

# Webarchiv

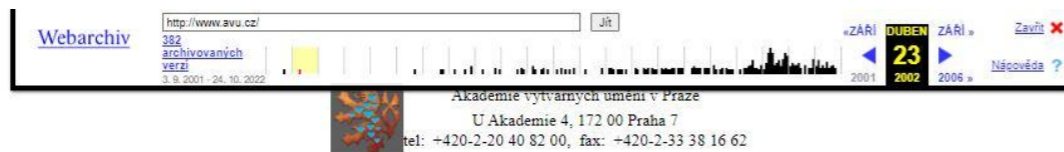
Archivace dat ve webovém archivu a možnosti jejich  
využití. Nové výstupy výzkumného projektu

Marie Haškovcová

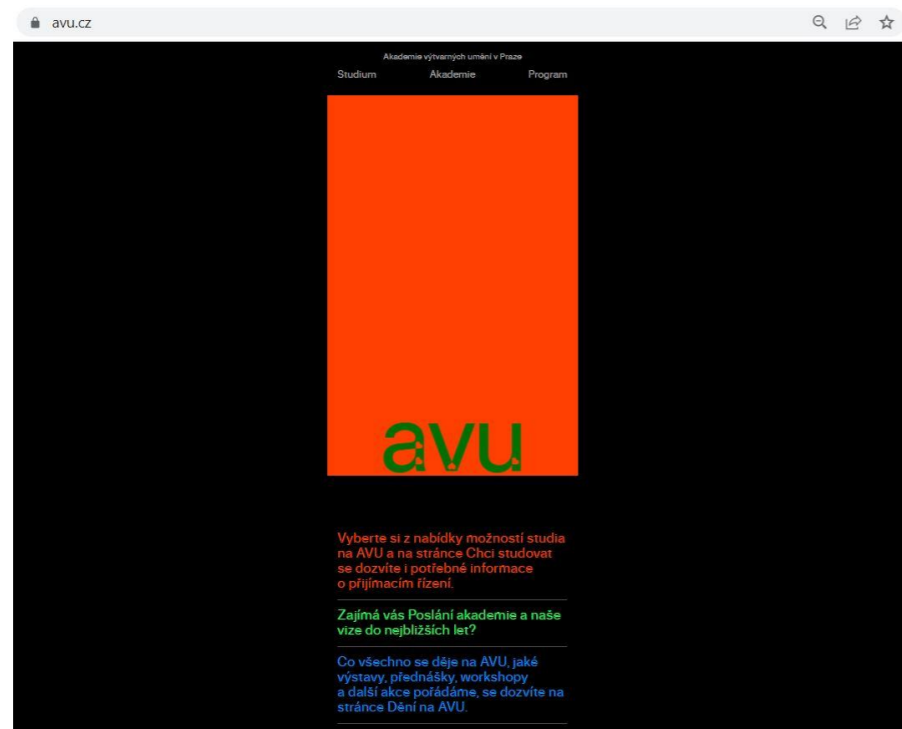
W  
W W  
W W W



<https://avu.cz>



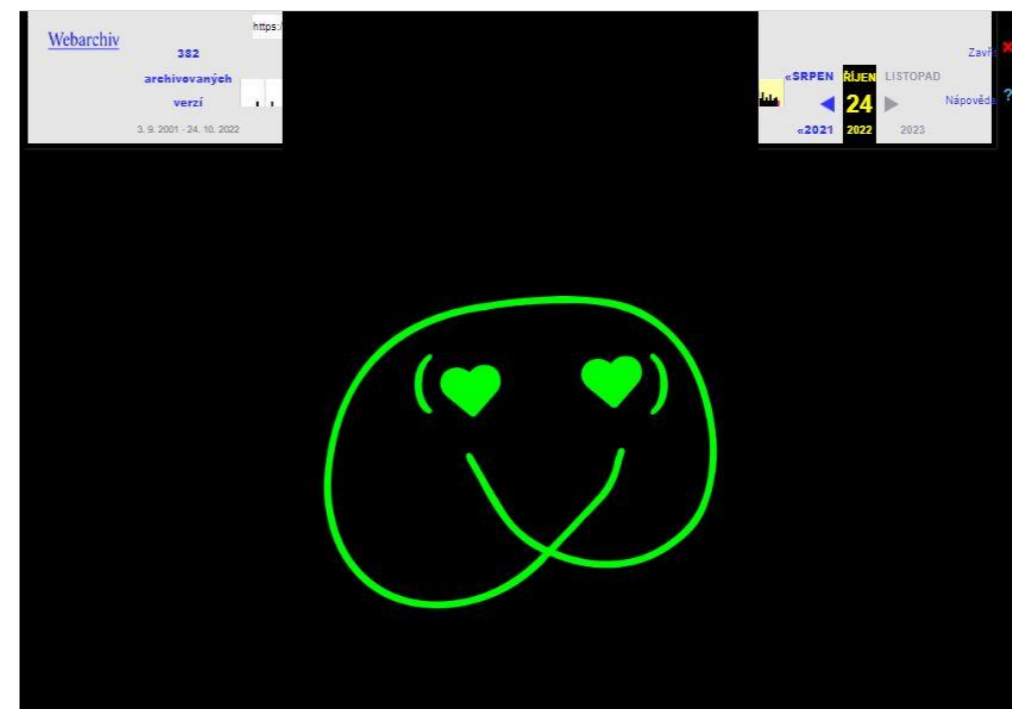
2002



2022



2021



2022

# *Legislativa*

- Autorský zákon 121/2000 Sb., Knihovní licence § 37 rozmnoženina díla, archivní a konzervační potřeby
- volný přístup – smlouva nebo Creative Commons  
**0,4 % archivu**
- trojnovela – právní úprava web-harvestingu
- implementace směrnice o autorském právu na jednotném digitálním trhu do české legislativy

# *Jak archivujeme? Nechte se Webarchivovat!*

- celoplošné sklizně  
.cz, CZ.NIC, cca 1,4 milionů domén
- výběrové sklizně  
kurátorský přístup
- tematické kolekce  
událost, téma

