NICOLAUS COPERNICUS
UNIVERSITY
IN TORUŃ
Faculty of History

Studying the past Web in Poland – current state and perspectives

Bartłomiej Konopa
NCU Toruń, State Archive in Bydgoszcz

06.06.2019

NICOLAUS COPERNICUS UNIVERSITY IN TORUŃ
Faculty of History

Table of contents

NICOLAUS COPERNICUS UNIVERSITY IN TORUŃ
Faculty of History

Program for digitizing cultural heritage and collecting, storing and accessing digital objects in Poland 2009-2020

Archiving polish Web as one of four task of expanding Polish digital resources

NICOLAUS COPERNICUS UNIVERSITY IN TORUŃ
Faculty of History

Institutions that could be responsible for archiving the Polish Web

National Library of Poland

National Digital Archives

NICOLAUS COPERNICUS UNIVERSITY IN TORUŃ
Faculty of History

Internet Archive by National Digital Archives

- launched in april 2009
- opened in march 2010
- last crawl in april 2011
- closed in end of 2015
- scope: 41 sites (governmental and State Archives websites)

NICOLAUS COPERNICUS UNIVERSITY IN TORUŃ
Faculty of History

archiwuminternetu beta

Wprowadź adres www: http:// 2009 Szukaj

Wyszukiwanie zaawansowane

Witamy w Archiwum Internetu!

Narodowe Archiwum Cyfrowe, aby zabezpieczyć informacje o historycznym znaczeniu dla państwa polskiego archiwizuje strony internetowe najważniejszych urzędów państwowych. W wersji beta Archiwum Internetu dostępne są strony poniższych instytucji.

- Naczelna Dyrekcja Archiwów Państwowych
- Archiwum Główne Akt Dawnych
- Archiwum Akt Nowych
- Narodowe Archiwum Cyfrowe
- Archiwum Państwowe w Białymstoku
- Archiwum Państwowe w Bydgoszczy
- Archiwum Państwowe w Częstochowie
- Archiwum Państwowe w Elblągu z siedzibą w Malborku
- Archiwum Państwowe w Gdańsku
- Archiwum Państwowe w Gorzowie Wielkopolskim
- Archiwum Państwowe w Kaliszu
- Archiwum Państwowe w Katowicach
- Archiwum Państwowe w Kielcach
- Archiwum Państwowe w Koszalinie
- Archiwum Państwowe w Krakowie
- Archiwum Państwowe w Lesznie
- Archiwum Państwowe w Lublinie
- Archiwum Państwowe w Łodzi

Projekt | Ludzie | Prawa autorskie

NDA's Internet Archive website, archive.org, captured at 21.12.2015

NICOLAUS COPERNICUS UNIVERSITY IN TORUŃ
Faculty of History

National Library of Poland

- took part in '09 European Election Web Harvesting Project
- only 16 websites captured
- unpublished

NICOLAUS COPERNICUS UNIVERSITY IN TORUŃ
Faculty of History

Anonymous Internet Archive

- at the turn of 2012 and 2013
- websites from polish ccTLD found in Alexa rankings
- 14 thousand websites (only homepages)
- crawl every hour
- collected data is not available

NICOLAUS COPERNICUS UNIVERSITY IN TORUŃ
Faculty of History

Archiwum Internetu

- Dane
- O Archiwum
- Kontakt

Dane

- Obejmują wszystkie adresy z końcówką .pl znajdujące się w rankingu miliona najczęściej odwiedzanych stron w Internecie wg alexa.com. W tej chwili jest to nieco ponad 14 tys. adresów.
- Strony główne tych serwisów kopiowane są co godzinę i pakowane do pliku.
- W tej chwili dostępne są tylko surowe dane, tj. **co godzinę** do poniższej listy **dopisywana jest aktualna paczka**.

