

Lithuanian Emigrant Newspaper Digital Archive

Jonas (Dr. John T.) Daugirdas

(Kristina Lapienytė, Administrator, Lithuanian Research Center)

Lithuanian diaspora

- Three main waves of emigration
 - **-** 1880**-**1920
 - 1947-1953
 - 1989-2000+
- Also to non-U.S. countries
 - Canada
 - United Kingdom
 - Australia
 - South America

Lithuanian emigrant newspapers

- These newspapers were quite numerous
- Various points of view
 - Communist/anarchist (especially first wave)
 - Laisvė (Freedom), Vilnis (The Surge)
 - Socialist
 - Keleivis (Traveler), Naujienos)The News)
 - Catholic
 - Katalikas (Catholic), Darbininkas (Worker), Garsas (Sound), Draugas (Friend), Vytis (Knight), Ateitis (Future),
 Aidai (Echoes), Eglutė (Fir Tree)
 - Other
 - Akiračiai (Horizons), Dirva (Soil), Vienybė (Unity), Tėvynė (Homeland), Saulė (The Sun), Skautų Aidas (Scout Echo), Mūsų Vytis (Our Knight), Lietuvių Dienos (Lithuanian Days), Lietuviai Amerikos Vakaruose (Lithuanians in Western U.S.).
- Non-U.S. countries
 - Britanijos / Europos Lietuvis (U.K.), Argentinos Balsas (Argentina), Mūsų Pastogė (Australia), Teviškės
 Aidai (Australia), Tėviškės Žiburiai (Canada), Nepriklausoma Lietuva (Canada), Mūsų Lietuva (Brazil).

How did this project start?



- Draugas (Friend) printed from 1909 m.
- Almost all issues from 1909 stored in binders at the Draugas publishing offices and at the Lithuanian Research Center.
- More than 230,000 pages!

 From 2008 m. original digital print files are available

Risk of losing the *Draugas* publishing building

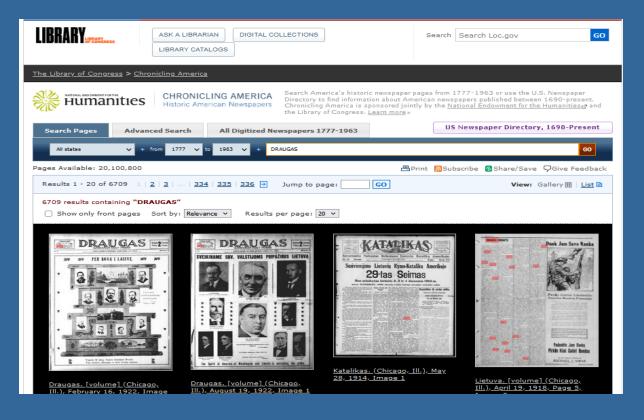
At the end of 2008, the Marian Fathers were considering selling the
 Draugas publishing building (they subsequently changed their minds in
 2009).



 The question arose: without the building, how could we secure the print archive of the newspaper?

Two government-run newspaper archiving programs 1. U.S. – Chronicling America

https://chroniclingamerica.loc.gov



The federal government (National Endowment for the Humanities) gives grants to State universities, who select newspapers and digitize them.

Chronicling America

Advantages

- High quality images
- Includes Optical Character Recognition (OCR), permitting copy-paste
- Set up for text-based searches

.

Disadvantages

- In the year 2008, Draugas had not been included in the list of newspapers processed by the University of Illinois Champaign-Urbana (the partner university in Illinois)
- Working with the University of Illinois Champaign-Urbana, we helped organize the digitization of *Draugas* via a supplemental grant to the NEH.
- but only from 1918 to 1924 and *Draugas* was printed in Pennsylvania from 1909 to 1912.
- The Chronicling America text-based search system is not very easy to use.

Two government-run newspaper archiving programs 2. Lithuania – Epaveldas.lt

Martynas Mažvydas Lithuanian National Library



Epaveldas scans not only newspapers, but also various printed materials, and also archives video and audio recordings.

Epaveldas.lt

Advantages

- High quality images, often scanned from originals rather than microfilm
- Very wide variety of newspapers scanned, published in both Lithuania and abroad.
- After copyright clearance, the Lithuanian National Library has been happy to hand over scanned images to the Lithuanian Research Center for digitization.