## Nechte se Webarchivovat!

Pokud hledáte certifikát či jiné naše grafické materiály pro umístění na vašem webu pokračujte [zde](#)

[Přidejte web](#) / [Creative Commons](#) / [Výběr zdrojů](#) / [Pomozte nám uchovat český web](#) / [FAQ](#)

### *Přidejte web*


URL

Mohu jednat za uvedené zdroje  Zdroj s licencí Creative Commons

Jméno

Kontaktní e-mail

Poznámka

Nejsem robot   
Ochrana soukromí · Smluvní podmínky

**Přidat web**

*Strategie budování sbírky Webarchivu*

<https://webarchiv.cz/static/www/download/collection-policy.pdf>

# Výběrové sklizně

- online přístup (smlouva, Creative Commons), 5300 zdrojů
- pravidelné a dlouhodobé sklizení
- katalogizace

## Webarchiv podle oborů

Přehled nasmlouvaných webů dle oborového třídění:

Vše 5416 / [Antropologie, etnografie](#) 165 / [Beletrie](#) 29 / [Biologické vědy](#) 230 / [Chemie, Krystalografie, Mineralogické vědy](#) 47 / [Divadlo, film, taneč.](#) 170 / [Ekonomické vědy, obchod](#) 325 / [Filozofie a náboženství](#) 241 / [Fyzika a příbuzné vědy](#) 91 / [Geografie, Geologie, Vědy o Zemi](#) 416 / [Historie a pomocné historické vědy, Biografické studie](#) 307 / [Hudba](#) 154 / [Jazyk, lingvistika, literární věda](#) 221 / [Knihovnictví, informatika, všeobecné, referenční literatury](#) 328 / [Lékařství](#) 290 / [Literatura pro děti a mládež](#) 1 / [Matematika](#) 45 / [Politické vědy, \(Politologie, politika, veřejná správa, vojenství\)](#) 357 / [Právo](#) 151 / [Psychologie](#) 84 / [Sociologie](#) 317 / [Technika, technologie, inženýrství](#) 289 / [Tělesná výchova a sport, Recreace](#) 216 / [Umění, architektura](#) 361 / [Výchova a vzdělávání](#) 241 / [Výpočetní technika](#) 162 / [Zemědělství](#) 178

### Biologické vědy /

Vše 230 / [Biochemie, Molekulární biologie, Biofyzika](#) 7 / [Biologické vědy](#) 14 / [Biotechnologie, Genetické inženýrství](#) 5 / [Botanika](#) 24 / [Buněčná biologie, Cytologie](#) 4 / [Mikrobiologie](#) 2 / [Obecná biologie](#) 5 / [Obecná botanika](#) 7 / [Obecná ekologie](#) 5 / [Obecná genetik, Obecná cytogenetika, Evoluce](#) 4 / [Obecná zoologie](#) 1 / [Virologie](#) 1 / [Životní prostředí a jeho ochrana](#) 109 / [Znečištění a poškození životního prostředí](#) 10 / [Zoologie](#) 31

Zobrazení: vizuální, textové



[Achatina web - chov afrických šneků](#) [Achatina achatina](#)



[Acta Entomologica Musei Nationalis Pragae](#)



[Acta Universitatis Agriculturae et Silviculturae Mendelianae Brunensis](#)



[Agentura ochrany přírody a krajiny ČR](#)



[Alky roku](#)



[Alliance Nebe nad Ostravou](#)



[Aranearium](#)



[Arnika](#)

# Tematické sbírky

- aktuální události (volby, povodně)
- dlouhodobě budované sbírky (česká média)
- spolupráce s IIPC

[úvod](#) / [o Webarchivu](#) / [katalog stránek](#) / [tematické sbírky](#) / [přidat web](#)

Hledání



## Sweb.cz

Klíčová slova sklizně:

[blogy](#)



Tematická sbírka shromažďuje všechny dostupné weby (cca 85 000) na doméně sweb, která ukončuje svůj provoz dne 30.11.2022. Služba sweb byla důležitým zástupcem tuzemských poskytovatelů webhostingů zdarma. Sweb funguje minimálně od roku 2000, kdy byla pořízena první archivovaná verze. Kolekce tak představuje důležitý obraz o raném tuzemském internetu. Zřejmě v roce 2009 došlo také ke změně v subdoméně - adresy sweb.cz/UZIVATEL byly deaktivovány a nahradila je varianta UZIVATEL.sweb.cz. Tato kolekce obsahuje pouze URL dostupné v říjnu 2022, přestože v archivu lze nalézt i weby se starou strukturou adresy.

[Seznam URL v CSV](#)

### [Dagblog : blog nejen o programování pro všechny kodery a koderky](#)

[dagblog.sweb.cz](#) [aktuální verze](#)

### [Databáze divadelních her a scének](#)

[divadelnihery.sweb.cz](#) [aktuální verze](#)

### [Fotografie Aleš Uhlíř](#)

[al.uhlir@sweb.cz](#) [aktuální verze](#)

### [Fraktály a chaos](#)

[chaos.fraktaly.sweb.cz](#) [aktuální verze](#)

### [Giza, Gizeh, Jizah, al-Jizeh & sluneční soustava, Sun system](#)

[fotografiek.sweb.cz](#) [aktuální verze](#)

### [Jawetta](#)

[jaweta.sweb.cz](#) [aktuální verze](#)

# *Data a metadata*

**české webové zdroje** - bohemikální charakter (území, autorství, jazyk, obsah), nejen na české doméně

sklizení: Heritrix 3.4, Webrecorder, archiveweb page / browsertrix (sociální média, dynamický obsah)

zpřístupnění: OpenWayback 3.0

ukládání: WARC

kurátorský nástroj: Seeder, <https://github.com/webarchivcz/>

25–40 TB ročně

## *metadata:*

- popisná / bibliografická
- technická a administrativní

# Metadata

## *Bibliografická* knihovní systém **Aleph** **MARC 21, RDA**, od 2015

*Katalogizační manuál*

<https://webarchivcz.github.io/katalogizacni-manual/>

*Technická a administrativní*  
údaje získané během sklizně: datum  
sklizně, typ, autor ad.

