for a single newspaper, and one whose print originals often bear mistakenly repeated date stamps and other typographical flaws, such errors naturally can occur). We appreciate your flagging this one, which we’ll promptly correct. Databases are more akin to live television than to books. They truly are “broadcast” and can fairly be understood to be in a perpetual case of minor flux. Errors can be fixed; at times electronic corruptions or interruptions in service can afflict proper use. Such is the state of the thing. We’re glad in this case, and the case of the пробел you stumbled on, that once a correction is made, it is immediately promulgated for all users. We’ll fix this.

Please be careful to draw a clear distinction between the function and features of East View’s UDB search engine (the interface’s main search boxes) and the search limitations of Adobe Acrobat. In several places, your article refers to “the search box at left” and similar phrases. Please note that at the page-level, there are two search boxes, with different functionality.
One is the UDB’s own (at top, outside the frame of the page). The other presents the search functionality of the Acrobat Reader browser plugin itself. The pages are presented in PDF, and within your browser, via the Acrobat plugin. That plugin is not authored or controlled by EVIS, of course, nor is it actually “preconfigured” into the database. Instead, when accessing the database, the user is required to install the Acrobat plugin if it is not already present. The advantage of this approach is that Acrobat is the most ubiquitous, free means for making PDF content viewable within a browser. A disadvantageous aspect, however, is that there is variability in each user’s computing environment. For example: whether a current and updated version of the plugin is installed, on which browser, and on which OS. More tangibly, the settings on the user’s own PC determine the behavior of several aspects of the plugin and how it behaves (default magnification, ability to marquis zoom, etc.). And in this specific case/content, the most important difference is that the EVIS UDB is, of course (and, as noted in your article) painstakingly optimized for Russian-language searching, including the ability to find multiple inflected forms of Russian words, with a pre-programmed understanding of Russian morphology as well. This is not true in Acrobat (nor, it is worthwhile saying here, in most Federated Search solutions). A search for Russian text in Adobe’s search box requires the savvy user to adopt conservative strategies (wildcarding before known morphological breaks, for example: searching for ‘пиро*’ instead of ‘пирог’ to ensure hits for forms such as ‘пирожное’, etc.).

East View very much agrees with your assessment of the continued role of microforms in the overall library mix. And that is why we still produce and offer it also!

... 
Kind regards,
Robert E. Lee
Director, Online Publishing
robert.lee@eastview.com
+1 (952) 252-4522
EAST VIEW INFORMATION SERVICES, INC.
10601 Wayzata Blvd, Minneapolis MN 55305-1526 USA
Office: +1.952.252.1201 | Fax: +1.952.252.1202
www.eastview.com
Appendix
«СЕАНС ЧЕРНОЙ МАГИИ»
НА ТАРАНКЕ
Театр

Прочитал повесть Эдуарда Успенского «Сеанс черной магии» и не мог не восторгаться его художественными возможностями. Успенский мастерски рассказывает о мирной жизни москвичей, о том, как они живут, думают, любят и смеются. Он пишет о людях, о их судьбах, о том, как они справляются с трудностями.

В этой повести Успенский использует элементы магии и фантастики, чтобы создать атмосферу загадочности и таинственности. Он делает это так мастерски, что читатель не может не быть вовлечен в происходящие события.

Вероятно, это связано с тем, что Успенский сам является ценителем магии и фантастики. Он хорошо знает, как создать атмосферу загадочности и таинственности, и удачно использует это в своей повести.

В целом, «Сеанс черной магии» — это удивительная книга, которая вдохновляет и заставляет задумываться о жизни и ее загадках. Каждый читатель сможет найти в ней что-то свое.

2 Thanks to Sarah Bazih for permission to use her name in this article.

3 OCR is a “technique for inputting text to a computer by means of a document reader. First, a scanner produces a digital image of the text; then character-recognition software makes use of stored knowledge about the shapes of individual characters to convert the digital image to a set of internal codes that can be stored and processed by a computer.” See p. 396 in Hutchinson Dictionary of Computing and the Internet (Abingdon, Oxfordshire, GBR: Helicon Publishing, 2005), s.v. “Optical Character Recognition,” accessed 26 April 2013 in Ebrary Academic Complete. The reader should bear in mind that the Archive was never intended to be a "full-text" database because of the exorbitant cost of manually correcting the OCR of such an immense and dense newspaper. Rather, the publisher's goal was to create a “maximally exhaustive, maximally legible full-image archive and to enhance it with the best possible unmitigated OCR.” For more details see the Response from East View Information Services (EVIS) which has been appended to the end of this article.