Disadvantages

- Images of pages only. No text. So, not searchable.
- For most U.S. published newspapers, the sets of issues are incomplete.
- Many issues that are archived at other libraries in Lithuania are not included
- Almost no issues published during the Second World War are available.

How digitization works

From microfilm

- Many U.S. libraries (including the Library of Congress) have microfilms of LT newspapers
 published in the USA, though onlyl for limited ranges of years.
- Strengths: Relatively cheap conversion to JPEG/TIFF files, then OCR (identification)
- Disadvantages: High contrast film was invariably used. <u>Very poor rendition of photographs</u>. Often, speckles and dark areas greatly reduce accuracy of OCR.

From originals

- Many libraries archive mainly originals.
- Strengths: Very high quality images of pages, including images of photos.
- Disadvantages: Some originals are very fragile. When photographing some extremely fragile originals, this may be the last opportunity to handle them.

Costs

- From microfilm: \$ 0.20 /page From originals: \$0.70 / page
- From previously scanned JPEG or TIFF images (e.g., Epaveldas.lt): \$0.20 / page (cost for doing OCR, combining page scans into issues, web-posting, etc.)

Where do we find these newspapers – 1?

Draugas (Friend)

 Draugas Publishing House (originals), US Library of Congress (microfilms), Centers for Research Libraries (originals and microfilms), Lithuanian Research Center (originals), epaveldas (scanned originals and microfilm).

• Dirva (Soil)

- 1943–2020: Dirva Publishing House (recent printer digital files), Cleveland Public Library (microfilm), Chronicling America (scanned microfilm).
- 1915–1942: Lithuanian Research Center (originals), Kaunas Technology University (KTU, originals), Lithuanian Literature and Folklore Institute (LLTI, originals), Klaipėda University (Pemkus archive, originals); Lithuanian Historical Institute (originals).

Ateitis (Future)

- 1911–1940: epaveldas.lt (scanned originals), Lithuanian Research Center (originals)
- 1946–1997: epaveldas.lt (scanned originals), Lithuanian Research Center (originals)

Where do we find these newspapers – 2?

Keleivis (The Traveler)

Boston Public Library, Lithuanian Literature and Folklore Institute (LLTI) -- originals

Saule (The Sun)

- Paul Coombe (Mahanoy Historical Society) 1. bought bound originals at a flea market!
 Saved originals strewn about in condemned building where Saule was once published!
- Vilnius University Library (originals); Kaunas Technology University Rare Document
 Section (originals); plus other libraries in Lithuania.
- ALKA (American Lithuanian Catholic Archive in Putnam, CT; originals)

Vienybė and Vienybė Lietuvninkų (Unity)

New York Public Library (microfilm), SLA (Susivienijimas Lietuvių Amerikoje [Lithuanian Alliance] – microfilm); various libraries in Lithuania, including the Lithuanian Historical Institute (originals).

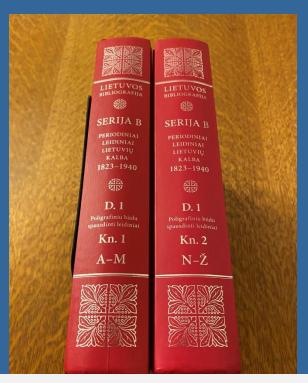
Where do we find these newspapers? – 3?

Lietuvos Bibliografija

 A list of ~130 libraries in the Baltics and Russia that have issues of Lithuanian language press in their archives.

E.g. "Dirva":

Egz(: C1 – 1925 (10); 1926 (1, 7, 9–11, 22, 25, 52, 53); 1927 (53); 2928 (1, 8); 1929 (36–52); 1930 (1–8, 10–17, 20, 22, 24–28, 30, 31, 38, 47, 48, 51, 52); 1931 (2–6, 8, 9, 13–16, 21–23, 25–31, 33–36, 42, 43, 45, 49–52); 1932 (1–14, 24, 27, 26–36); 1933 (2–11, 16, 17, 21, 23, 27, 30–33, 35–43, 45–51); 1934 (1–4, 6, 10, 12, 14, 16–19, 22–24, 26, 28, 31, 32, 34–36, 39, 40, 44–47, 49–52); 1935 (1, 2); 1936 (46, 52); 1937 (1, 2, 53); 1938; 1939 (1–33, 36, 38–47, 49–51); 1940 (1–5, 7–20, 22–27, 29, 30); C2 – 1918 (19, 25, 29, 30, 34, 38, 39, 46–52); 1920 (6, 8, 12–17, 20, 29, 34, 35); 1924 (11–31, 33–37); 1926 (2–9, 11, 12, 19–37, 39, 40); 1928



C1 = Lith. National Library Mažvydas C2 = Vrublevskiu library

Who is financially supporting the digitization project?