*Metodika:*

<https://invenio.nusl.cz/record/432325?ln=cs>

### WebArchiv - Úplné zobrazení záznamu

Zvolte formát: Standardní -- Katalogizační záznam -- Stručný záznam -- MARC -- [Citace](#)

#### Záznam 1 z 2

```
LDR -----nai-a22-----i-4500
FMT SE
BAS |a 01
BAS |a WAR
001 web20172886912
003 CZ-PrNK
005 20170320175238.0
006 m-----o-d-----
007 cr-cn-
008 170320c:20069999xr--x-w-o-----0---b2cze--
015 |a cnb002886912
040 |a ABA001 |b cze |e rda
0410 |a cze |a eng
043 |a e-xr--- |b e-xr-pl |2 czenas
072 7 |a 785 |x Instrumentální hudba. Symfonická hudba. Hudba pro více nástrojů |2 Konspekt |9 9
072 9 |a 784-788 |x Instruments and instrumental ensembles and their music |2 Conspectus |9 9
080 |a 785.11.071 |2 MRF
080 |a 785.11 |2 MRF
080 |a 78 |2 MRF
080 |a 78-027.22 |2 MRF
080 |a 78.082.4 |2 MRF
080 |a (0.034.2)004.738.12 |2 MRF
1102 |a Plzeňská filharmonie |7 ko2004209449 |4 pbl
24510 |a Plzeňská filharmonie
264 1 |a Plzeň : |b Plzeňská filharmonie, |c [2006]-
300 |a 1 online zdroj
310 |a Aktualizováno nepravdělně
336 |a text |b txt |2 rdaccontent
337 |a počítač |b c |2 rdamedia
338 |a online zdroj |b cr |2 rdacarrier
520 |a Stránky zahrnují základní informace o Plzeňské filharmonii, o orchestru, údaje koncertech a vystoupeních.
588 |a Název z titulní obrazovky (náhled z 20.03.2017) |5 CZ-PrNK
61027 |a Plzeňská filharmonie |7 ko2004209449 |2 czenas
65007 |a symfonické orchestry |7 ph135216 |2 czenas
65007 |a orchestrální hudba |7 ph123783 |2 czenas
65007 |a hudba |7 ph114719 |2 czenas
65007 |a hudební život |7 ph135177 |2 czenas
65007 |a koncerty |7 ph121790 |2 czenas
65009 |a symphonic orchestras |2 eczenas
65009 |a orchestral music |2 eczenas
65009 |a music |2 eczenas
65009 |a musical life |2 eczenas
65009 |a concerts |2 eczenas
651 7 |a Plzeň (Česko) |7 ge130439 |2 czenas
651 9 |a Plzeň (Czechia) |2 eczenas
655 7 |a www dokumenty |7 fd186892 |2 czenas
655 9 |a www documents |2 eczenas
85640 |u http://www.plzenskafilharmonie.cz |q text/html |4 N
85640 |u http://wayback.webarchiv.cz/wayback/http://www.plzenskafilharmonie.cz |q text/html |2 archivní verze stránek |4 N
929 |a Text
SYS 002886912
```

[ [http://aleph.nkp.cz/F/?func=direct&doc\\_number=002886912&local\\_base=Web](http://aleph.nkp.cz/F/?func=direct&doc_number=002886912&local_base=Web) ]

[Citace](#)

```
1 {
2   "_id": "5dcd98a695bcf5a1a194f0be",
3   "recType": "harvest",
4   "author": "NKCR",
5   "date": "2019-11-14T19:10:46.983Z",
6   "standard": "Grainery 0.4",
7   "harvest": {
8     "harvestPrefix": {
9       "harvestNameStand": "V6M_2017-10-05",
10      "harvestFromWarcinfo": "V6M_2017-10-05",
11      "harvestNameFNtrunc": "V6M_2017-10-05",
12      "harvestDirsName": "V6M_2017-10-05",
13      "harvestType": "?výběrová",
14      "harvestSuffix": ["V6M", "2017-10-05"]
15    },
16    "date": "2017-10-05T11:26:00.000Z",
17    "harvestID": "105f23c9-b037-4c1d-901c-dcf272877d9f",
18    "size": 618029,
19    "warcsNumber": 12092
20  },
21  "harvestCrawl": {
22    "logs": true,
23    "path": "logs/crawl",
24    "fileName": ["crawler00.tar.gz", "crawler01.tar.gz", "crawler03.tar.gz"]
25  },
26  "paths": {
27    "cdxsID": ["105f23c9-b037-4c1d-901c-dcf272874d9f"],
28    "warcsID": ["105f23c9-b037-4c1d-901c-dcf272877d9f"],
29    "warcsFileNames": ["V6M_2017-10-05-crawler00.webarchiv.cz-warcs.gz"]
30  },
31  "revision": {
32    "dateOfValidation": "2019-12-04T19:10:46.983Z",
33    "statusOfValidation": "NA",
34    "nextLastDateOfValidation": "2021-12-03T19:10:46.983Z",
35    "hashOrig": "NA",
36    "hashLast": "NA",
37    "commentaries": { "exists": false, "text": "NA" }
38  }
39 }
```



# Jak data využívat?


Webarchiv <http://www.laboratory.cz/projects/1853-moderni-zena-architektka-vnitrek> Jít

6 archivovaných verzí 23. 4. 2022 - 26. 9. 2022

ČERVENEC ZÁŘÍ ŘÍJEN 2021 2022 2023 Zavřít Nápověda

laboratoř index < 1187 / 1187 en

**moderní žena - architektka vnitřek** (2022)  
výchozím bodem publikace je emancipační potenciál ženských protagonistek moderny a bauhausu, jejich role a úspěchy na příkladech z české republiky, německa, polska, maďarska a estonska v posledních více než sto letech ... koncept a design: petr babák, martin ponec (laboratoř), foto: eda babák  
[http://www.umprum.cz/web/c...\\_kika-12197](http://www.umprum.cz/web/c..._kika-12197)



