

Wg opisu

Wg identyfikatora

Zaawansowane

Opis publikacji

Szukaj

Google
Książki beta

CYFROWA BIBLIOTEKA NARODOWA

Strona główna Kolekcje tematyczne Podział formalny CBN BN Kontakt



Advanced Search

LIBRARY OF CONGRESS

ASK A LIBRARIAN

DIGITAL COLLECTIONS

LIBRARY CATALOGS

Our [Online Book Catalog](#) offers additional ways to search

The Library of Congress > Digital Collections

DIGITAL COLLECTIONS

Digital Collections & Services

Access to print, pictorial and audio-visual collections and other digital

Featured Digital Collections & Services

American History & Culture
A digital library of historic maps, photos, documents, audio and video from American Memory.

Historic Newspapers
Enhanced access to America's historic newspapers through the *Chronicling America* project.

Digital Collections
Open shared materials and digital preservation from...

Performing Arts
Collections, articles, presentations on theater and dance.

Prints and Photographs
Catalog of about 1 million of the Library's pictorial holdings over 1 million digital...

Veterans History
Experience the history of...

Arkadiusz Pulikowski
Instytut Bibliotekoznawstwa i Informatyki
Uniwersytet Śląski

Wyszukiwanie pełnotekstowe w zasobach bibliotek cyfrowych

Plan wystąpienia

- biblioteki cyfrowe w Polsce – stan obecny,
- wyszukiwanie pełnotekstowe w zasobach polskich bibliotek cyfrowych,
- wskazanie problemów,
- przedstawienie propozycji ich rozwiązania bazującej na doświadczeniach zagranicznych.



Biblioteki cyfrowe w Polsce – ważne daty krótkiej historii

- październik 2002
 - Wielkopolska Biblioteka Cyfrowa na platformie dLibra,
- grudzień 2002
 - Polska Biblioteka Internetowa,
- październik 2006
 - Cyfrowa Biblioteka Narodowa Polona,
- czerwiec 2007
 - Federacja Bibliotek Cyfrowych,
- czerwiec 2007
 - CBN Polona w FBC,
- lipiec 2008
 - Biblioteka Narodowa przejmuje PBI,
- czerwiec 2009
 - PBI w FBC.



Federacja Bibliotek Cyfrowych (FBC)

- została utworzona w czerwcu 2007,
- „ma na celu wirtualne połączenie bibliotek cyfrowych i repozytoriów dostępnych w polskim Internecie oraz udostępnienie nowych zaawansowanych funkcji i usług sieciowych realizowanych w tym środowisku”,
- liczy 40 uczestników, z których 36 wykorzystuje oprogramowanie dLibra,
- 19 regionalnych bibliotek cyfrowych, 21 instytucjonalnych,
- na 19 regionalnych bibliotek cyfrowych składa się sto kilkadziesiąt bibliotek i innych instytucji różnej wielkości,
- udział w FBC jest bezpłatny.



