

MUNI  
ARTS

# Analýza textů závěrečných prací studentů FF MU

PhDr. Markéta Bočková

# O CO NÁM VLASTNĚ JDE...



CO KTERÁ ČÍSLA VLASTNĚ ŘÍKAJÍ?



A PROČ NÁM NESTAČÍ  
PŮJČOVANOST?

nejpůjčovanější znamená  
nejužitečnější?

vybíráte si věci co jsou  
nejprodávanější nebo na doporučení?



A EXISTUJE OPRAVDU  
MOŽNOST JAK TO  
ZJISTIT?

teda až budeme vědět co vlastně  
chceme?

# RELEVANCE / PERTINENCE



## RELEVANCE

formální

- shoda dotazu se selekčními údaji nalezeného dokumentu

věcná

- shoda dotazu s obsahem z pohledu uživatele



## PERTINENCE

zhodnocení z pohledu uživatele

kritéria:

- Tématicnost- do jaké míry obsah dokumentu odpovídá tématu
- Spolehlivost- do jaké míry je obsah dokumentu vnímán jako pravdivý, přesný a spolehlivý
- Pochopitelnost- do jaké míry je obsah uživateli srozumitelný
- Novost- do jaké míry přináší dokument uživateli nové informace
- Rozsah- do jaké šíře a hloubky je téma zpracováno

# 2 kroky vpřed, 1 vzad, co týden nový problém

## Co máme dělat?

- protože nevíme jak se k datům dostat
- protože nevíme v čem to budeme dělat
- protože nevíme kdo to bude dělat

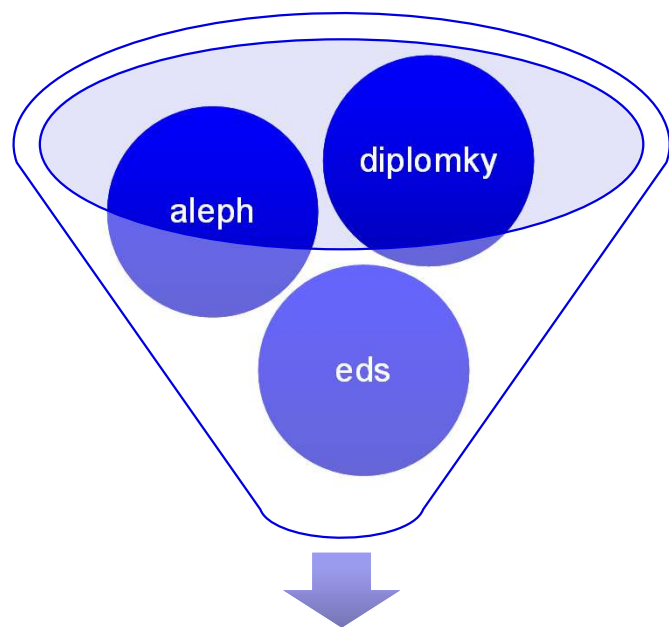
## Jak to uděláme?

- Python / Excel
- SQL / Python
- Python – jupyter notebook

## ...kdo to vlastně udělá?

- dis Markét
- kolega co má dělat úplně něco jiného?
- Markét co nemá co dělat a kolega co má dělat něco jiného?
- nikdo

# Výstup



Dataset pro analýzu (excel)

ISSN	ISBN	DOI	URL
máme nemáme	máme nemáme	máme nemáme	co navštěvují
akvizice	akvizice	akvizice lekce IV	lekce IV

# Analýza závěrečných prací

bc, mgr, nav\_mgr, rigo  
po exportu (hrubá data)

279 013 řádků

**isbn**

19 731

**issn**

3 102

**doi**

8 762

**url**

41 104

# výstup ze závěrečných prací.txt

```
In [71]: #načítáme id diplomky
df = pd.read_csv("id_diplomky.csv", encoding = "utf-8", sep= "\t", header= None,
result = df.head(10)
print("Toto je ukázka prvních 10 řádků v souboru id.txt")
print(result)
```