Oto one:

2012-11-17 20:00:01
2012-11-17 19:00:02
2012-11-17 17:00:01
2012-11-17 16:00:02
2012-11-17 14:00:02
2012-11-17 13:00:01
2012-11-17 12:00:01
2012-11-17 11:00:01
2012-11-17 10:00:02
2012-11-17 09:00:01
2012-11-17 08:00:02
2012-11-17 07:00:01
2012-11-17 06:00:01
2012-11-17 05:00:02
2012-11-17 03:00:02
2012-11-17 00:00:02
2012-11-16 22:00:01
2012-11-16 21:00:01
2012-11-16 20:00:02
2012-11-16 18:00:02
2012-11-16 17:00:02
2012-11-16 16:00:02
2012-11-16 15:00:01
2012-11-16 14:00:02

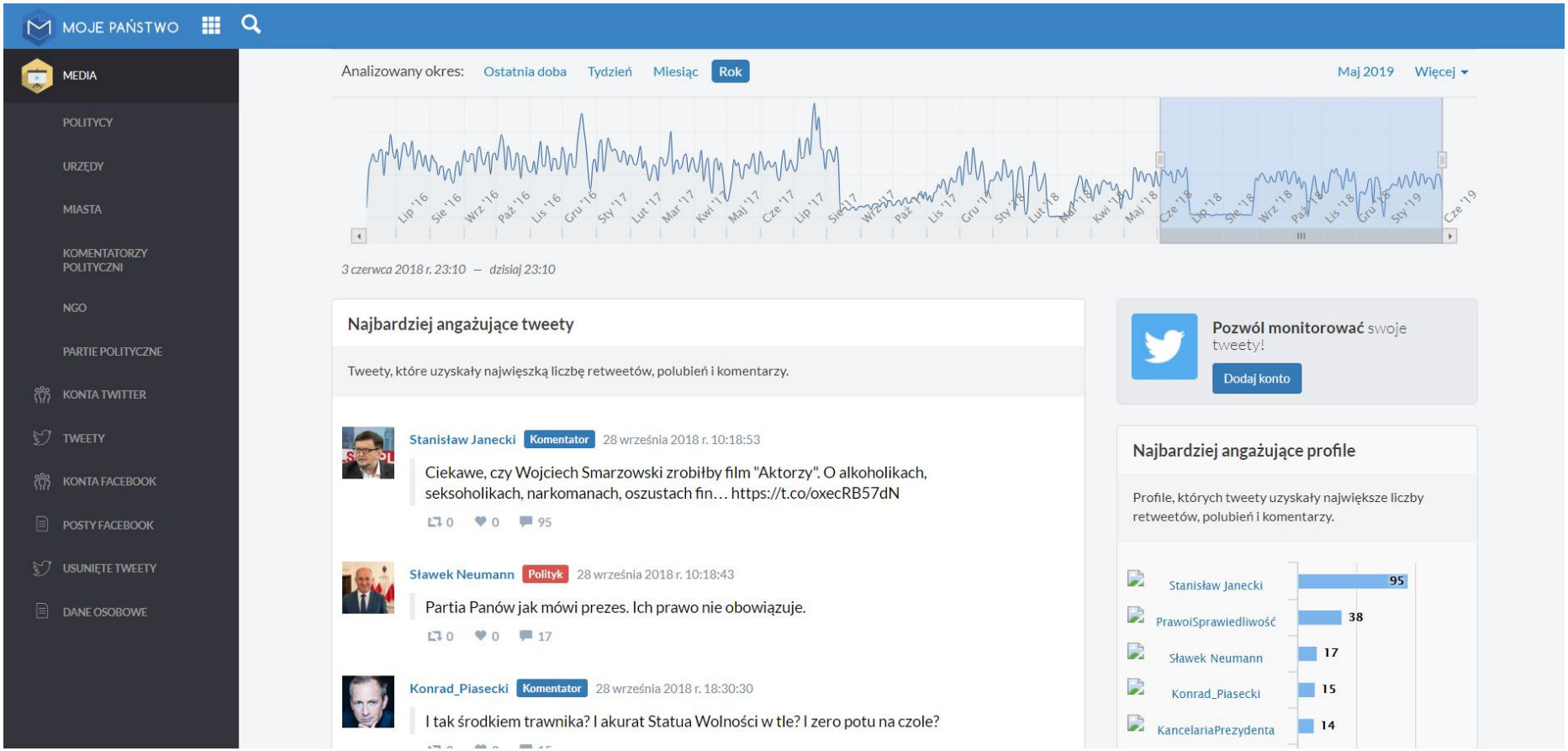
Anonymous Internet Archive website, archive.org, captured at 17.11.2012