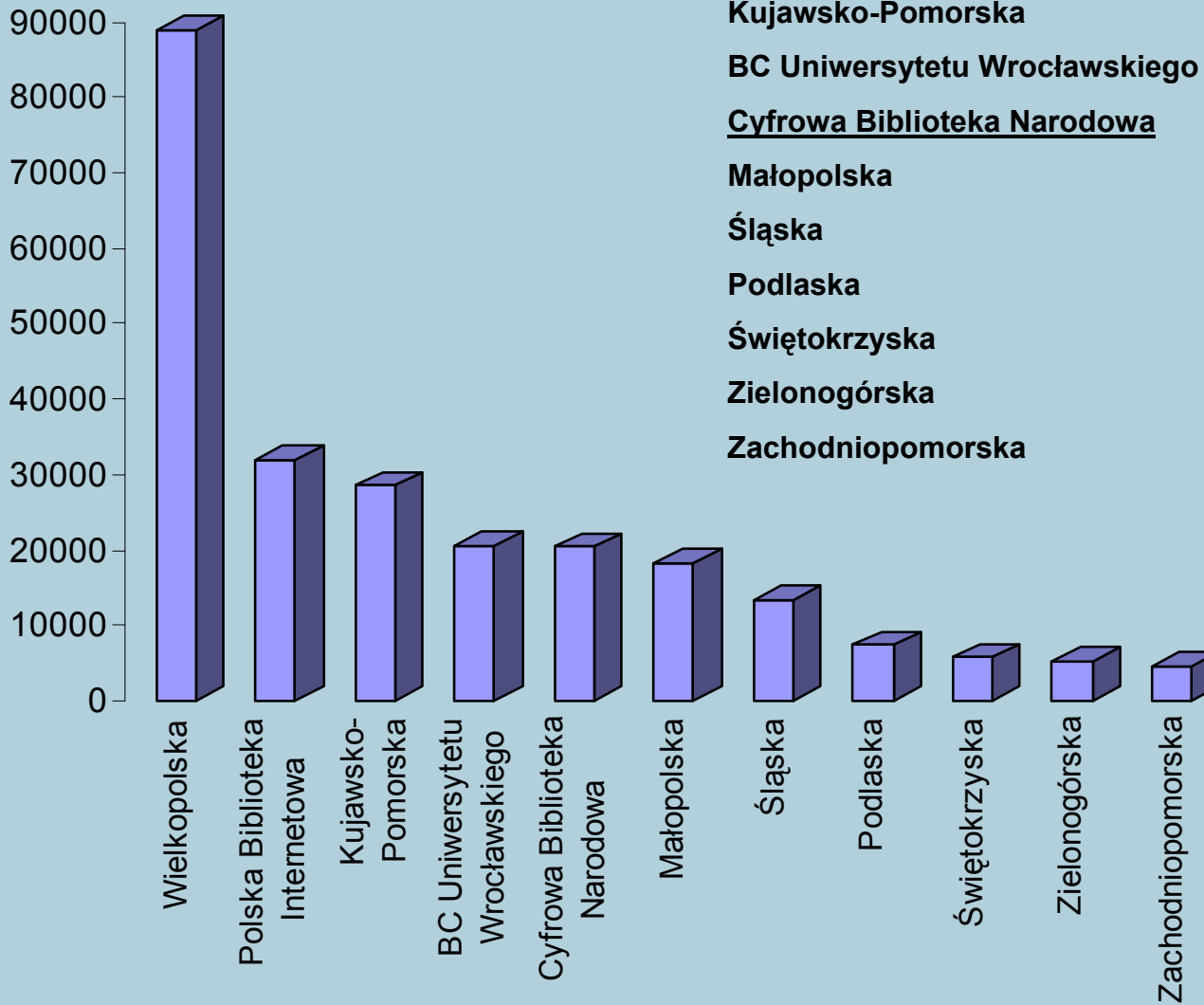
Serwis FBC

- jest rozwijany i utrzymywany przez Poznańskie Centrum Superkomputerowo Sieciowe,
- serwis FBC umożliwia przeszukiwanie z jednego miejsca publikacji zgromadzonych w bibliotekach cyfrowych należących do federacji,
- co ważniejsze, wyszukiwarka FBC pozwala kontrolować duplikaty i minimalizować ich występowanie dzięki umożliwieniu przeszukiwania wykazów publikacji przeznaczonych do digitalizacji,
- dane o zdigitalizowanych i planowanych do digitalizacji obiektach są pobierane dzięki protokołowi OAI-PMH, który jest wymagany do uwidocznienia zasobów w serwisie FBC.



Liczba publikacji zgromadzonych w największych bibliotekach cyfrowych

Łącznie w FBC: **284 504**
(20.09.2009 r.)