## Umění, architektura /

Vše 361 / [Architektura](#) 56 / [Fotografie](#) 46 / [Grafické umění](#) [Grafika](#) 13 / [Kresba](#) [Umělecká řemesla](#) 15 / [Malířství](#) 13 / [Sochařství](#) [keramika](#) [porcelán](#) [umělecké zpracování kovů](#) 17 / [Umění](#) 75 / [Územní plánování](#) [Urbanismus](#) [Památková péče](#) 22 / [Výtvarné umění](#) 103

Zobrazení: vizuální, [textové](#)



[14|15 Bařův institut](#)



[abcd : art brut](#)



[Action galleries](#)



[Adolf Wölfli : stvořitel universa](#)



[Akademie výtvarných umění v Praze](#)



[Alena Schulz](#)



[Alice Nikitinová](#)



[Alšova jihočeská galerie v Hluboké nad Vltavou : AIG](#)



[Amatérská obrazárna](#)



[Archeologické oddělení NPU Praha](#)



[Architecture week](#)



[Archiweb.cz](#)

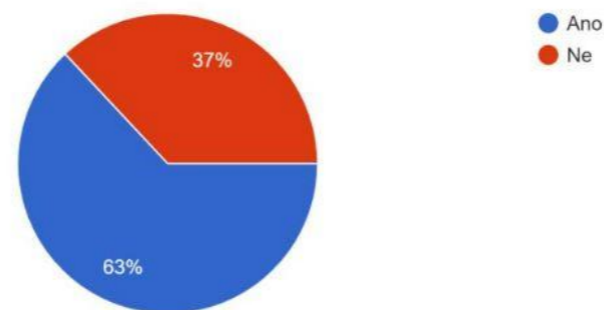
1 2 3 4 5 6 7 8 9 10 ... 31

# Potřeby uživatelů webového archivu

- povědomí o Webarchivu, zájem o data a nároky uživatelů
- formáty a nástroje
- legislativní restrikce, omezený přístup k datům
- vývoj uživatelsky přívětivých nástrojů pro pokročilou práci s daty

Víte o existenci českého webového archivu NK ČR - <https://webarchiv.cz?>

146 odpovědí



*Používáte Webarchiv? Průzkum potřeb uživatelů českého webového archivu ProInflow, <https://doi.org/10.5817/ProIn2022-2-2>*

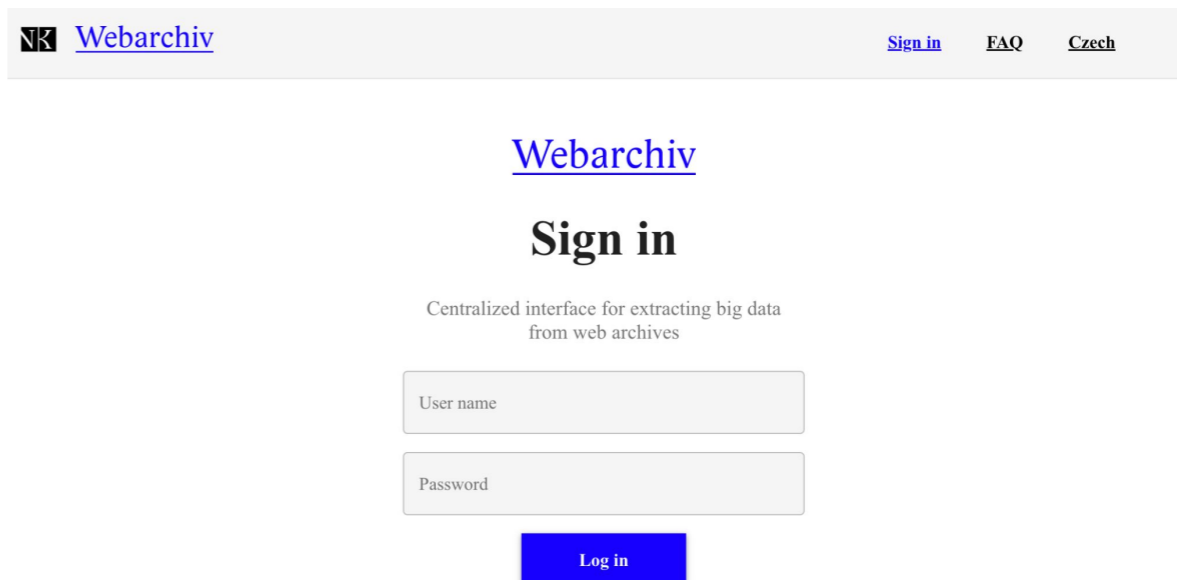
# *Vývoj centralizovaného rozhraní pro vytěžování velkých dat z webových archivů - WACloud*

- **Národní knihovna ČR**  
(data, zkušenosti s archivací webu, infrastruktura)
- **Západočeská univerzita v Plzni, Fakulta aplikovaných věd, Katedra kybernetiky**  
(strojové zpracování dat, přístupy založené na hlubokých neuronových sítích pro klasifikaci dokumentů)
- **Sociologický ústav AV ČR, v.v.i.**  
(výzkumníci v oblasti sociálních věd)

MK ČR – Program na podporu aplikovaného výzkumu a experimentálního vývoje národní a kulturní identity, 2018 – 2022 (NAKI II)

# *Cíle a výstupy výzkumného projektu*

- umožnit přístup k datům Webarchivu badatelské komunitě
- rozšíření technologické infrastruktury pro analytické zpracování velkého objemu dat
- přehodnocení právního rámce



[Webarchiv](#) [Sign in](#) [FAQ](#) [Czech](#)

[Webarchiv](#)