Draugas digitization

Lietuvių Fondas (Lithuanian Foundation) is the main donor

Oak Tree Philanthropic Foundation

Lietuvių Katalikų Mokslo Akademija (Lithuanian Catholic Science Academy)

Lietuvių Katalikų Religinė Šalpa (Lithuanian Catholic Religious Aid)

Ateitis Foundation

Tautos Fondas (Lithuanian National Foundation)

Individual donors

Who is financially supporting the digitization project?

Spauda.org (via Lithuania Research Center)

- Lithuanian Foundation (main donor)
- Lithuanian Catholic Science Acatemy (Catholic-oriented newspapers, including Amerika,
 Darbininkas, Tėviškės Aidai, Garsas, Mūsų Lietuva, Eglutė, Tėvynės Sargas, Vytis)
- Cleveland Foundation Pletra (Dirva)
- Canadian Lithuanian Foundation and Montreal Lithuanian Canadian Community
 (Tėviškės Žiburiai and Nepriklausoma Lietuva)
- Ateitis Foundation (Ateitis, Naujoji Vaidilutė)
- Australian Lithuanian Foundation (Mūsų Pastogė)
- Individual donors: Albinas Markevičius, Donatas Januta, Lembertų šeima, and Saulys
 Fondas -- (Lietuvių Dienos)
- Kazickas Family Foundation (Vytis, DP publications)

How does optical character recognition (OCR) happen?

Two-layer PDF files

- Bottom layer = image (JPEG)
- Top layer = text that has been interpreted from the image below.
- All PDF page images from a given issue combined into one larger PDF file. This is different from standard practice of one file for each page.

ABBY Fine Reader

- Company founded by Russian immigrants?
- Very accurate in terms of recognizing special Lithuanian letters with diacritical markings.
- Allows text recognition using more than one language simultaneously: e.g. Lithuanian +
 English, or for Mūsų Lietuva published in Brazil, Lithuanian + Portuguese.
- Disadvantages: Does not identify closely spacees newspaper columns; Sometimes
 (rarely) mixes up special Lithuanian characters such as ų vs. y, etc.; Difficulty interpreting
 larger headline and title fonts;. OCR training option does not work through version 15.

How is search implemented – 1?

Google webcrawlers

- Google webcrawlers automatically index double-layered PDF files containing text.
- When you upload a newspaper's PDF files to www.spauda.org or www.draugas.org, the indexed pages are automatically added to the Google Search database.
- Google Custom Search: Google allows non-profit organizations to create Custom Search
 pages where searches are limited to files that are located in a single "subdirectory". For
 example, search limited to *Draugas* issues published in the year 1934.
- Advantages: All of the information in the million pages of emigrant Lithuanian newspapers uploaded to www.spauda.org and www.draugas.org is now available to the whole world via Google Search (and other search engines as well).
- One major drawback: Google does not index all the pages submitted to it. (We routinely submit lists of URLs (in the form XML files) that point to uploaded PDF files to Google search for indexing).

How is search implemented – 2?