Wielkopolska	88873	31,2 %
<u>Polska Biblioteka Internetowa</u>	32071	11,3 %
Kujawsko-Pomorska	28608	10,1 %
BC Uniwersytetu Wrocławskiego	20497	7,2 %
<u>Cyfrowa Biblioteka Narodowa</u>	20387	7,2 %
Małopolska	18337	6,4 %
Śląska	13378	4,7 %
Podlaska	7333	2,6 %
Świętokrzyska	5751	2,0 %
Zielonogórska	5303	1,9 %
Zachodniopomorska	4579	1,6 %

Search within fields (help)
 Use a word or phrase, e.g.,
 Entire Record
 Entire Record

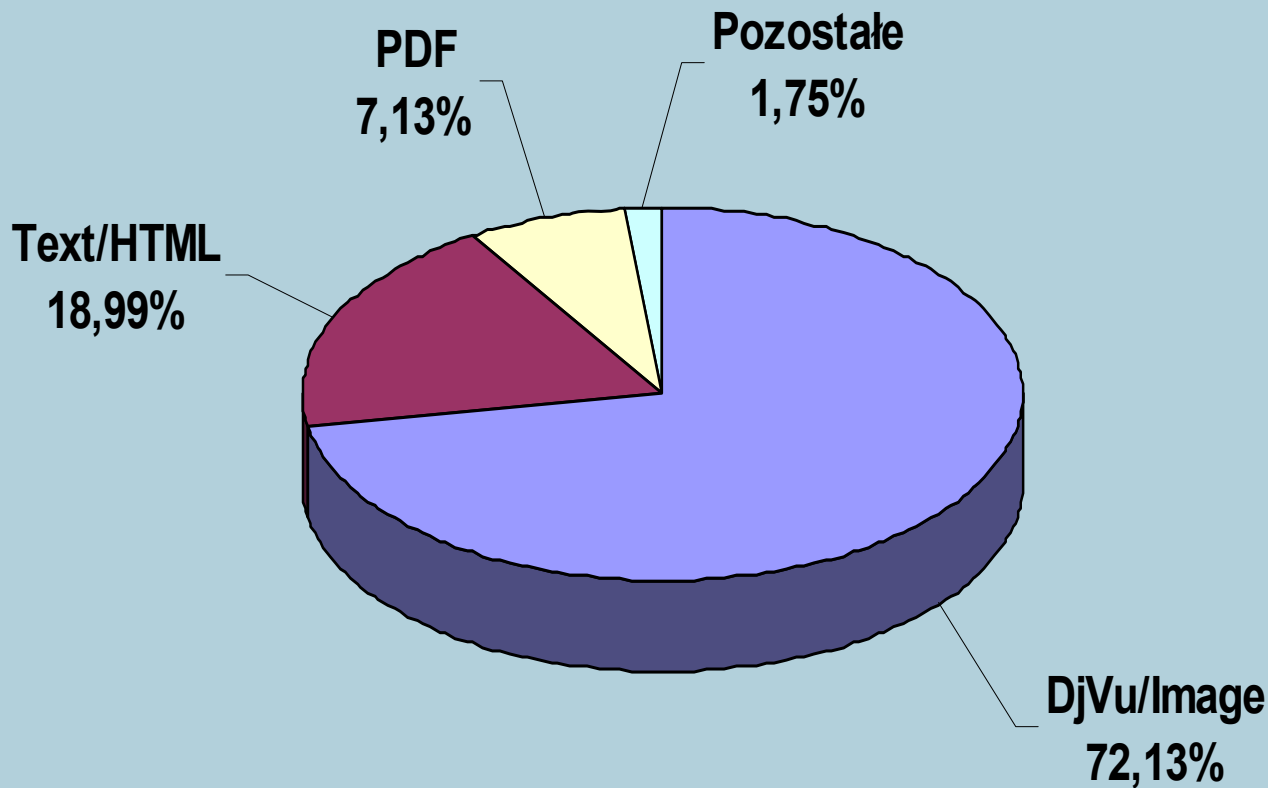
NARODOWA Biblioteka Cyfrowa
 alny CBN BN Kontakt

obecnie czytających: 182
 Śląska Biblioteka Cyfrowa

Wg opisu
 Wg identyfikatora

FEDERACJA BIBLIOTEK

Formaty plików stosowane w FBC



- 72,13% przypadające na DjVu to ponad 205 tysięcy dokumentów z 284 tysięcy dostępnych,
- to blisko $\frac{3}{4}$ zasobów, z którymi występują problemy przy wyszukiwaniu pełnotekstowym.

