

ProQuest Historical Newspapers: Page-level and Article-level Compared

Background

When the ProQuest Historical Newspapers program was launched over 15 years ago, all newspapers were digitized at the article-level. This means that every page of newspaper content was “zoned” into its distinct articles and other component parts (editorials, advertisements, cartoons, etc.), and each of those component parts was then run through OCR and treated as an individual entity in the database. Beginning in 2016, some new historical newspaper titles in the ProQuest Historical Newspapers program are being digitized at the page-level. This means that the full-page image is run through OCR, and the full page of content is stored in its entirety in the database. This article examines the differences between article-level and page-level digitization in searching, search results, and content display.



Article-level Zoning



High-Resolution Page-level Image

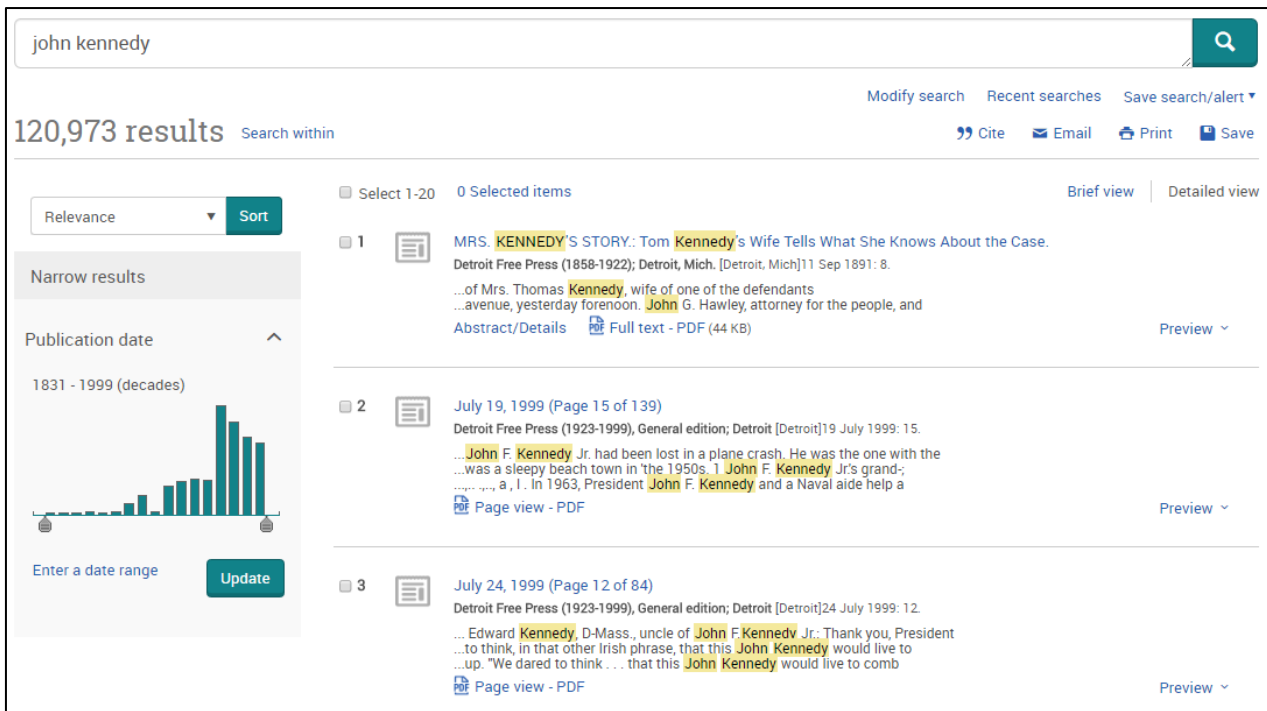
Differences in Searching

The basic searching function is identical for both article-level and page-level. **Every part of every page of ProQuest Historical Newspapers is full-text searchable, whether they are digitized at the article-level or page-level.** If you search for a term such as the name “John Kennedy” and it appears in the OCR text any place on a page, it will generate a hit for that page—whether it is in an article title, in an article, in an advertisement, etc. Further, both article-level and page-level titles are fully cross-searchable with all other historical newspapers, contemporary

newspapers, and any other content on the ProQuest Platform. The primary difference in searching article-level titles is in the Advanced Search: because they include article-level metadata, newspapers digitized at the article-level provide users with the ability to restrict search results to different portions of the newspaper (articles, advertisements, cartoons, etc.).