Elastic search

- A search program that can be adapted to an individual website
- Our programmers routinely index the files on www.draugas.org and www.spauda.org
- Strengths:
- 1) Full control of which PDFs are indexed avoids failure-to-index issue with Google.
- 2) Easy to specify different ranges of search years
- 3) Custom search form allows selection of subset of newspapers to search
- Disadvantages: Elastic search results do not appear on the public internet

Elastic search – three custom search form starting points

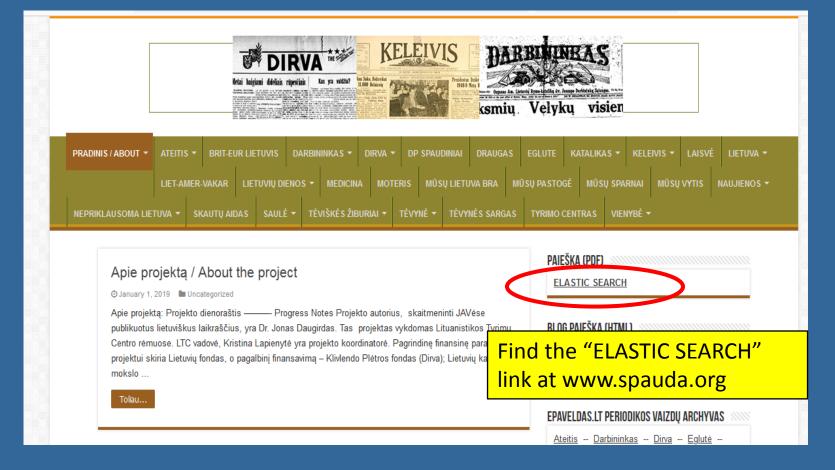
- draugas.org/elastic-search DRAUGAS and KULTŪRINIS PRIEDAS (The Friend and its cultural supplement)
- draugas.org/news/elastic-search DRAUGAS NEWS and LITHUANIAN HERITAGE
- spauda.org/elastic-search all other newspapers uploaded to www.spauda.org

How is search implemented – 3?

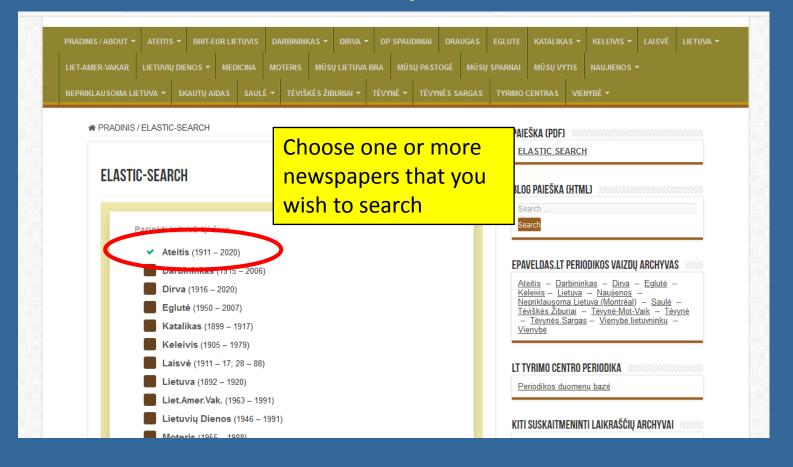
- Two steps required:
- Step 1) Find those PDF files (newspaper issues) that contain the inputted search terms.
- Step 2) Within each PDF file, find the exact location (page, column, line) where the search term appears.

- Example: Step 1.
- Choose a newspaper: e.g. ATEITIS (Future)
- Choose a desired time interval: e.g., 1911–1946

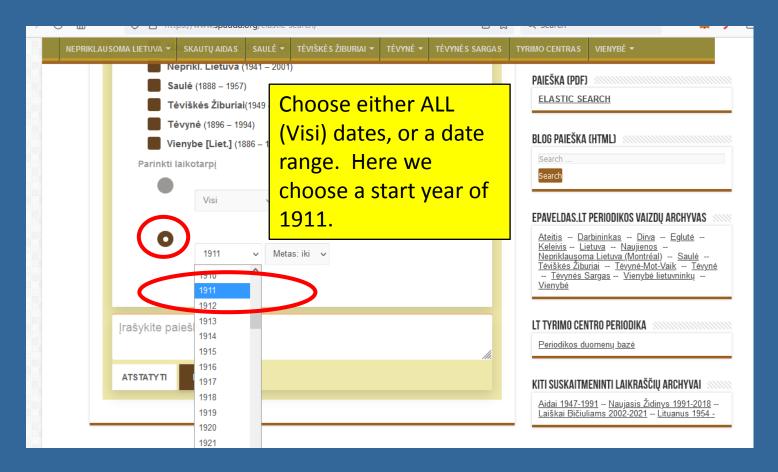
How is search implemented – 4?