Toto je ukázka prvních 10 řádků v souboru id.txt

	type	id	file name
0	url	<a href="http://www.malice-radhost.cz">http://www.malice-radhost.cz</a>	bc_274054_t_121322316.txt
1	url	<a href="https://www.kpufo.cz">https://www.kpufo.cz</a>	bc_274054_t_121322316.txt
2	url	<a href="https://moravskoslezsky.denik.cz">https://moravskoslezsky.denik.cz</a>	bc_274054_t_121322316.txt
3	url	<a href="http://www.frenstatpr.ez">http://www.frenstatpr.ez</a>	bc_274054_t_121322316.txt
4	url	<a href="https://www.idnes.cz">https://www.idnes.cz</a>	bc_274054_t_121322316.txt
5	url	<a href="https://moravskoslezsky.denik.cz">https://moravskoslezsky.denik.cz</a>	bc_274054_t_121322316.txt
6	issn	1213-	bc_274054_t_121322316.txt
7	issn	1213-8371	bc_274054_t_121322316.txt
8	issn	1213-8371	bc_274054_t_121322316.txt
9	issn	1213-8371	bc_274054_t_121322316.txt

# Data z knihovního systému

## Z\_0\_0 z bibliografických záznamů – ISBN, ISSN

The screenshot shows the phpMyAdmin interface for a database named 'Knihomol'. The current table is 'isbn\_aleph'. The interface includes a navigation sidebar on the left, a top menu with options like 'Projit', 'Struktura', 'SQL', 'Vyhledávání', 'Vložit', 'Export', 'Import', and 'Úpravy', and a main content area. A warning message at the top states: 'Aktuální výběr neobsahuje unikátní klíč. Editování v mřížce, zaškrťovací políčka nebo odkazy na editaci a mazání proto nejsou k dispozici.' Below this, a green bar indicates 'Zobrazeny záznamy 0 - 24 (433461 celkem, Dotaz trval 0,0217 sekund.)'. The SQL query shown is 'SELECT \* FROM `isbn\_aleph`'. The table view shows columns 'isbn', 'bibno', and 'bibref'. The first few rows are:

isbn	bibno	bibref
000163222	241514	BLYTON, Enid. <i>Mystery Stories : the secret of c...
000184202	229087	GARNER, Alan. <i>Elidor.</i> London: Collins, 1965...
000184289	254120	GARNER, Alan. <i>The stone book quartet.</i> Londo...
000195655	6466992	ALLEN JENSEN, Virginia. EDMAN, Polly. <i>Red threa...
000196522	13407	FARNDON, John. <i>Children&apos;s encyclopedia.</i>...
000197170	456571	KERMAN, Joseph. <i>Musicology.</i> London: Fontana...
000197172	177978	HOSKING, Geoffrey A. <i>A history of the Soviet un...
000200020	196612	HILL, Lawrence. <i>Black berry, sweet juice : on b...
000200501	196301	WRIGHT, Richard Bruce. <i>Clara Callan : a novel.</i>...
000200510	471058	GOODWEATHER, Hartley. <i>DreadfulWater shows up.</i>...
000200557	6347193	HOLLINGSHEAD, Greg. <i>Bedlam : a novel.</i> First...
000200568	398726	BEZMOZGIS, David. <i>Natasha and other stories.</i>...
000200586	399969	WRIGHT, Richard Bruce. <i>Adultery : a novel.</i> ...
000200616	557846	KING, Thomas. <i>A short history of Indians in Can...
000200674	480518	MCGOOGAN, Kenneth. <i>Lady Franklin&apos;s revenge...
000200722	20372	SYKES, Christopher. <i>Evelyn Waugh : a biography...