**Sign in**

Centralized interface for extracting big data  
from web archives

User name

Password

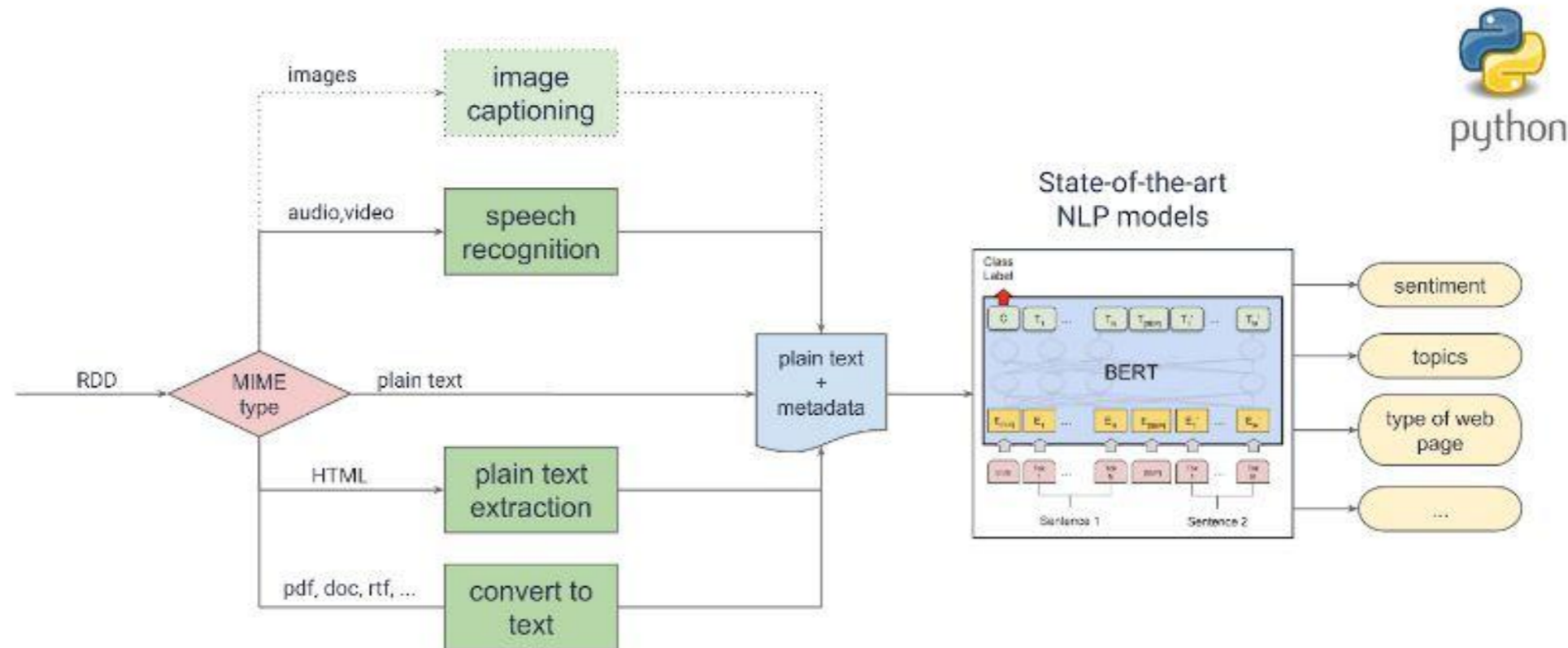
Log in

- uživatelsky přívětivé grafické rozhraní
- fasetový a fulltextový vyhledávač, který umožní výzkumníkům definovat část dat pro jejich výzkum, exportní aplikace

# Infrastruktura, technologie, UI

- **infrastruktura:** HADOOP a Hbase
- **technologie:** WARC – PySpark – Python – Scikit-Learn – Apache Hbase – Hadoop – Apache Spark – TensorFlow – PyTorch

## AI Algorithms Pipeline



analýza témat textových dokumentů a jejich automatická detekce, analýza audiosouborů, přístupy založené na hlubokých neuronových sítích pro klasifikaci dokumentů, automatické analytické nástroje, automatické přiřazování metadat

# Dotaz, filtry, nastavení limitů, sklizně

topics:"knihovna a muzeum" OR topics:"literatura" AND date:[2018-11-01T00:00:00Z TO 2022-11-30T00:00:00Z] AND url:/.\*nkp.cz.\*/ AND sentiment:[0 TO 1]

The screenshot shows the Webarchiv search interface. On the left is a sidebar with filters: **FILTRY** (Téma, Typ stránky, Datum sklizně, URL, Sentiment) and **NASTAVENÍ LIMITŮ** (Stop slova, Počet záznamů). The main area is titled **DOTAZ** and contains the search query: `topics:"knihovna a muzeum" OR topics:"literatura" AND date:[2018-11-01T00:00:00Z TO 2022-11-30T00:00:00Z] AND url:/.*nkp.cz.*/ AND sentiment:[0 TO 1]`. Below the query are buttons for **Importovat** and **Exportovat**, and a section for **Logické operátory** (AND, OR, NOT, (, )). The **SKLIZNĚ** section displays four harvest cards:

- Te** (checkbox):
  - Název sklizně: Topics\_2022-09-T-TestRozhrani
  - Velikost: -
  - Počet WARCu: 126 646
  - Datum spuštění: 12. 9. 2022
- Se** (checkbox):
  - Název sklizně: Serials-2021-02-1M-2M\_OneShot-crawler00
  - Velikost: -
  - Počet WARCu: 4 735 803
  - Datum spuštění: 26. 2. 2021
- Se** (checkbox):
  - Název sklizně: Serials-2021-01-1M-4M\_OneShot-crawler00
  - Velikost: -
  - Počet WARCu: 4 315 890
  - Datum spuštění: 22. 1. 2021
- To** (checkbox):
  - Název sklizně: Topics-2018-01-T2018VolPrezII-crawler01\_
  - Velikost: -
  - Počet WARCu: 11 852
  - Datum spuštění: 12. 1. 2018

A blue **Pokračovat** button is located at the bottom right.