Format DjVu

- odpowiedni dla skanowanych dokumentów z uwagi na rozdzielenie warstwy tła od warstwy treści,
- warstwa treści jest wyodrębniana ze skanów z wykorzystaniem wbudowanego modułu OCR,
- dzięki temu możliwe jest wyszukiwanie pełnotekstowe,
- największe dwie zalety:
 - niewielkie rozmiary,
 - szybkość dostępu do poszczególnych stron dokumentów opublikowanych online,
- największe dwie wady:
 - słaby OCR,
 - nieindeksowanie przez Google i inne wyszukiwarki.



OCR DjVu a OCR FineReader

- porównanie na fragmencie z książki „Przebieczany. Wieś w powiecie wielickim” Stanisława Cercha z 1899 roku,
- poniżej wycinek oryginału, a pod nim efekty pracy OCR Document Express Prof. i FineReadera wklejone do MS Word.

Prawie do każdej chaty przytyka sad, dzielący ją od chaty sąsiadów. Wszystkie chaty zwrócone są jednak ku stronie południowej, bez względu na to, czy mają bliskie, czy dalekie sąsiedztwo.

OCR odczytany DjVu Viewera

Prawie do każdej chaty przytyka sad, dzielący ją od chaty sąsiadów.

Słód, v. 'V

Wszystkie chaty zwrócone są jednak ku stronie południowej, bez względu na to, czy mają bliskie, czy dalekie sąsiedztwo.

OCR z FineReadera 9

Prawie do każdej chaty przytyka sad, dzielący ją od chaty sąsiadów- Wszystkie chaty zwrócone są jednak ku stronie południowej, bez względu na to, czy mają bliskie, czy dalekie sąsiedztwo.



Wyszukiwanie rozproszone w serwisie FBC

- wyszukiwanie prowadzone jest wśród metadanych pobieranych do FBC od wszystkich uczestników,
- istnieje możliwość zawężenia wyszukiwania do wybranych bibliotek cyfrowych,
- dostępnych jest 15 pól wyszukiwawczych w obrębie opisu publikacji,
- nie jest możliwe wyszukiwanie w pełnym tekście.

The screenshot shows the search interface of the FBC (Federacja Bibliotek Cyfrowych) website. At the top, there are three tabs: "Wg opisu" (selected), "Wg identyfikatora", and "Zaawansowane". Below the tabs is a search bar with a dropdown menu currently open, showing the following options: "Opis publikacji" (selected), "Tytuł", "Twórca", "Temat i słowa kluczowe", "Opis", "Wydawca", "Współtwórca", "Data", "Typ zasobu", "Format", "Identyfikator zasobu", "Źródło", "Język", "Powiązania", "Zakres", and "Prawa". To the right of the search bar is a "Szukaj" button. Below the search bar, there are checkboxes for "publikacje dostępne" (checked) and "publikacje planowane". A link "Wybierz biblioteki cyfrowe" is also visible. The main content area shows the number of publications: "Wszystkich publikacji: 284 504" and "Planowanych publikacji: 4 175". At the bottom, there is a section titled "Baza" with a description: "Baza Bibliotek Cyfrowych" and "Baza Bibliotek Cyfrowych (w skrócie) to baza danych i udostępnianie danych na temat bibliotek cyfrowych w Polsce, także będących w trakcie powstawania."

Wyszukiwanie lokalne na stronie jednej z bibliotek cyfrowych na platformie dLibra

- wyszukiwanie prowadzone jest wśród tych samych elementów opisu co w FBC, dodatkowo pojawiają się tagi dodawane przez użytkowników danej biblioteki,
- istnieje możliwość wyszukiwania pełnotekstowego,
- domyślnie każde zapytanie jest rozszerzane przy użyciu synonimów (możliwe wyłączenie).