# Data z EDS

## nefunguje

- seznam dostupných titulů (časopisy, knihy apod.) ve Full Text Finderu
  - bez DOI, ale má ISSN a ISBN
- nemáme přístup k centrálnímu indexu (dump)
- dotaz musí vrátit přesnou shodu ne částečnou (např. 550 jiných výsledků)
- API
  - zkusíme předtím něco jiného
- placený produkt LinkIQ od EBSCO

## funguje (asi)

- Full Text Finderu pro ISBN a ISSN
- OpenURL přímo Full Text Finderu
  - ☹ HTTPSPool(host='resolver.ebscohost.com', port=443): Max retries exceeded with url:.... >: Failed to establish a new connection: [Errno 11002] getaddrinfo failed'))

# complete-holdings-MU-20220401.csv

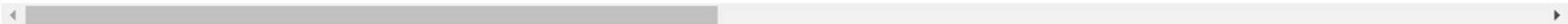
0	KBID	1445264	non-null	int64	21	CoverageStatement	35	non-null	object
1	Title	1445263	non-null	object	22	Embargo	19916	non-null	object
2	AlternateTitle	1445264	non-null	object	23	CustomEmbargo	56	non-null	object
3	PackageName	1445264	non-null	object	24	Description	2	non-null	object
4	URL	1440778	non-null	object	25	Subject	1153158	non-null	object
5	ProxiedURL	1445264	non-null	object	26	ResourceType	1445264	non-null	object
6	Publisher	1357824	non-null	object	27	PackageContentType	1445264	non-null	object
7	Edition	377830	non-null	object	28	CreateCustom	1445264	non-null	object
8	Author	901045	non-null	object	29	HideOnPublicationFinder	1445264	non-null	object
9	Editor	133986	non-null	object	30	Delete	1445264	non-null	object
10	Illustrator	2448	non-null	object	31	OrderedThroughEBSCO	1445264	non-null	object
11	PrintISSN	157561	non-null	object	32	IsCustom	1445264	non-null	object
12	OnlineISSN	137561	non-null	object	33	UserDefinedField1	0	non-null	float64
13	PrintISBN	913976	non-null	object	34	UserDefinedField2	0	non-null	float64
14	OnlineISBN	929932	non-null	object	35	UserDefinedField3	0	non-null	float64
15	DOI	0	non-null	float64	36	UserDefinedField4	0	non-null	float64
16	PeerReviewed	1445264	non-null	object	37	UserDefinedField5	0	non-null	float64
17	ManagedCoverageBegin	1314916	non-null	object	38	PackageType	1445264	non-null	object
18	ManagedCoverageEnd	1314916	non-null	object	39	AllowEBSCOtoSelectNewTitles	1445264	non-null	object
19	CustomCoverageBegin	22930	non-null	object	40	PackageID	1445264	non-null	int64
20	CustomCoverageEnd	22930	non-null	object	41	VendorName	1445264	non-null	object
21	CoverageStatement	35	non-null	object	42	VendorID	1445264	non-null	int64
22	Embargo	19916	non-null	object					

# complete-holdings-MU-20220401.csv

Out[3]:

	KBID	Title	AlternateTitle	PackageName	URL	ProxiedURL	Publisher	Edition	Ar
0	2837	AJR, American Journal of Roentgenology	AJR American Journal of Roentgenology AJR AJR ...	Publisher's Site	<a href="https://www.ajronline.org/loi/ajr">https://www.ajronline.org/loi/ajr</a>	<a href="http://ezproxy.muni.cz/login?url=https://www.a...">http://ezproxy.muni.cz/login?url=https://www.a...</a>	American Roentgen Ray Society	NaN	
1	3865	American Journal of Public Health	American Journal of Public Health JPH Am J Pub...	Publisher's Site	<a href="https://ajph.aphapublications.org/loi/ajph">https://ajph.aphapublications.org/loi/ajph</a>	<a href="http://ezproxy.muni.cz/login?url=https://ajph...">http://ezproxy.muni.cz/login?url=https://ajph...</a>	American Public Health Association	NaN	
2	25262	Education	Education	Publisher's Site	<a href="https://www.ingentaconnect.com/content/prin/ed">https://www.ingentaconnect.com/content/prin/ed</a>	<a href="http://ezproxy.muni.cz/login?url=https://www.i...">http://ezproxy.muni.cz/login?url=https://www.i...</a>	Project Innovation of Mobile	NaN	
3	26723	Environmental Values	Environmental Values 环境价值 Environ Values	Publisher's Site	<a href="https://www.ingentaconnect.com/content/whp/ev%...">https://www.ingentaconnect.com/content/whp/ev%...</a>	<a href="http://ezproxy.muni.cz/login?url=https://www.i...">http://ezproxy.muni.cz/login?url=https://www.i...</a>	White Horse Press	NaN	
4	39260	The Journal of Biological Chemistry	Journal of Biological Chemistry JBC J Biol Che...	Publisher's Site	<a href="https://www.jbc.org/issues">https://www.jbc.org/issues</a>	<a href="http://ezproxy.muni.cz/login?url=https://www.j...">http://ezproxy.muni.cz/login?url=https://www.j...</a>	Elsevier	NaN	