# Analytické dotazy - frekvence, kolokace, sítě, fulltext

Webarchiv [Nový dotaz](#) [Oblíbené](#) [Moje dotazy](#) [FAQ](#) [Uživatel](#)

> DOTAZ

topics:"knihovna a muzeum" OR topics:"literatura" AND data:[2018-11-01T00:00:00Z TO 2022-11-30T00:00:00Z] AND url:/\*nkp.cz/\* AND sentiment:[0 TO 1]

Dotaz je **validní** a zpracuje se s ním **100 záznamů**

Logické operátory: AND OR NOT ( )

SKLIZNĚ [Upravit](#)

ANALYTICKÉ DOTAZY [+ Přidat další dotaz](#)

Typ dotazu Frequency	Zadejte text pro vyhledávání	<a href="#">+</a>	Seznam výrazů: knihovna	<a href="#">Odstranit</a>
	Počet záznamů 2			
Typ dotazu Colocation	Zadejte text pro vyhledávání	<a href="#">+</a>	Seznam výrazů: čtenář	<a href="#">Odstranit</a>
	<input checked="" type="checkbox"/> Přidat kontext		Počet slov kontextu 3	
Typ dotazu Network	Zdrojové uzly	<a href="#">+</a>	Zdrojové uzly nkp.cz	<a href="#">Odstranit</a>
	<input type="checkbox"/> Seskupit podle domény			
	Cílové uzly	<a href="#">+</a>	Cílové uzly nkp.cz	
	<input type="checkbox"/> Seskupit podle domény			
Typ dotazu Fulltext	Zadejte text pro vyhledávání	<a href="#">+</a>	Seznam výrazů: kniha	<a href="#">Odstranit</a>
	Počet záznamů 5	Řazení dle Rok vzestupně	Formát JSON	

[Vyhledat](#)

# Sítě

```
[ {
  "year" : 2021,
  "nodes" : [ {
    "name" : "dnnt.nkp.cz/seznam.html",
    "links" : [ {
      "name" : "https://aleph.nkp.cz/F/?func=file&file_name=find-b&local_base=DNT",
      "count" : 1
    }, {
      "name" : "https://aleph.nkp.cz/F/?func=file&file_name=find-b&local_base=DNTP",
      "count" : 1
    }, {
      "name" : "https://aleph.nkp.cz/F/?func=file&file_name=help-1",
      "count" : 1
    }, {
      "name" : "https://aleph.nkp.cz/web/download/sdnt.zip",
      "count" : 1
    }, {
      "name" : "https://aleph.nkp.cz/web/download/sdnt_navrhly.zip",
      "count" : 1
    }, {
      "name" : "https://dnnt.nkp.cz/autor.html",
      "count" : 1
    }, {
      "name" : "https://dnnt.nkp.cz/ctenar.html",
      "count" : 1
    }, {
      "name" : "https://dnnt.nkp.cz/downloads.html",
      "count" : 1
    }, {
      "name" : "https://dnnt.nkp.cz/downloads/vyrazeni.xlsx",
      "count" : 1
    }, {
      "name" : "https://dnnt.nkp.cz/downloads/zarazeni-hrom.xlsx",
      "count" : 1
    }, {
      "name" : "https://dnnt.nkp.cz/downloads/zarazeni.xlsx",
      "count" : 1
    }, {
      "name" : "https://dnnt.nkp.cz/index.html",
      "count" : 1
    }
  ]
}
```

# Kolokace

```
{
  "full.nkp.cz/nkkr/knihovna141/141130.htm" : {
    "000" : [ "s celkem 25 000 <em>čtenáři</em>, což odpovídá" ],
    "fondy" : [ "\můstky\ mezi fondy a <em>čtenáři</em>, jsou hlavními" ],
    "hlavními" : [ "a <em>čtenáři</em>, jsou hlavními nástroji knihkupce zejména" ],
    "jednoho" : [ "vedou <em>čtenáře</em> od jednoho textu k druhému," ],
    "knihy" : [ "že doporučíme určité knihy, <em>čtenářům</em> usnadňujeme výběr" ],
    "malé" : [ "online nástroji. Pro malé <em>čtenáře</em> byl vytvořen" ],
    "měla" : [ "nabídka však měla u <em>čtenářů</em> úspěch. Co" ],
    "navede" : [ "označení snadno <em>čtenáře</em> navede k tomu, co" ],
    "odpovídá" : [ "000 <em>čtenáři</em>, což odpovídá asi 17 %" ],
    "pohybu" : [ "v pohybu, což má <em>čtenáře</em> \vnořit do" ],
    "setkání" : [ "jezdí na setkání se <em>čtenáři</em>, účastní se" ],
    "snadno" : [ "online. Účinné označení snadno <em>čtenáře</em> navede k" ],
    "usnadňujeme" : [ "určité knihy, <em>čtenářům</em> usnadňujeme výběr a pomáháme" ],
    "vedou" : [ "materiálu\". Odkazová slova vedou <em>čtenáře</em> od jednoho" ],
    "vhodné" : [ "zájmům <em>čtenářů</em>. Je vhodné najít takový informační" ],
    "vnořit" : [ "což má <em>čtenáře</em> \vnořit do textového materiálu\." ],
    "vytvořen" : [ "malé <em>čtenáře</em> byl vytvořen blog \Čte se" ],
    "zájmům" : [ "měl odpovídat rozličným zájmům <em>čtenářů</em>. Je vhodné" ],
    "úspěch" : [ "měla u <em>čtenářů</em> úspěch. Co se týká" ],
    "účastní" : [ "setkání se <em>čtenáři</em>, účastní se festivalů apod." ]
  },
  "full.nkp.cz/nkkr/knihovnaplus61/chalupova.htm" : {
    "mohou" : [ "noví <em>čtenáři</em> se mohou přihlásit zdarma, uděluje" ],
    "například" : [ "<em>čtenářů</em>, ale i například ve spolupráci s" ],
    "nejmladší" : [ "možnostech pořadatelů. Na nejmladší <em>čtenáře</em> se zaměřili" ],
    "noví" : [ "burzy vyřazených knih, noví <em>čtenáři</em> se mohou" ],
    "představil" : [ "<em>čtenář</em> se zde představil zpěvák Sámeer Issa." ],
    "skupiny" : [ "na různé věkové skupiny <em>čtenářů</em>, ale i" ],
    "slavný" : [ "knih; jako první slavný <em>čtenář</em> se zde" ],
    "zaměřili" : [ "nejmladší <em>čtenáře</em> se zaměřili rovněž pořadatelé soutěže" ]
  },
  "full.nkp.cz/nkkr/knihovnaplus92/senf.htm" : {
    "besedou" : [ "ke <em>čtenářům</em> před besedou. Vystoupení byla snímána" ],
    "chtějí" : [ "opravdu chtějí na své <em>čtenáře</em> působit sympaticky," ],
    "konfliktním" : [ "zabýval Rozhovorem s konfliktním <em>čtenářem</em>. Na něm", "či" ],
    "kontaktu" : [ "Při prvním kontaktu se <em>čtenářem</em> nejvíce o" ],
    "nejvíce" : [ "kontaktu se <em>čtenářem</em> nejvíce o knihovníkovi vypoví" ],
    "něm" : [ "konfliktním <em>čtenářem</em>. Na něm již všichni mohli" ],
    "obsahu" : [ "pro <em>čtenáře</em>. V obsahu nemohly chybět ani" ],
    "potřebuje" : [ "kdo ji dnes potřebuje. <em>čtenáři</em> či sami" ],
    "promluvu" : [ "knihovně, úvodní promluvu ke <em>čtenářům</em> před besedou." ],
    "proškolily" : [ "konfliktním <em>čtenářem</em> se proškolily přímo v městských" ],
    "přitažlivosti" : [ "čtivosti a přitažlivosti pro <em>čtenáře</em>. V obsahu" ],
    "působit" : [ "na své <em>čtenáře</em> působit sympaticky, chtějí být" ],
    "sami" : [ "potřebuje. <em>čtenáři</em> či sami spisovatelé? Věřme, že" ]
  }
}
```