NICOLAUS COPERNICUS UNIVERSITY IN TORUŃ
Faculty of History

My Country by ePanstwo

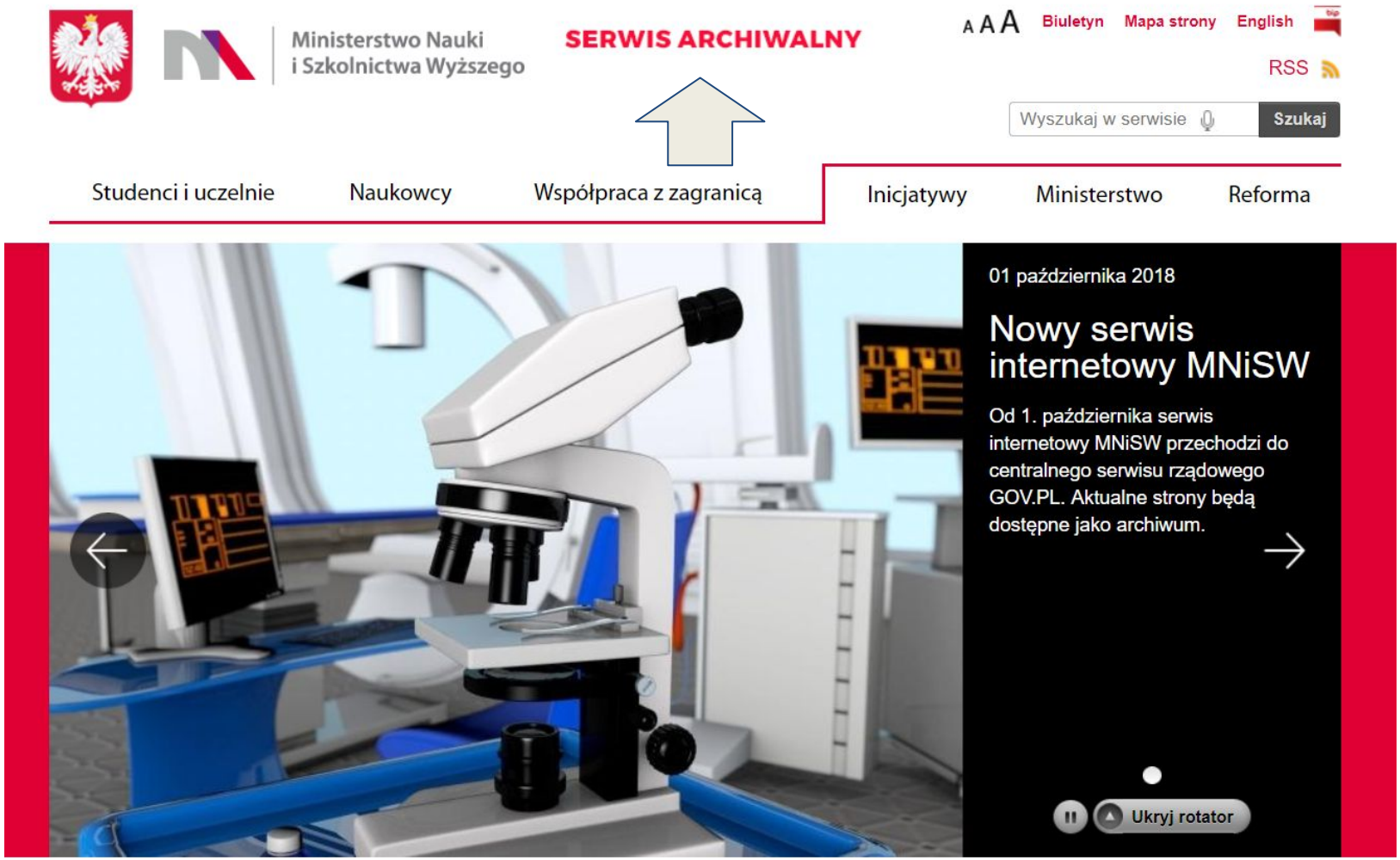
- political related content gathered from Facebook and Twitter