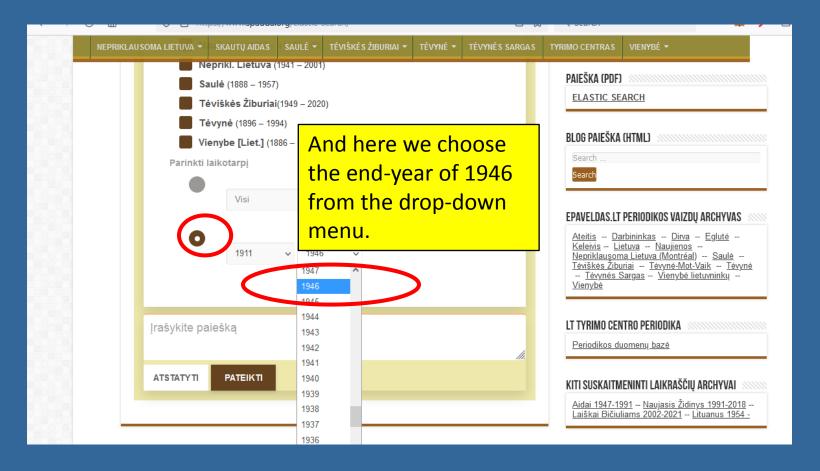
How is search implemented – 5?



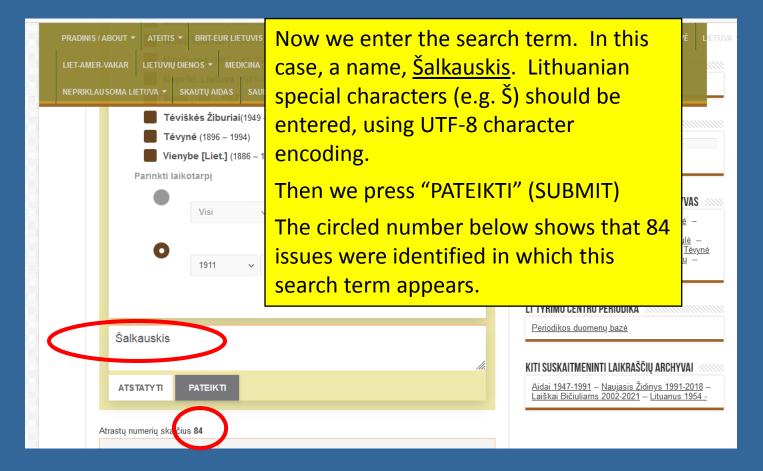
How is search implemented – 6?



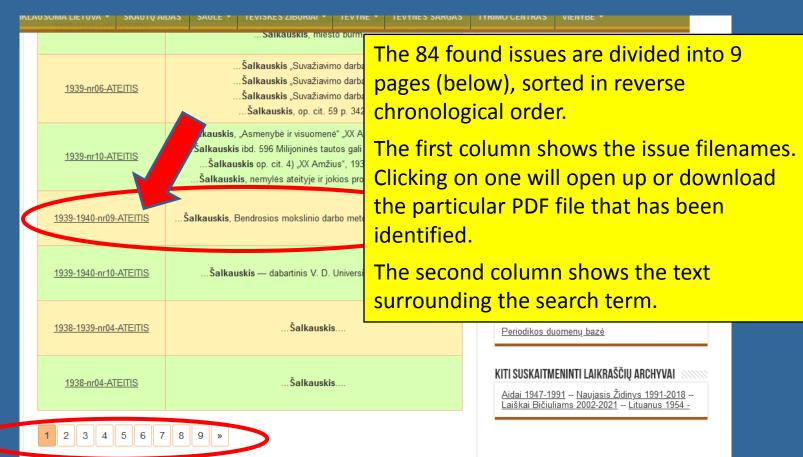
How is search implemented – 7?