4 Thanks to East View Information Services (East View or EVIS) for answering questions about the quality of OCR and to Dima Frangulov, Vice President of EVIS, for identifying the likely print source of the EVIS microfilm copy that was used to create the Pravda Digital Archive. For more information about the Pravda Digital Archive, see http://www.eastview.com/Files/EastViewPravdaDigitalArchive.pdf, accessed 15 March 2013.

5 Woland and his retinue (Azazello, Abadonna, Behemoth, Hella, and Koroviev) are characters in Mikhail Bulgakov’s novel, the Master and Margarita. Woland, in contrast to the devil in Goethe’s Faust, is not the personification of evil but instead plays a somewhat ambiguous role in the novel, acting as both an architect of mischief and as an agent for good. Operating within the plane of the fantastic, Woland and his entourage execute a variety of tricks that torment Moscow’s literary apparatchiki. For more information, see pp. 202-203 of Reference.
“False hits” can be defined loosely as irrelevant results of a query or results in which the designated term or terms appear but which are unrelated to the topic. In the case above, the results of the query for мастер AND маргарита is the clearest example of a “false hit” because the articles in which instances of the terms appear are completely unrelated to the object of the query—the review of Liubimov’s production of Bulgakov’s novel Master and Margarita. The notion of “false hits” is a little more subtle in the results of the query for таганка AND мастер because some of the results at least relate to the topic in a general sense (i.e., these results identify the Taganka Theater’s schedule for Liubimov’s production of Master and Margarita); A “lighter shade” of false hits, these results also do not include the actual object of our search—the review of Liubimov’s production.

7 Please refer to East View’s Response for a discussion of the distinction between the Archive’s search engine and the text-search function of the Adobe Acrobat Reader. Because the text-search function of Adobe Acrobat Reader does not have the ability to find multiple inflected forms of Russian words; truncated forms of words should be used in the Adobe Acrobat Reader’s text-search function.

8 See East View’s Response for an explanation of the pagination error, which they readily agree to correct and for which they express appreciation for identifying the problem.

9 Email from Catherine Jansen to Jon Giullian, 3 September 2010, in the author’s possession.


13 Ibid.

14 Ibid.

15 Ibid.


17 The second article poses the question: “How, he was asked, is it possible ‘Rocky’ has earned $110 million in six months and that he is still broke… even though Stallone owns 10% of the film’s profits?” Scott Vernon, “Rocky road to Oscar didn’t change Stallone,” Montreal Gazette, 3 August 1977, accessed 15 March 2013, http://news.google.com/newspapers?id=IAsyAAAAIBAJ&sjid=1qEFAAAAIBAJ&pg=2388%2C528256. Is it any wonder this theme finds its way into a Rocky sequel?

18 Email from Dima Frangulov to Jon Giullian 27 April, 2012, in the author’s possession.

19 Историчка (Историчка) is the nickname for the State Historical Library of Russia (Государственная публичная историческая библиотека России – Gosudarstvennaiia publicheia istoricheskaia biblioteka Rossiia). Leninka (Ленина) is the nickname for the Russian State Library (Российская Государственная Библиотека имени Ленина – Rossiiskaia Gosudarstvennaiia Biblioteka imeni Lenina). See also note 4.

20 N. Potapov, “Teatr na Taganke [Spektakli]: ‘Seans chernoi magii’ na Taganke,” Pravda, 29 May 1977, as reproduced on the Taganka Theatre website, accessed 15 March 2013, http://taganka.theatre.ru/performance/master_margarita/4161/. Potapov’s review appears on the Taganka Theater website and is linked to the web page devoted to the theater’s production of the Master and Margarita. Go to http://taganka.theatre.ru, then click on Спектакли > Мастер и Маргарита > then scroll down to the bottom of the page to the list of links. Potapov’s review is the second link from the bottom.

21 Comments from reviewer number two who reviewed this article for SEEIR.

22 For example, some in the tech-world speculate that the iPhone 3S will soon become obsolete as the resolution standard for the next “generation” of iPhones climbs higher and as new apps are designed for the new standard.
Furthermore, the iPhone 3GS will not be compatible with iOS 5. These two factors foreshadow an early death for the iPhone 3S. See: Jon Dick, “iPhone 3GS’s days are numbered?” The TechBlock, 8 February 2012, accessed 15 March 2013, http://thetechblock.com/iphone-3gss-days-numbered. See also Bryan M. Wolfe, “The iPhone 3GS Might Not Play Nice with iOS 5 (Or At All),” 24 March 2011, accessed 15 March 2013, http://appadvice.com/appnn/2011/05/days-iphone-3gs.

23 See http://www.lockss.org

24 Upon completion of this article, both the Editor and I thought it would be a good idea to invite East View to comment on the article. Thanks to Robert E. Lee, Director of Online Publishing at East View for his thoughtful analysis, explanations, responsiveness to user feedback, and continued efforts to improve the Pravda Digital Archive and other electronic products produced by East View.