

Old News, New Tricks:

How AI Breathes Life into Newspaper Historic Headlines

By: Daniel Horowitz - Daniel@MyHeritage.com - Twitter: [@MyHChiefGen](https://twitter.com/MyHChiefGen) - Instagram: [horowitz_daniel](https://www.instagram.com/horowitz_daniel)

Every genealogist's dream is to find newspaper articles mentioning their families. Newspapers provide not just the facts, but also the story — filling in more details to paint a richer and more colorful picture of our ancestors' lives.

Unlike vital records like birth, marriage, and death (BMD) — which are usually strictly structured, containing only facts and no photos, and are often handwritten — newspaper articles are unstructured, providing details and narrative, and often include photos and graphics. They are also printed, which makes them much easier to read.

Newspapers were the social media channels of yesteryear. They contain announcements of all sorts of life events, including births, marriages, engagements, graduations, travels, court news, land disputes, local celebrities, advertisements, and many more. Each of these announcements can contain plenty of important information we can extract and add to our family research — but they're not always easy to find.

MyHeritage has millions of newspapers from all over the world in its database, thanks to important partnerships with genealogical societies, archives, libraries, and various repositories. This makes MyHeritage and its search engine (<https://www.myheritage.com/research>) the best tool for finding family members.

The screenshot shows the MyHeritage website's search interface. At the top, there is a navigation bar with the MyHeritage logo and links for Home, Family tree, Discoveries, Photos, DNA, and Research. Below this is a search bar titled "Search historical records" with the subtitle "Explore your family history in MyHeritage's database of 18.7 billion historical records". The search bar contains four input fields: "First and middle name(s)", "Last name", "Year of birth" (with a calendar icon), and "Place" (with a location pin icon). Below the search bar are several filter buttons: "ADD DETAILS", "Father", "Mother", "Spouse", "Death", and "+ More". To the right of the search bar is a dropdown menu with the following options: "Search all records", "Collection Catalog", "Birth, Marriage & Death", "Census records", "Family trees", "Newspapers" (which is highlighted), "Immigration records", "Hire a researcher", and "Census Helper™" with a "NEW" badge.

What type of records are you looking for?

A carousel of record categories with icons and record counts:

- Immigration & Travel: 202,790,540 records
- Newspapers: 2,404,820,292 records
- Books & Publications: 855,523,603 records
- Public Records: 1,069,982,173 records
- Schools & Universities: 276,819,816 records

Because MyHeritage has international coverage, it has content relevant to all countries. We're rapidly expanding the data sets with the ambition to grow the search engine over time to become the largest worldwide collection of historical documents available online.

The MyHeritage search engine contains records from all over the world, and with the location selector, you can automatically focus your search on the continent, country, or state that interests you. All results will be related to the place you selected, including both historical records and family tree profiles from this area.

OldNews.com

Moving forward, OldNews.com will serve as MyHeritage's focal point for historical newspapers. Existing MyHeritage users can log in to OldNews.com with their MyHeritage account credentials, and new users registering via OldNews.com can use the same login credentials to access MyHeritage. OldNews.com supports Two-Factor Authentication, for enhanced security.