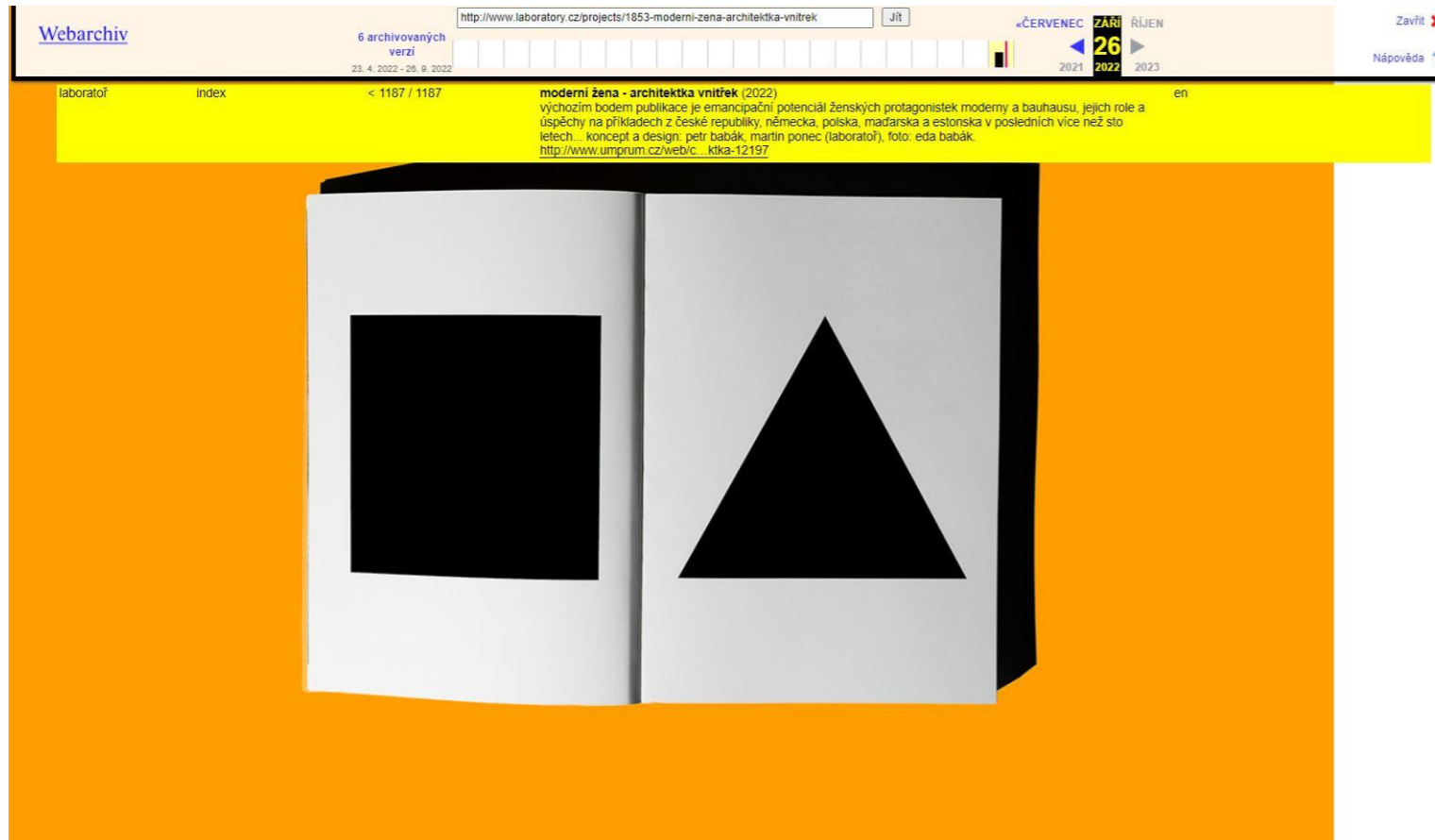
## Fulltext

```
[ {
  "id" : "0a227c78-f372-4410-9446-3a105ba56206",
  "url" : "skip.nkp.cz/bulletin/popis_kkl.htm",
  "urlDomain" : "skip.nkp.cz",
  "urlDomainTopLevel" : "nkp.cz",
  "title" : "Bulletin SKIP, Vydavatelské údaje",
  "language" : "cs",
  "plainText" : "Katalog knihovny knihovnické literatury v úplnosti od r. 1990. Průběžně se doplňuje a informuje o tom, jaké údaje lze použít pro prohlížení\nzadáním rešeršního dotazu - volba Vyhledávání",
  "year" : 2021,
  "headlines" : null,
  "links" : null,
  "linksDomain" : null,
  "linksDomainTopLevel" : null,
  "topics" : [ "literatura" ],
  "sentiment" : 0.8981181826809512
}, {
  "id" : "203f23a6-894c-424d-863f-c54d5f7e00f8",
  "url" : "full.nkp.cz/nkkr/nkkr0103/0103205.html",
  "urlDomain" : "full.nkp.cz",
  "urlDomainTopLevel" : "nkp.cz",
  "title" : "Personálie v knihovnách (J. Bínová)",
  "language" : "cs",
  "plainText" : "Cílem česko-slovenského semináře pořádaného 5.-6. září 2001 v Moravské zemské knihovně regionálních bibliografů v Hradci nad Moravicí uspořádaného v roce 1975 Slezským muzeem v Olomouci (v některých případech) velmi rozsáhlé personální bibliografie vydané v posledních letech, např. o Vlastimíru z NK ČR se ve svém vystoupení věnoval problematice souborů autorit a stručně představil projekt (t). Součástí ikonografických sbírek je i soupis pamětních desek a tabulí.\nShromážděný materiál je z roku 1994 začala vyrůstat generace slovenských regionálních a personálních bibliografů v kraji, která nazvala svůj referát Elektronické zdroje personální bibliografie v ČR, referát byl jejich skutečným výsledkem regionálních knihoven: Ľudmily Strónerové z KK Trenčín, Eleny Matiskové z KK Zvolen a Jany Amrichové v Martine, s informací o zásadních rozhodnutích IFLA v letošním roce. IFLA od roku 2003 přestala publikovat bibliografie - J. Amrichové, E. Matiskové, Ľ. Strónerové - budou publikovány rovněž v časopise Čtenář.",
  "year" : 2021,
  "headlines" : [ "Personálie v knihovnách" ],
  "links" : [ "http://full.nkp.cz/nkkr/NKKR0103.html", "http://full.nkp.cz/nkkr/NKKR0103.html" ],
  "linksDomain" : [ "full.nkp.cz", "full.nkp.cz" ],
  "linksDomainTopLevel" : [ "nkp.cz", "nkp.cz" ],
  "topics" : [ "Slovensko", "literatura" ],
  "sentiment" : 0.8525089837981403
}
```