Differences in Search Results

The search results interface has been designed so both article-level and page-level results are presented side-by-side in an integrated and intuitive way. Most users are unaware that there is any difference between the following results:



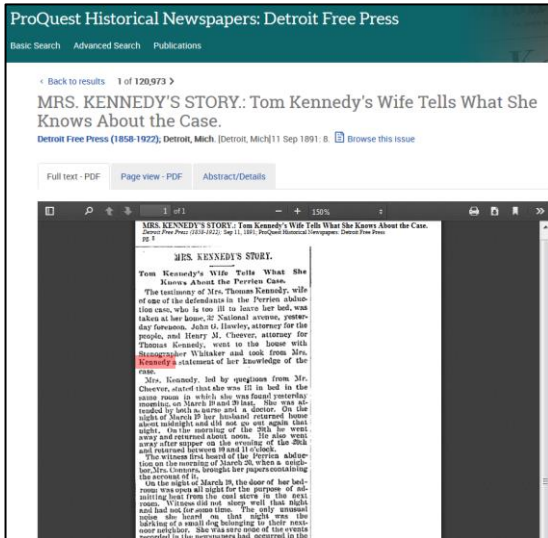
The screenshot shows a search interface for 'john kennedy' with 120,973 results. On the left, there is a 'Narrow results' sidebar with a 'Publication date' histogram showing a distribution from 1831 to 1999. The main results area shows three items:

- Item 1:** Article-level result. Title: **MRS. KENNEDY'S STORY: Tom Kennedy's Wife Tells What She Knows About the Case.** Source: Detroit Free Press (1858-1922); Detroit, Mich. [Detroit, Mich]11 Sep 1891: 8. Snippet: ...of Mrs. Thomas Kennedy, wife of one of the defendants ...avenue, yesterday forenoon. John G. Hawley, attorney for the people, and. Options: Abstract/Details, Full text - PDF (44 KB), Preview.
- Item 2:** Page-level result. Title: July 19, 1999 (Page 15 of 139). Source: Detroit Free Press (1923-1999), General edition; Detroit [Detroit]19 July 1999: 15. Snippet: ...John F. Kennedy Jr. had been lost in a plane crash. He was the one with the ...was a sleepy beach town in the 1950s. John F. Kennedy Jr.'s grand- ...up. In 1963, President John F. Kennedy and a Naval aide help a. Options: Page view - PDF, Preview.
- Item 3:** Page-level result. Title: July 24, 1999 (Page 12 of 84). Source: Detroit Free Press (1923-1999), General edition; Detroit [Detroit]24 July 1999: 12. Snippet: ... Edward Kennedy, D-Mass., uncle of John F. Kennedy Jr.: Thank you, President ...to think, in that other Irish phrase, that this John Kennedy would live to ...up. "We dared to think... that this John Kennedy would live to comb. Options: Page view - PDF, Preview.

In the example above, the first result is article-level title, and the second and third results are page-level. In each case the user is presented with context and clues that help them quickly determine their interest level in the result. The search text is highlighted to show the keywords in context. The difference is that the article-level result includes the article name, while the page-level result includes the issue date and page.

Differences in Content Display

When a user selects an article-level search result, that article image is displayed with the search term highlighted. When a user selects a page-level result, the entire page of content is displayed with the search term highlighted¹:



Article-level Result



Page-level Result

The article-level result is a bitonal (black and white) image displayed at 300dpi (dots per inch). The page-level result is a high-resolution greyscale image displayed at 400dpi, which provides almost photographic-like quality of the microfilm images. In both cases the images may be scrolled and zoomed as needed, and saved to the user's local storage.

¹ Hit-term highlighting currently requires that users have the Adobe PDF plug-in installed as the default PDF viewer for their browser. Firefox and Internet Explorer 8 have the Adobe PDF plug-in installed by default. This link describes the use of the Adobe Plug-in with various browsers: <https://helpx.adobe.com/acrobat/using/display-pdf-in-browser.html>

New Enhanced Page-Level Viewing Interface

The new page-level interface provides an enhanced browsing experience. When a user selects “Browse this Issue” they are presented with a new newspaper browsing interface optimized for page-level newspaper content. This interface features a scrolling thumbnail section at the bottom of the screen that enables users to quickly skim through a newspaper issue, and a highly intuitive interface that allows a user to manipulate the page image:

The screenshot displays a web browser window showing a newspaper page from the Detroit Free Press, dated Monday, July 19, 1999. The page is titled "6A DETROIT FREE PRESS" and "LEGACY LOST". The main headline is "Saddest words: It might have been" by Ron Dzwonkowski. The article discusses the death of John F. Kennedy Jr. and includes a photograph of him and his wife, Carolyn Bessette Kennedy, walking in New York City. The interface includes a search bar, navigation icons, and a scrolling thumbnail section at the bottom labeled "FRIENDS" and "QUOTABLE".

Page 15

ProQuest Historical Newspapers: Detroit Free Press

6A DETROIT FREE PRESS

LEGACY LOST

MONDAY, JULY 19, 1999

RON DZWONKOWSKI

Saddest words: It might have been

OH, NO. Not this Kennedy. A lot of hopes and dreams were dashed and too many painful memories were stoked with news that John F. Kennedy Jr. had been lost in a plane crash.

He was the one with the potential to atone for all the troubled Kennedys, the one who handled all the Kennedy attention with the most grace, the one who seemed capable of even becoming the president that history revealed his father never really was — and never had a chance to be.