The screenshot shows the search interface of the Wielkopolska Biblioteka Cyfrowa. At the top, there are two tabs: "Wielkopolska Biblioteka Cyfrowa" and "Federacja Bibliotek Cyfrowych". Below them is a search input field and a "Szukaj" button. The search criteria are set to "tylko w tekście publikacji". A dropdown menu is open, showing various search criteria: "tylko w opisie bibliograficznym" (selected), "Tytuł", "Twórca", "Temat i słowa kluczowe", "Opis", "Wydawca", "Współtwórca", "Data", "Typ zasobu", "Format", "Identyfikator zasobu", "Źródło", "Język", "Powiązania", "Zakres", "Prawa", and "Tagi". To the right of the dropdown, there is a link "Poprawne formułowanie zapytań". Below the search area, there are sections for "Kolekcje okolicznościowe" and "Wiadomości". The "Kolekcje okolicznościowe" section lists several collections, including "90 lat Politechniki Poznańskiej", "90 lat Uniwersytetu im. Adama Mickiewicza w P...", "Powszechna Wystawa Krajowa - 80 lat temu", "Polonia Maior-fontes", and "Powstanie Wielkopolskie 1918/19". The "Wiadomości" section shows "90 lat Politechniki Poznańskiej" with the date "[09.09.2009]".

Prezentacja wyników lokalnego wyszukiwania pełnotekstowego

- lista trafień odsyła do dokumentów zawierających poszukiwane terminy,
- niestety, po wybraniu publikacji nie jesteśmy przenoszeni do strony zawierającej poszukiwany termin,
- nie wiemy ile razy termin pojawił się w dokumencie,
- musimy samodzielnie przeszukiwać dokument, korzystając z narzędzi dostępnych w czytniku danego formatu, np. Djvu Viewer (czasochłonne, szczególnie przy dużej liczbie stron).

Wyniki wyszukiwania

Wyniki wyszukiwania ogólnego w kolekcji - Wielkopolska Biblioteka Cyfrowa

Zapytanie: 

Lista publikacji spełniających warunek: **Tekst publikacji: stanisławów** 

Wyniki: Znaleziono obiektów: 113

Widok: **Prosty** | Rozbudowany

1.  **Kościuszek w Sosnowicy - Stodor Adam** [■■■■■]
2.  **Urzędowy spis: lekarzy, lekarzy-dentystów, farmaceutów, felczerów, pielęgniarek, położnych, uprawnionych i samodzielnych techników dentystycznych oraz wykazy: aptek, szpitali, ubezpieczalni społecznych, ośrodków zdrowia, przychodni samodzielnych oraz centrali i filii Państwowej Szkoły Higieny Cz. VIII Apteki** [■■■■■]
3.  **Urzędowy spis lekarzy uprawnionych do wykonywania praktyki lekarskiej oraz aptek w Rzeczypospolitej Polskiej** [■■■■■]
4.  **Księga pamiątkowa ku czci Władysława Abrahama T.2** [■■■■■]

Wyszukiwanie lokalne na stronie Cyforwej Biblioteki Narodowej

- CBN Polona wykorzystuje format PDF zamiast DjVu,
- PDF z dobrym OCRem zapewnia dużo lepsze wyniki wyszukiwania pełnotekstowego.

CYFROWA BIBLIOTEKA NARODOWA

Strona główna Kolecje tematyczne Podział formalny CBN BN Kontakt

Szukaj w CBN Polona

Tekst publikacji
katowice

Użyj synonimów

Publikacja

Tytuł :
Dziennik Krakowski. 1939, nr 3 (11 września)

Menu

Opis publikacji
Treść
Wyniki wyszukiwania
Powrót do kolekcji

Treść w nowym oknie
Wersja tekstowa (PDF)
Wersja graficzna (JPG)

Strona tytułowa

Wojaska niemiecka, wkraczające do Krakowa, rozlepiły na ulicach miasta i w bramach domów ogłoszenie Komendanta naczelnego wojsk gen. von Brauchitscha w dwu językach. W ogłoszeniu ten dowódca wojsk stwierdza dwie zasady: 1) walczą oni z wojskiem, a nie z bezbronną ludnością (nie „z narodem”) i 2) uszanują wszystkie międzynarodowe zwyczaje i prawa wojenne.