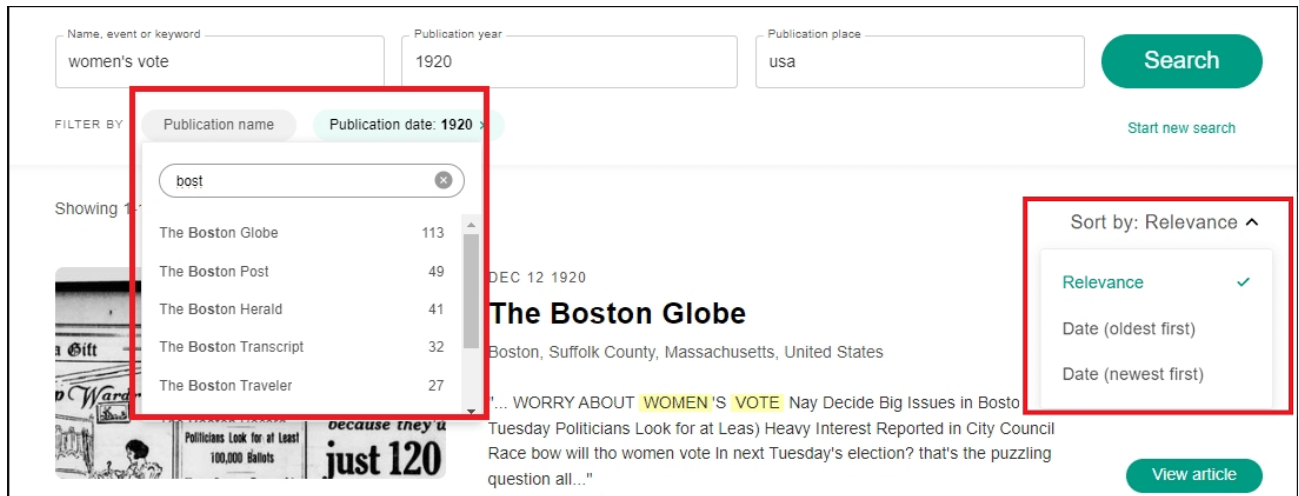


OldNews.com offers a wide range of publications from local, national, and international news, with extensive coverage of the 1800s and 1900s. The website includes newspapers from across the U.S., Canada, U.K., Austria, Germany, Czechia, the Netherlands, and Australia. Millions of newspaper pages are added each month, including additional countries. The OldNews.com website is available in 11 languages: English, French, German, Dutch, Danish, Norwegian, Swedish, Finnish, Italian, Spanish and Portuguese.

The new content on OldNews.com was scanned using best-in-class optical character recognition (OCR) technology and enhanced with sophisticated algorithms developed in-house by MyHeritage. The website also offers optimized performance through a new technical implementation where the content is served from a Content Distribution Network, through a point of presence closest to the user. This means that users from all over the world can access OldNews.com at an extremely fast speed.

We've enhanced the publication name filter so that once you enter at least 3 letters in your search, it will autocomplete the query and display a list of publication names that include your query. We've also added new sorting capabilities when viewing search results. From the search results page, click the new "Sort by" dropdown on the top right

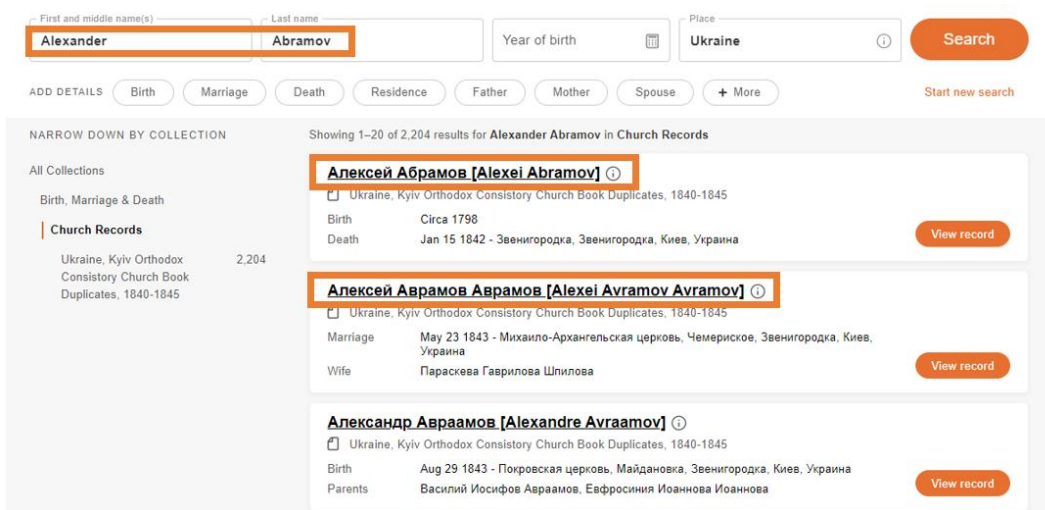
of the results. The default setting is to sort by relevance. You can also sort the results by publication date, either from oldest to newest, or newest to oldest.