NICOLAUS COPERNICUS UNIVERSITY IN TORUŃ
Faculty of History



My Country web service

NICOLAUS COPERNICUS UNIVERSITY IN TORUŃ



Archival website of Ministry of Science and Higher Education

NICOLAUS COPERNICUS UNIVERSITY IN TORUŃ
Faculty of History

Polish scientific literature about Web archiving projects

- mostly about Internet Archive foundation
- about using Wayback Machine
- also introducing other national or grassroot initiatives

NICOLAUS COPERNICUS UNIVERSITY IN TORUŃ
Faculty of History

Polish scientific literature about Web archiving

- Web archiving as a process
- methods and tools
- legal issues
- grassroot archiving
- cooperation with users
- information management

NICOLAUS COPERNICUS UNIVERSITY IN TORUŃ
Faculty of History

Polish scientific literature about archived Web as sources for researches

- Web archeology as a way to study websites
- important source for contemporary history and humanistics
- going to replace old sources
- requires new skills

NICOLAUS COPERNICUS UNIVERSITY IN TORUŃ
Faculty of History

webArch
Pracownia archiwizacji Webu LaCH UW

O pracowni | Blog | About | @DHLab_UW

Szukaj ...
Szukaj

O PRACOWNI
- Zespół

ABOUT

BAZA WIEDZY
- Archiwizacja Webu
- Periodyzacja archiwizacji Webu
- Krótka historia
- Bibliografia (pol)
- Do pobrania
- Seminarium (2017)
- Tworzenie przypisów
- Jak zadbać o stabilne linki?
- Sprawdź swoja strone

webArch LaCH UW

Celem pracowni jest gromadzenie wiedzy i rozwijanie kompetencji w zakresie profesjonalnej archiwizacji Webu i zaawansowanej pracy z archiwami WWW.

Archiwizacja Webu

Archiwizacja Webu to proces gromadzenia części zasobów World Wide Web (WWW) w celu zachowania ich dla przyszłych badaczy, historyków i opinii publicznej.

Warsztaty

Zachęcamy do udziału lub współorganizacji warsztatów z zagadnień archiwistyki Webu.

webArch's blog

NICOLAUS COPERNICUS UNIVERSITY IN TORUŃ
Faculty of History

M. Roszkowski, B. Wlodarczyk

Web Citations in Polish Library and Information Science Journals: the Analysis of URLs' Validity (2016)

- 4 scientific journals analized
- 4593 unical URLs from 670 articles - approx. 30% inactive
- 53% of inactive was found in the Internet Archive resources

NICOLAUS COPERNICUS UNIVERSITY IN TORUŃ
Faculty of History

M. Król, D. Zdonek and J. Gorzelny
Information value of Internet domain (2017)

- searching for methods to evaluate reliability
- comparison of WHOIS search engine and Wayback Machine

NICOLAUS COPERNICUS UNIVERSITY IN TORUŃ
Faculty of History

B. Michalska-Bednarek

***History of websites of NCU Library* (2016)**

- based on official reports, interviews and archived Web

A. Surdyk

***The activities of academic associations on the internet based on the example of PTBG. A summary of seven years of the existence of the association on the net* (2017)**

NICOLAUS COPERNICUS UNIVERSITY IN TORUŃ
Faculty of History

M. Król

From the archives of the Internet: changes in the mode of presentation of agrotouristic offer (2017)