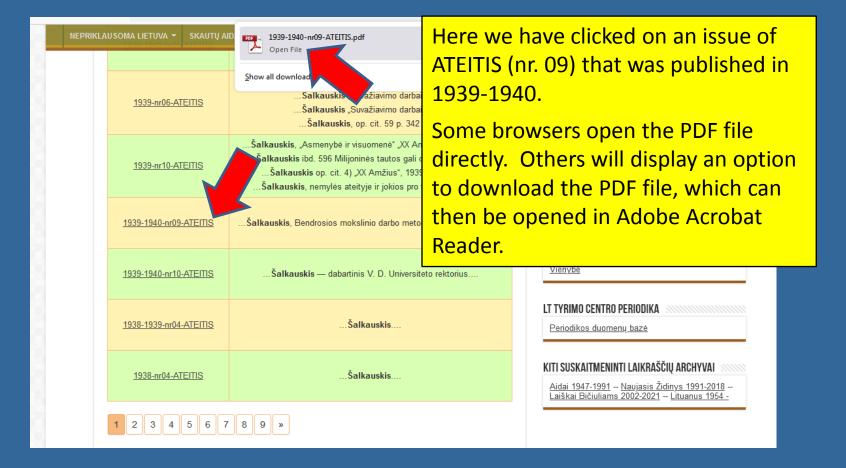
How is search implemented – 8?



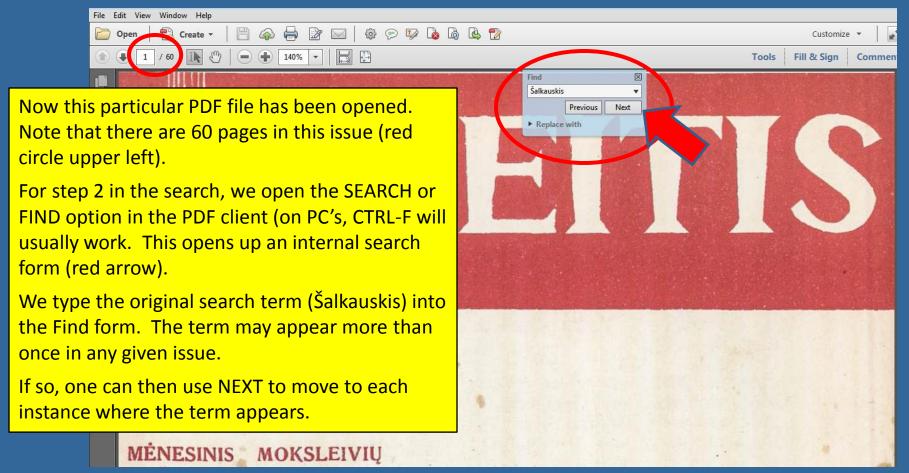
How is search implemented – 9?



How is search implemented – 10?



How is search implemented – 11?



How is search implemented – 12?

3. "Bepigu jiems — man tr Salkauskis · u". Taip save teisina, apgaudinėja tie, kurie niekam resperious Next ko nesiekia. Pamylėk darbą, kaip lakūnas myli erdvę, Replace with Find next (Ctrl+G) nubrauksi šį priekaištą. Gyvenimą kuria ir nugali ne ištyžę žodžiai, bet darbai. Britu admiraliteto pirmasis lordas Churchillis, prisimines dabartinius vokiečiu grasinimus ir gasdinimus, sako: "Jei žodžiai galėtu užmušti, britai jau seniai būtų negyvi". Vienais žodžiais neužmušim studiju kliūčiu. "Apskritai, reikia žinoti, kad gabumai galutinoje saskaitoje sveria gyvenime, ir veikime mažiau, negu pastovus, patvarus, samoningas ir metodingas darbas" (Prof. St. Šalkauskis, Bendrosios mokslinio darbo metodikos pradai, 52 psl.). Lietuviai gabūs, tik reikia išmokti Here the searched for text is highlighted

4. "Ten tai gerai — where it appears. In this case, the term altorius, ant kurio žmog appeared only once in the entire issue. bei" (Sailer). Ten gera, kur musų nera! Kiekvienas darbas, kurį žmogus dirba su meile, atsidėjimu ir nuoširdumu, reikalauja jėgos, kantrybės, išsižadėjimo.

Štai mintys, kurios turi lydėti jaunuoli, sprendžiant pašaukimo

How is search implemented – 13?