Global Name Translation Technology™

This technology, unique to MyHeritage, helps break through language barriers in your quest to uncover your past. The technology automatically translates names found in historical records and family trees from one language into another, with very high accuracy, generating all plausible translations to facilitate matches between names in different languages. In addition, a manual search on MyHeritage's search engine will return results in other languages, automatically translated into your language.

For example, perhaps your American family has Russian roots. Previously, you would have needed to search in Russian to find all information available about your ancestors. Now you can search in English, and receive results in Russian, translated back into English for your convenience.



You can also view search results in Table view, which organizes the results in a more compact table, making it easier to scan.

You can run an advanced search to narrow the results. Using the docked search bar on the top of the page, you can add filters for events (i.e. birth, marriage, death, immigration, and others), details on the relatives of the person you are looking for, and keywords.

Sometimes you may not know the exact name of the person you're looking for, or that person may be found in historical records with a different spelling than the one you had in your records. Don't worry. When using the MyHeritage search engine you can control how you'd like it to search names. It can match the name exactly, or use the default setting that will automatically catch first name variations (such as Bill and William) and likely last name variations. Separate matching options are provided for first names and last names in the advanced search form. This lets you run exact searches with pinpointed accuracy, or cast a wide net to catch as many plausible variations as possible to find those elusive ancestors you've never been able to find before.

The search engine will even figure out the likely gender of the person you're looking for. So, if you search for a person called George, the search engine automatically infers that you are searching for someone male. Such automatic deduction of gender is among the many unique features exclusive to the MyHeritage search engine.

Once you receive results, you can narrow them down by collection or category, or add filters for additional names, places, and other common fields.

Keyword searches are often very useful because they cover many fields, so they can be used to look for details such as a town or a person's occupation. For example, you can use the keyword "Wedding" to search within the MyHeritage photo collection, and it will find photos with this word mentioned in the title or description of the photo. You can enhance this search by adding more criteria, such as places and dates.

Newspaper Matching

MyHeritage's unique technology applies semantic analysis to newspapers and understands what they are about. Obituaries, weddings, graduations, etc. are all automatically detected, parsed, and understood.

Most of the data in newspapers is indexed using Optical Character Recognition (OCR) technology that has some challenges of its own when you are trying to find your individual (Daniel Horowitz) even when it's written in different ways, for example:

- Horowitz Daniel (last name + first name)
- Danniell X. Horowitz (not the exact spelling and middle names or initials)
- Mr. Horowitz D. or Attorney D. Horowitz (title and last name only)
- Dan and Charles Horowitz (different but related names)
- Horowitz House Family Ties Two Dads A Father's Homecoming said Daniel

The quality of the paper, style of print, and quality of the digital copy can affect the OCR technology and make it harder for the technology to recognize characters. Among the most common OCR mistakes in English are:

- Capital R \Leftrightarrow B
- v \Leftrightarrow y
- c \Leftrightarrow e
- Capital D \Leftrightarrow O \Leftrightarrow 0
- Capital S \Leftrightarrow 8
- Capital S \Leftrightarrow 5
- Capital Z \Leftrightarrow 2 or 7
- Capital G \Leftrightarrow 6 or 8
- Capital B \Leftrightarrow 6 or 8
- Capital K \Leftrightarrow |<
- rn \Leftrightarrow m
- ar n \Leftrightarrow em
- h \Leftrightarrow b
- r \Leftrightarrow n
- [,] \Leftrightarrow l (el)
- nl \Leftrightarrow m
- en \Leftrightarrow em
- n \Leftrightarrow ri
- en \Leftrightarrow ar
- i, l, 1, /, ! are often interchanged

Be sure to consider other external factors when analyzing a newspaper piece: who submitted the information, who wrote and edited it, possible typos, and space available, to mention a few.