- analysis of websites of agritourism centers
- author observed changes in technical issues and designing style

The use of archived Web in scientific research - survey results

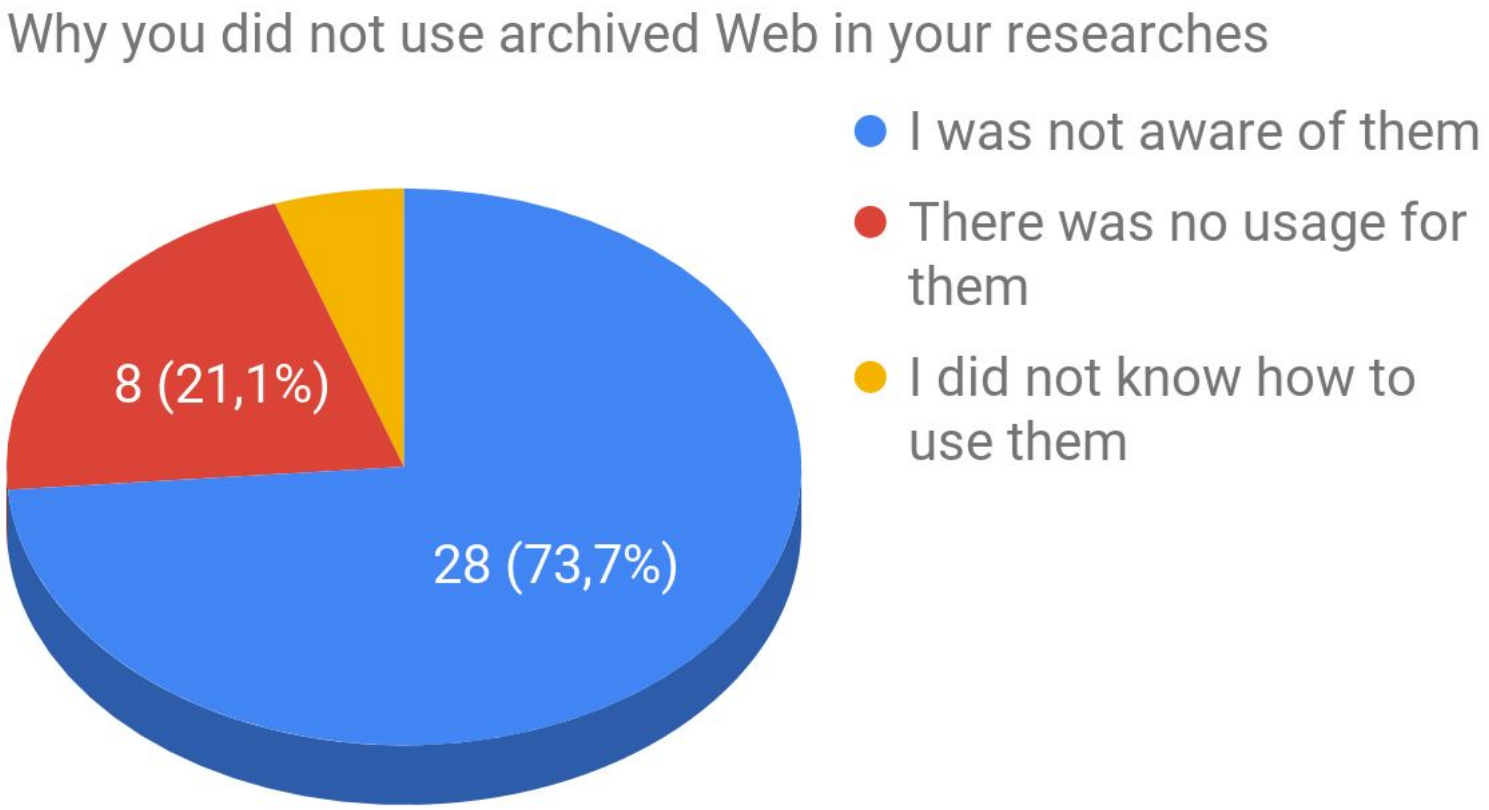
- survey was addressed to scholars and PhD students
- division into users and nonusers
- 52 responses were collected