Wojaska niemieckie, wkraczające do Krakowa, rozlepiły na ulicach miasta i w bramach domów ogłoszenie Komendanta naczelnego wojsk gen. von Brauchitscha w dwu językach. W ogłoszeniu ten dowódca wojsk stwierdza dwie zasady: 1) walczą oni z wojskiem, a nie z bezbronną ludnością (nie „z narodem”) i 2) uszanują wszystkie międzynarodowe zwyczaje i prawa wojenne.

Podsumowanie możliwości wyszukiwania pełnotekstowego w zasobach polskich bibliotek cyfrowych

- Google oraz inne wyszukiwarki nie są w stanie bezpośrednio indeksować plików djvu; z innymi formatami (1/4) radzą sobie bardzo dobrze,
- w związku z powyższym można przyjąć, że około 3/4 zasobów polskich bibliotek cyfrowych należy do tzw. niewidocznego Internetu,
- jakość OCR w plikach djvu w znacznym stopniu ogranicza wykorzystanie wyszukiwania pełnotekstowego dostępnego lokalnie w większości bibliotek na platformie dLibra,
- to poważny problem, gdyż nawet gdyby Google zaczął indeksować pliki djvu, kompletność wyników byłaby daleka od oczekiwanej,
- sposób prezentacji wyników wyszukiwań w znacznym stopniu utrudnia dotarcie do fragmentu tekstu zawierającego poszukiwaną frazę.



Próby rozwiązania aktualnie występujących problemów

- opisane problemy są znane specjalistom zajmującym się na co dzień bibliotekami cyfrowymi,
- pisze się o tych problemach na forach Biblioteka 2.0 oraz Digitalizacja i biblioteki cyfrowe (Ebib),
- sposobem na uwidocznienie zasobów w Google jest ekstrakcja warstwy tekstowej z dokumentów djvu do osobnego (ukrytego) pliku i podpięcie do opisu,
- dLibra 4.0 ma to już wbudowane, ale skuteczność rozwiązania jest ograniczona,
- słaby OCR można zamienić na bardzo dobry wykonując skany w FineReaderze jako PDF dwuwarstwowy, a następnie konwertując programem pdf2djvu (czasochłonne).



Zagraniczne sposoby udostępniania skanowanych dokumentów – źródło inspiracji

- Google Books <http://books.google.pl>
- Internet Text Archive <http://www.archive.org>
- Gallica <http://gallica.bnf.fr>
- Digitale bibliotheek voor Nederlandse letteren <http://www.dbnl.org>
- Austrian Literature Online <http://www.literature.at>
- Hungarian Electronic Library <http://mek.oszk.hu>
- National Library of Australia <http://www.nla.gov.au/digicoll>
- Biblioteca nacional digital <http://purl.pt>
- The Digital Christian Library <http://thedcl.org>
- aDigital Library of India <http://dli.iiit.ac.in>
- Botanicus <http://www.botanicus.org>



Google Books – books.google.pl

- nawet dla książek dostępnych w pełnym widoku nie ma możliwości zaznaczania,
- wygoda...