With unlimited options in life, this young Kennedy seemed determined to be as regular a guy as possible, or at least as regular a very rich guy as possible. He worked for a living. He was an assistant district attorney in New York City — going 6-0 as a criminal prosecutor — and then became a magazine publisher, a strange leap made possible by his fame. But Kennedy did more than lend a presence to George, the monthly magazine which takes an often-irreverent look at politics as part of pop culture. He did inter-

views, wrote a column and edited George writers, all with absolutely no need of a paycheck.

He used in-line skates — arriving on wheels at his mother's apartment upon news of her death — and he threw a Frisbee around Central Park in New York.

Kennedy had his father's knack for disarming, self-deprecating humor and a remarkable tolerance for the media and the just plain folks who invaded his space. A political career seemed to be his for the asking. Kennedy always declined, but never said never.

Perhaps he was afraid, and justifiably so, because of what assassins did to his father and his uncle.

Still, I had a sense, especially from his work at George, that young John was studying the art of politics, getting himself ready to follow so many other Kennedys into public service. And it seemed he would do so with no skeletons in his corner of the family closet and no public pronouncements that would tie him to the raging liberal politics of the New England Kennedys. He would do it his way, on his terms, and he would do it with commitment and class.

The good looks would certainly be worth some votes. But he also would command the usual Kennedy vote, much of the young vote, and the votes of many baby boomers who, like me, remember his father as their president, even though they were, like me, too young to vote for him.

President John F. Kennedy came after President Dwight D. Eisenhower, which for American kids in 1960 was a little like trading in your grandfather in the black Dodge for your hula-hooping uncle with the red Thunderbird. I liked JFK before I knew there was any real difference between Democrats and Republicans. I liked what he said and the way he said it. I liked the way he always seemed to be pushing America to tackle tough issues and reach beyond its grasp.

We know now, of course, that JFK played the media like a violin, that there was a whole lot going on in the White House that would make even Bill Clinton blush. We know that JFK was not a well man and was probably in a little over his head, especially on foreign policy.

There was an unmistakable glow to John F. Kennedy that still overshadows his faults. There was leadership that you would follow anywhere, into a bomb shelter during the Cuban missile crisis, onto the segregated campus of a southern school, into space, even into the quagmire of Vietnam. I have not seen such leadership since.

Later would come the tarnish on his legacy, and the litany of tragedies and scandals that would generate such a wild mix of sympathy and shame for the Kennedy clan. But always there was JFK Jr., biding his time, giving us reason to think he would finally be the Kennedy to do the Kennedy thing right.

Instead, he is another Kennedy tragedy, taking with him perhaps the last best hope of rekindling the Kennedy magic.

It is fair to ask why such a promising young man who could afford to be flown anywhere he chose in absolute privacy and comfort would insist on piloting his own small plane.

The answer is that John Jr. was, even in his own reserved way, a true Kennedy — never content to just be along for the ride.

RON DZWONKOWSKI is editor of the Free Press editorial page. He can be reached at 313-222-8535 or rdzwonk@freepress.com.

John F. Kennedy Jr. and his wife, Carolyn Bessette Kennedy, stroll near their New York apartment in October of 1997. Kennedy managed to stay fairly private in spite of his famous name.

LAWRENCE LAVIND/Associated Press

FRIENDS

QUOTABLE

Searchable PDF Images

The last difference between article-level and page-level digitization is with the image itself. The new page-level images are all searchable PDF files, which means that when a user downloads the PDF file it can be manipulated using the local PDF interface. The downloaded PDF image can be searched again and again using the PDF Reader search function. The OCR text can also be copy/pasted into other documents:



He was the one with the potential to atone for all the troubled Kennedys, the one who handled all the Kennedy attention with the most grace, the one who seemed capable of even becoming the president that history revealed his father never really was and never had a chance to be.

Conclusion

While there are differences between the article-level and page-level digitization process, most users will not notice any difference in searching or search result interpretation. All portions of every page are full-text searchable regardless of the treatment. All ProQuest Historical Newspaper titles are fully cross-searchable with other historical newspapers, contemporary newspapers, and non-newspaper content on the ProQuest Platform. In addition, the new page-level images are high-resolution 400dpi greyscale searchable PDF files that offer end users exciting new capabilities.

	Article Level	Page Level
Full text searchable, every part of every page	Yes	Yes
Newspaper browsing	Yes	Yes
Cross searchable with other newspapers	Yes	Yes
Cross searchable with non-newspaper content	Yes	Yes
Initial view	Article	Page
Image type	300 dpi Bi-tonal	400 dpi Grey Scale
Article threading?	Yes	No
Content available for PAL ownership	Limited years	All years
Content Delivery Service available	Limited years	All years
Article level metadata searching	Yes	No
Open or closed?	Mixture	All closed
Pricing	Full list	20% less than article level for similar titles
Hit term highlighting	Most content	All content
Searchable PDFs	No	Yes

FOR MORE INFORMATION

Contact your [ProQuest Account Manager](#)