The use of archived Web in scientific research - users

- 11 respondents were using archived Web among others from social communication sciences, IT, economics, archeology and fine arts
- mostly using Internet Archive (project Minerva and Google Cache also mentioned)
- commonly used to obtain inaccessible materials
- average rating 3.9 (1-5 scale)

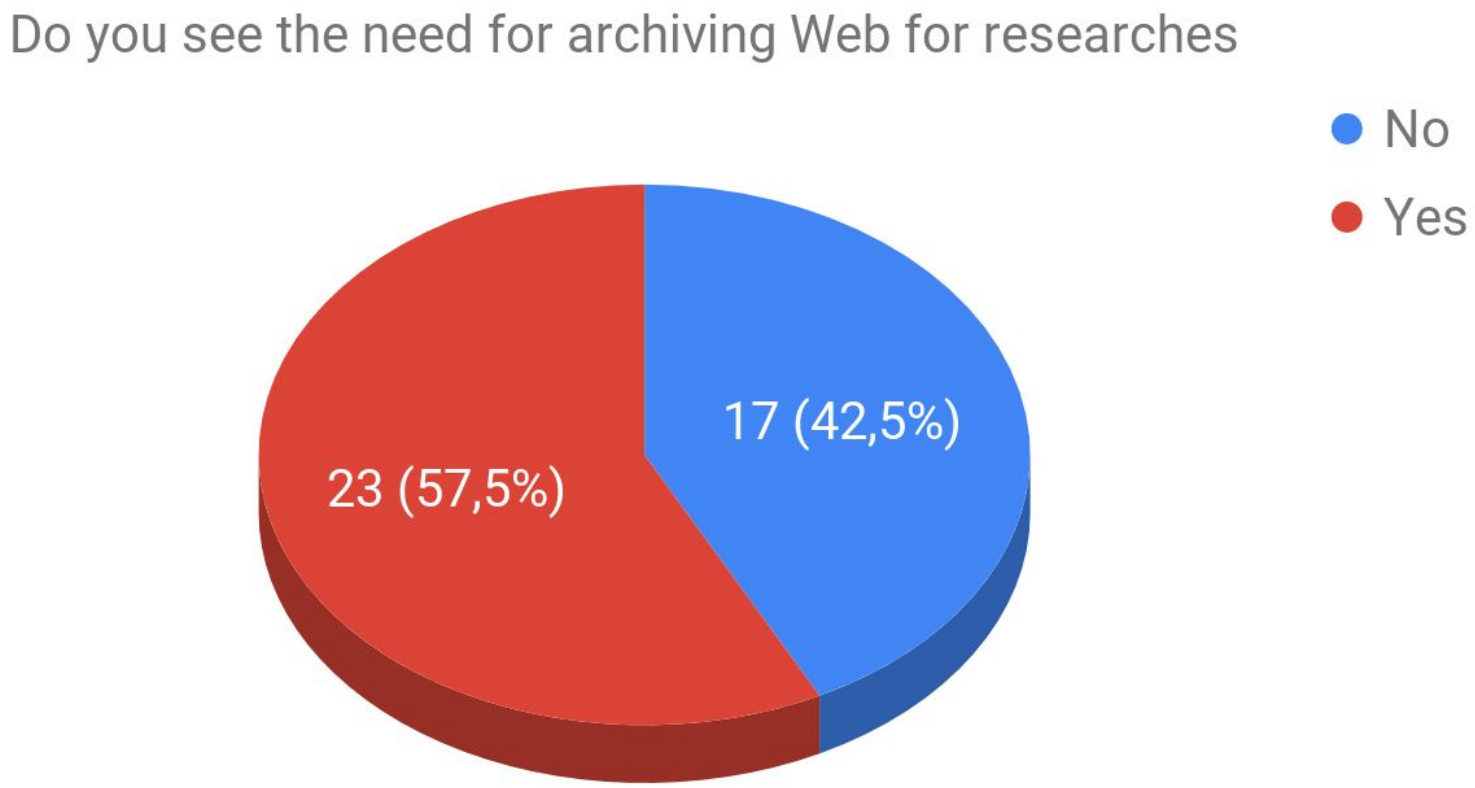
NICOLAUS COPERNICUS UNIVERSITY IN TORUŃ
Faculty of History