The screenshot shows a Google Books search for 'katowice'. The search results page displays the book 'Political trials in Poland, 1981-1986' by Andrzej Swidlicki. The search results show 'Wynik 4 z 21 w tej książce dla zapytania katowice'. The search results are: '< Poprzednia', 'Następna >', and 'Wyświetl wszystko'. The search results are displayed in a yellow box. The search results are: 'to transfer the factory Committee documents to the management. The workers then demanded that the union banner be protected from the ZOMO. He agreed to this and took the banner out of its show-case. In a spontaneous movement, 400 people gathered around the banner and it was carried to the Katowice Steelworks. As they approached the steelworks, the workers sang the hymn 'God Save Poland' and were greeted by representatives of the Strike Committee. It was decided that Kowalski and four other representatives of the Steelworks Repair Enterprise would join the Strike Committee while the rest would disperse and go home. Whoever wished to return to work on the following day would be free to do so. Kowalski, who was released on 24 January 1983 on the strength of the medical experts' report pending the court's decision on whether to continue the trial, disappeared in unexplained circumstances on 7 February 1983. His body was found in a river on 31 March 1983.'

Internet Text Archive – www.archive.org/details/texts

- wśród sponsorów Microsoft i Yahoo,
- ambicją przegonienie Google Books,
- najważniejsze kolekcje:
 - American Libraries, ponad milion pozycji,
 - Canadian Libraries, ponad 200 tys. poz.,
 - Universal Library, ponad 30 tys. poz.,
 - Project Gutenberg, ponad 20 tys. poz.

View the book



Read Online	(~132 pg)
PDF	(6.11 MB)
BMW PDF	(5.89 MB)
Full Text	(124 KB)
DjVu	(3.48 MB)

All Files: [HTTP](#)

[Help reading texts](#)

Problemy społeczno-wychowawcze w środowisku młodzieży (1915)

Author: [Bobrowska-Nowak, Wanda](#); [Radziejewicz-Winnicki, Andrzej](#)
Subject: [Educational sociology](#)
Publisher: [Katowice : Uniwersytet Slaski](#)
Year: [1915](#)
Language: [Polish](#)
Call number: 1011374
Digitizing sponsor: [Sloan Foundation](#)
Book contributor: [The Library of Congress](#)
Collection: [americana](#)

 This book has an [editable web page](#) on [Open Library](#).

Description

Summaries in English and Russian



Internet Text Archive (2)

- wyszukiwanie pełnotekstowe można realizować tylko w trakcie przeglądania dokumentów,
- dopracowane jest tak, jak to z Google Books,
- obecność PDF, HTML, TXT pozwala Google na indeksowanie.

Zoom: Auto Problemny społeczno-wychowawcze w środowisku modz... Page: n17

socialized go

Page n10 - of the Socialized Recitation. ...
Page n14 - "otation Work Socialized *honetie Spelling Game
Page n16 - . The Socialized Recitation emphasizes the pupil and uses the
Page n17 - . The Socialized Recitation does not neglect any of the
Page 1 - of the Socialized Recitation. — To
Page 2 - snbjects are socialized will guard against any disorganization of the
Page 4 - THE SOCIALIZED RECITATION. criticism should be shown.
Page 6 - THE SOCIALIZED RECITATION standpoint, it g apos ives
Page 8 - be termed socialized experience that is, not merely
Page 10 - THE SOCIALIZED RECITATION. morals. If money interests
Page 12 - THE SOCIALIZED RECITATION. others. He should receive
Page 14 - THE SOCIALIZED RECITATION. dimly conscious. The feeling
Page 16 - THE SOCIALIZED

x INTRODUCTION.

result, the old method does not object, but does not concern itself.

The Socialized Recitation Makes the Schoolroom Real, Life-Like and Natural.—The subject-matter of a given lesson is so planned by the teacher that it becomes material to be used by the child in creating experiences, and in giving and receiving impressions. Thus to a very large degree drill is eliminated, but the facts are nevertheless fixed, because of the concrete situations in which the child uses them. The Socialized Recitation avoids the artificial conditions of the old classroom and recitation. The teacher is no longer the standard and arbiter of all things. The children become members of a working community which adopts the principles of character and of good citizenship as the standard of living and working. The teacher becomes a better planner and guide, but is less active in the classroom during recitations. The teacher's work must be done before school opens, and once in the classroom, she becomes only a member of the class with more or less authority as required.