Fill & Sign Comme 3. "Bepigu jiems — man trūksta gabumu". Taip save teisina, tie, kurie niekam nesiryžta, nieko nesiekia. Pamylėk lvę, ir drąsia ranka nubrauksi ši prie-You can print out the paragraph where the nugali ne ištyže žodžini bot arbai. search term has appeared. Using your lordas Churchillis. parti-Copy With Formatting mouse, highlight the text that you desire to gasdinimus, sako: alėtu Edit Text Export Selection As... tu negyvi". Vienais žmucopy, and (in PCs) right click on the mouse Highlight Text tai, reikia žinoti, ka aluti-Strikethrough Text to get the Copy-Past menu and Choose me, ir veikime maži: Add Note to Replace Text ovus Add Note to Text odingas darbas" (Pr ıskis. Copy. Add Bookmark metodikos pradai, uviai Create Link Otherwise, CTRL-C will often work.

4. "Ten tai gerai — maza darbo". Nereikia užmiršti, kad darbo, vargo, aukos visur yra! "Kiekvienas luomas, pašaukimas yra altorius, ant kurio žmogus turi aukotis kitų gerovei ir Dievo garbei" (Sailer). Ten gera, kur mūsų nėra! Kiekvienas darbas, kurį žmogus dirba su meile, atsidėjimu ir nuoširdumu, reikalauja jėgos, kantrybės, išsižadėjimo.

How is search implemented – 14?

Programmer's Notepad - Untitled *]



<u>File Edit Search View Tools Window</u>

Untitled *

"Bepigu jiems – man trūksta gabumų". apgaudinėja tie, kurie niekam nesiryžta darbą, kaip lakūnas myli erdvę, ir drąs: Gyvenimą kuria ir nugali ne ištyžę žodž: Britų admiraliteto pirmasis lordas Churc vokiečių grasinimus ir gasdinimus, sako užmušti, britai jau seniai būtų negyvi" studijų kliūčių. "Apskritai, reikia žino sąskaitoje sveria gyvenime, ir veikime patvarus, sąmoningas ir metodingas darba Bendrosios mokslinio darbo metodikos pra gabūs, tik reikia išmokti dirbti!

H Next, open up a plain text editor. (We like Programmer's Notepad, downloadable for free from pnotepad.org).

Click on File – Encoding and choose UTF-8.

Then File - New - Plain Text.

Then Paste in the text from the PDF file.

You can use CTRL-V to paste, or right click to get the COPY-PASTE menu.

In this particular case, OCR was very good.

Sometimes the recognized text will spill over from one column to the next. In such cases, you need to copy and paste line by line for those lines where spillover to another column has occurred.

What's next on our agenda – 1?

- Naujienos (The News, 30,000 pages, 1959–1970)
- Laisvė (Freedom, 20,000 pages, 1918–1928)
- Argentinos Lietuvių Balsas (Argentinian Lithuanian Voice)
- Lietuvis Brazilijoje (Lithuanian in Brazil)
- Garsas (Voice)
- Amerikos Lietuvis (American Lithuanian)
- Vilnis (The Surge)
- Sandara
- Žvaigždė

What's next on our agenda – 2?

Displaced Persons camp refugee newsletters (1946 – 1952)

 Aušros belaukiant 	<u> </u>	Aušros i	belaui	kiant
---------------------------------------	----------	----------	--------	-------

BDPS (Bendrojo demokratinio pasipriešinimo sąjūdžio) užsienio delegatūros biuletenis

- DP Aitvaras
- DP Journal
- DP Jubiliejinė kultuvė
- Dragsbaek lietuvių stovyklos informacinis biuletenis
- Dypukas
- Gintaras
- Kelyje tėvynėn

Lietuvis

Lietuvių žinios – Copenhagen

Lietuvių žinios – Hanau

Mes stovyklaujam

Mūsų kelias

Mintis

Tremtinio balsas

Tremtinio dienos

Tremtis

Tuebinger DP Wochenblatt

Uchtės lietuvių stovyklos VYTIS biuletenis

LSDP organizacinio komiteto Didžiojoje Britanijoje biuleten

Žiburiai

Summary

- The digital archive of newspapers published by Lithuanian emigrants helps us to understand their hardships, their mindsets, and their worldview during the conditions extant at that time.
- Digitized texts with search facilitate access to these important sources, not only for Lithuanians, but for scholars across the world.