The use of archived Web in scientific research - nonusers



NICOLAUS COPERNICUS UNIVERSITY IN TORUŃ
Faculty of History

The use of archived Web in scientific research - nonusers



NICOLAUS COPERNICUS UNIVERSITY IN TORUŃ
Faculty of History

The use of archived Web in scientific research - user and nonusers

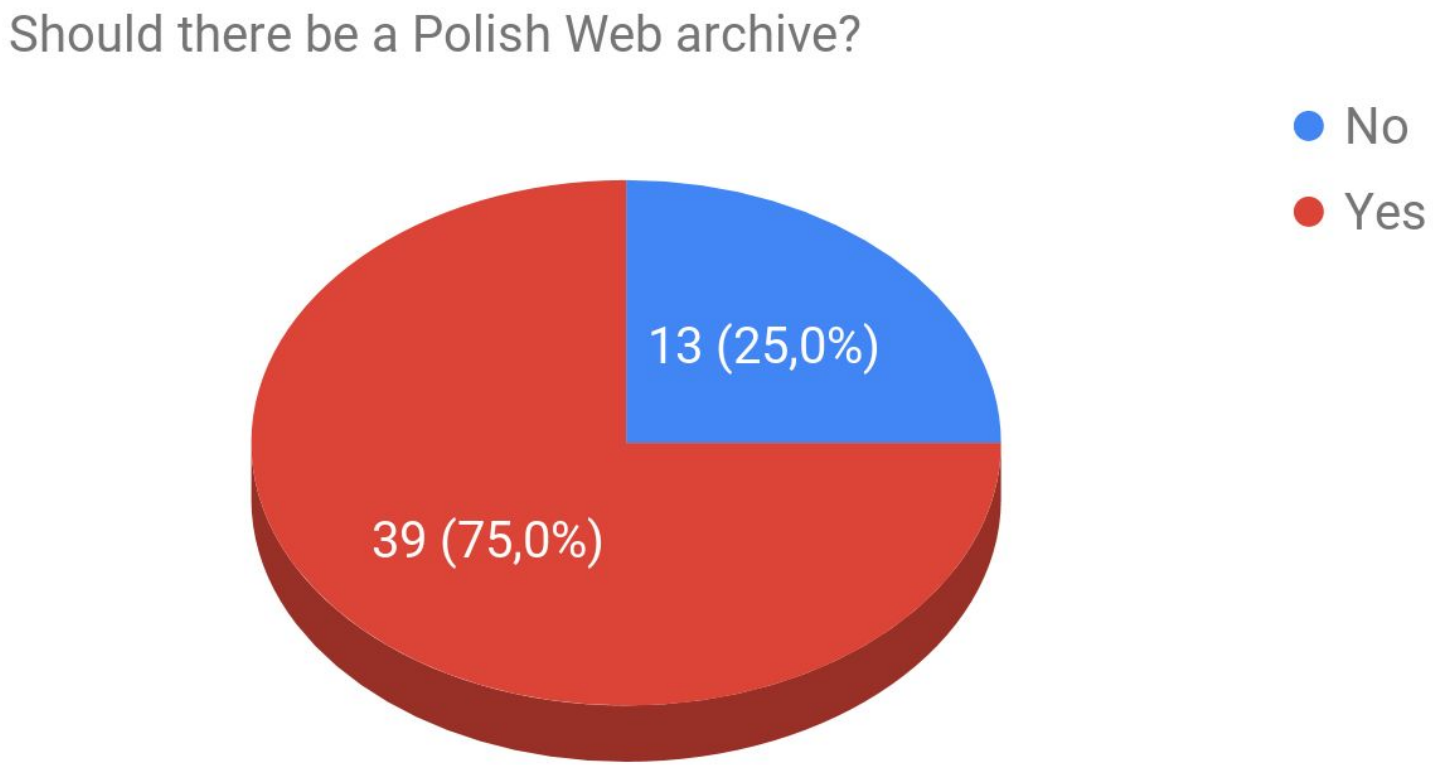
Which scientific disciplines could use the archived Web?