5 rows × 43 columns



# 80/20

- diplomky
  - špatně zadané identifikátory v textech
  - chyba při extrahování z textů
  - různé znakové sady
  - různé formáty (př. ISBN)
  
- data z knihovny
  - více identifikátorů v jednom poli
  - různé formáty (př. ISBN)
  - různé formáty souborů na vstupu

# ISBN

diplomky

• 19 731 → 47 688 → 28 044

aleph

• 433 461 → 416 201

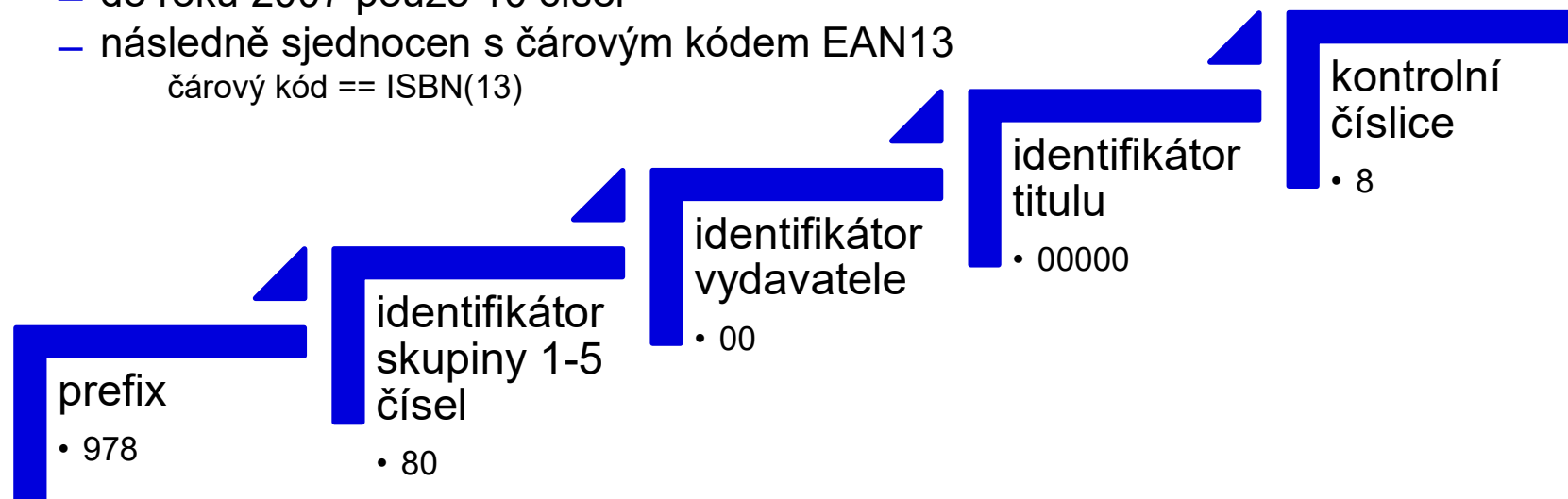
discovery

• 1 694 058 → 1 677 374

# ISBN

– ISBN – 13 čísel - 978-80-00-00000-8 == EAN-13

- do roku 2007 pouze 10 čísel
- následně sjednocen s čárovým kódem EAN13  
čárový kód == ISBN(13)