## Frekvence

```
{
  "full.nkp.cz/nkkr/knihovnaplus51.html" : {
    "ktere" : 7,
    "knihovny" : 5
  },
  "skip.nkp.cz/bulletin/popis_kkl.htm" : {
    "1990" : 4,
    "zaznamy" : 4
  },
  "wwwold.nkp.cz/pages/page.php3?page=slov_vznik_a_vyvoj.htm" : {
    "knihovna" : 9,
    "knihovny" : 6
  },
  "nkp.cz/?sindex=2" : {
    "kultury" : 2,
    "lubomir" : 2
  },
  "wwwold.nkp.cz/pages/page.php3?page=slov_fond.htm" : {
    "knihovny" : 8,
    "knihovna" : 4
  },
  "skip.nkp.cz/mezsve.htm" : {
    "2005" : 42,
    "informacni" : 21
  },
  "text.nkp.cz/o-knihovne/o-knihovne/odborne-cinnosti/doplnovani-formacni" : {
    "narodni" : 13,
    "cr" : 10
  },
  "full.nkp.cz/nkkr/knihovna132/132005.htm" : {
    "naprstkova" : 48,
    "muzea" : 45
  },
  "text.nkp.cz/o-knihovne/zakladni-informace/verejne-zakazky/o-knihovne" : {
    "narodni" : 13,
    "cr" : 10
  },
  "full.nkp.cz/nkkr/nkkr0401/0401045.html" : {
    "katalogu" : 11,
    "narodni" : 11
  },
}
```

# Archivní kopie - Dataset



```
{
  "full.nkp.cz/nkkr/knihovnaplus51.html" : {
    "ktere" : 7,
    "knihovny" : 5
  },
  "skip.nkp.cz/bulletin/popis_kkl.htm" : {
    "1990" : 4,
    "zaznamy" : 4
  },
  "wwwold.nkp.cz/pages/page.php3?page=slov_vznik_a_vyvoj.htm" : {
    "knihovna" : 9,
    "knihovny" : 6
  },
  "nkp.cz/?sindex=2" : {
    "kultury" : 2,
    "lubomir" : 2
  },
  "wwwold.nkp.cz/pages/page.php3?page=slov_fond.htm" : {
    "knihovny" : 8,
    "knihovna" : 4
  },
  "skip.nkp.cz/mezsve.htm" : {
    "2005" : 42,
    "informacni" : 21
  },
  "text.nkp.cz/o-knihovne/o-knihovne/odborne-cinnosti/doplnovani-for" : {
    "narodni" : 13,
    "cr" : 10
  },
  "full.nkp.cz/nkkr/knihovna132/132005.htm" : {
    "naprstkova" : 48,
    "muzea" : 45
  },
  "text.nkp.cz/o-knihovne/zakladni-informace/verejne-zakazky/o-knih" : {
    "narodni" : 13,
    "cr" : 10
  },
  "full.nkp.cz/nkkr/nkkr0401/0401045.html" : {
    "katalogu" : 11,
    "narodni" : 11
  },
}
```

*Děkuji za pozornost*

W  
W W  
W W W

[www.webarchiv.cz](http://www.webarchiv.cz)

[www.facebook.com/webarchivcz](https://www.facebook.com/webarchivcz), [https://twitter.com/webarchiv\\_cz](https://twitter.com/webarchiv_cz)

[marie.haskovcova@nkp.cz](mailto:marie.haskovcova@nkp.cz)