The Socialized Recitation does not neglect any of the formal or fundamental principles of good teaching. It employs them to far better advantage and does away with the *academized recitation* conducted by the teacher to the exclusion of the pupil and his participation in the work.

WILLIAM T. WHITNEY.

Port Chester.

EXPLANATORY CHART
OF THE
SOCIALIZED RECITATION

Purpose of the Socialized Recitation.—To do away with passivity in the classroom; to provide opportunity for the natural development of initiative, of activity resulting in *originality*, of the imaginative powers, and of the realization and assumption of responsibility; to give opportunity for the child to *do* and *to be*, rather than merely to know.

Change Necessitated.—The teacher's work becomes the background rather than the working center of the class period. *She plans* ways of developing the pupils' initiative and vitalizing their experience. She herself must be the *instigator*, but not the entire power, in arranging and encouraging discussions and continuing effort. Just as the successful play organizer succeeds best when she reaches the stage where the children naturally lead, so the teacher of the day school must withdraw from dominating the game, and must arrange methods of work so that pupils become personalities showing reaction, not mere receptacles.

1

Propozycja

- poszerzenie dostępnych formatów pliku dla tego samego dokumentu,
- oprócz dotychczasowego DjVu – PDF i TXT,
- PDF dwuwarstwowy, wygenerowany przez FineReadera z pliku DjVu,
- taki PDF będzie charakteryzował się świetnym OCR
- Google zaindeksuje zarówno PDFy jak i TXT,
- odzyskamy zasoby dla widocznego Internetu,
- wszystkie dotychczasowe problemy znikną,
- do tego zyskamy Google Custom podpięte do FBC,
- przykład dla dokumentu 32 stronicowego:
 - DjVu 2,5 MB,
 - PDF:
 - jakość wysoka 2 MB (B&W bez tła),
 - jakość średnia 1,4 MB (B&W bez tła),
 - jakość niska 4 MB (Grayscale z tłem),



Wg opisu

Wg identyfikatora

Zaawansowane

Opis publikacji

Szukaj

Biblioteka

- Śląska Biblioteka Cyfrowa
 - .Dolny Śląsk
 - .Górny Śląsk
 - .Kresy Wschodnie

Śląska Biblioteka Cyfrowa

Federacja Bibliotek Cyfrowych

Szukaj

Google
Książki beta

CYFROWA BIBLIOTEKA NARODOWA

Strona główna Kolekcje tematyczne Podział formalny CBN BN Kontakt



BN POŁO

Project
Gutenberg

Online Book Catalog

Author:

Title Word(s):

EText-No:

Go!

Advanced Search

Recent Books

Top 100

Online Catalog

Library Catalog

Advanced Search

LIBRARY OF CONGRESS

ASK A LIBRARIAN

DIGITAL COLLECTIONS

LIBRARY CATALOGS

Our [Online Book Catalog](#) offers additional ways to search

Author:

Title:

Subject:

Language:

Category:

LoCC:

Filetype:

Etext-No:

Full Text

The Library of Congress > Digital Collections

DIGITAL COLLECTIONS



- Digital Collections Home
- About Digital Collections
- Digital Reference
- Digital Library Standards
- Digital Preservation

More Resources

Bibliographies and Guides

Digital Collections & Services

Access to print, pictorial and audio-visual collections and other digital resources

Featured Digital Collections & Services



American History & Culture
A digital library of historic maps, photos, documents, audio and video from American Memory.



Performing Arts Collections
Collections, articles, presentations on theater and dance.



Historic Newspapers
Enhanced access to America's historic newspapers through the *Chronicling America* project.



Prints and Photographs
Catalog of about 1.5 million of the Library's pictorial holdings over 1 million digital images.



International Collections
International materials and bilingual pressings from foreign countries.



Veterans History
Experiences from the American Civil War to the present.

Dziękuję za uwagę

e-mail: arkadiusz.pulikowski@us.edu.pl