(most commonly responses):

- history (22)
- social communication and media sciences (13)
- sociological sciences (9)
- cultural and religious sciences (8)
- legal sciences (6)
- economics (6)
- political sciences (5)
- linguistics (5)
- computer sciences (5)

NICOLAUS COPERNICUS UNIVERSITY IN TORUŃ
Faculty of History

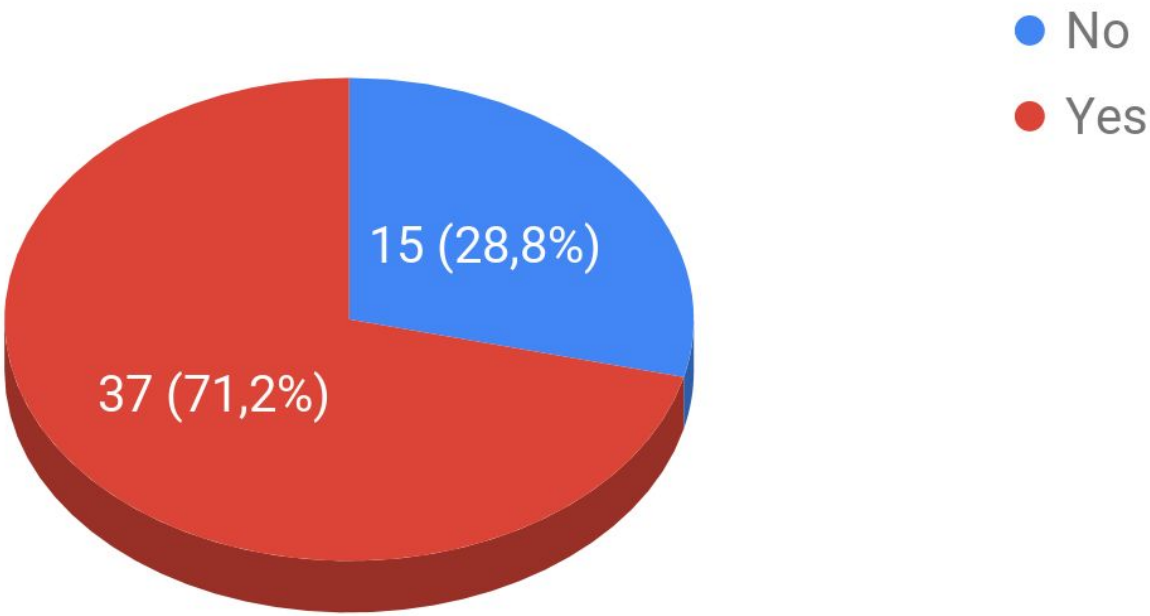
The use of archived Web in scientific research - user and nonusers



NICOLAUS COPERNICUS UNIVERSITY IN TORUŃ
Faculty of History

The use of archived Web in scientific research - user and nonusers

Does the issue of archived Web require popularization (eg in the form of trainings, workshops, etc.)?



NICOLAUS COPERNICUS UNIVERSITY IN TORUŃ
Faculty of History

Main aims for the future of studying past Web in Poland

establishing Polish Web archive

popularization and educating

NICOLAUS COPERNICUS UNIVERSITY IN TORUŃ
Faculty of History

Thank you

Bartłomiej Konopa
bartlomiejkonopa@gmail.com