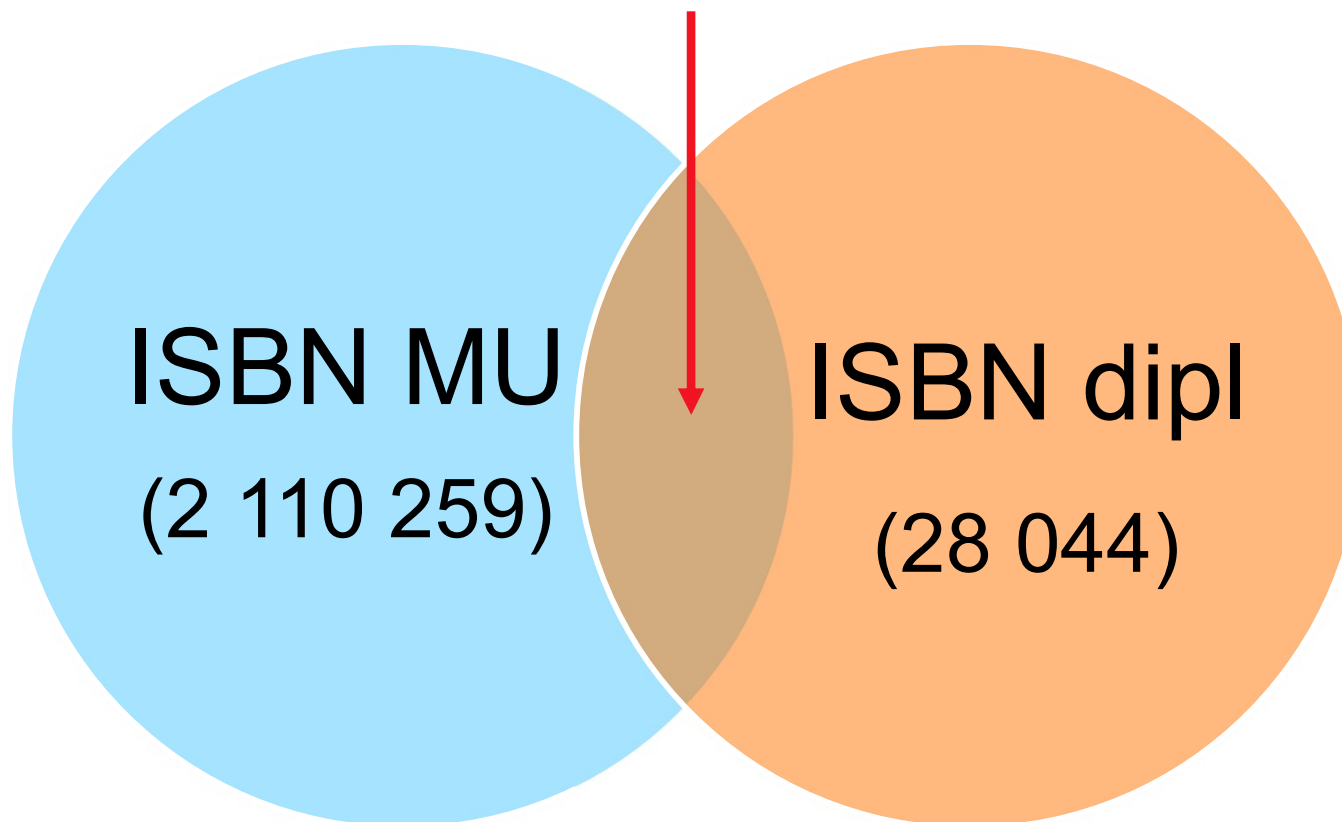


- odděláváme prefix 978/979 (knihy/hudebniny)
- odděláváme poslední kontrolní znak
  - u ISBN(10) a ISBN(13) se počítá jiným způsobem

<https://text.nkp.cz/sluzby/sluzby-pro/isbn-ismn-issn/isbn/isbn-5>  
[International Standard Book Number – Wikipedie \(wikipedia.org\)](https://en.wikipedia.org/wiki/International_Standard_Book_Number)

# ISBN výstup

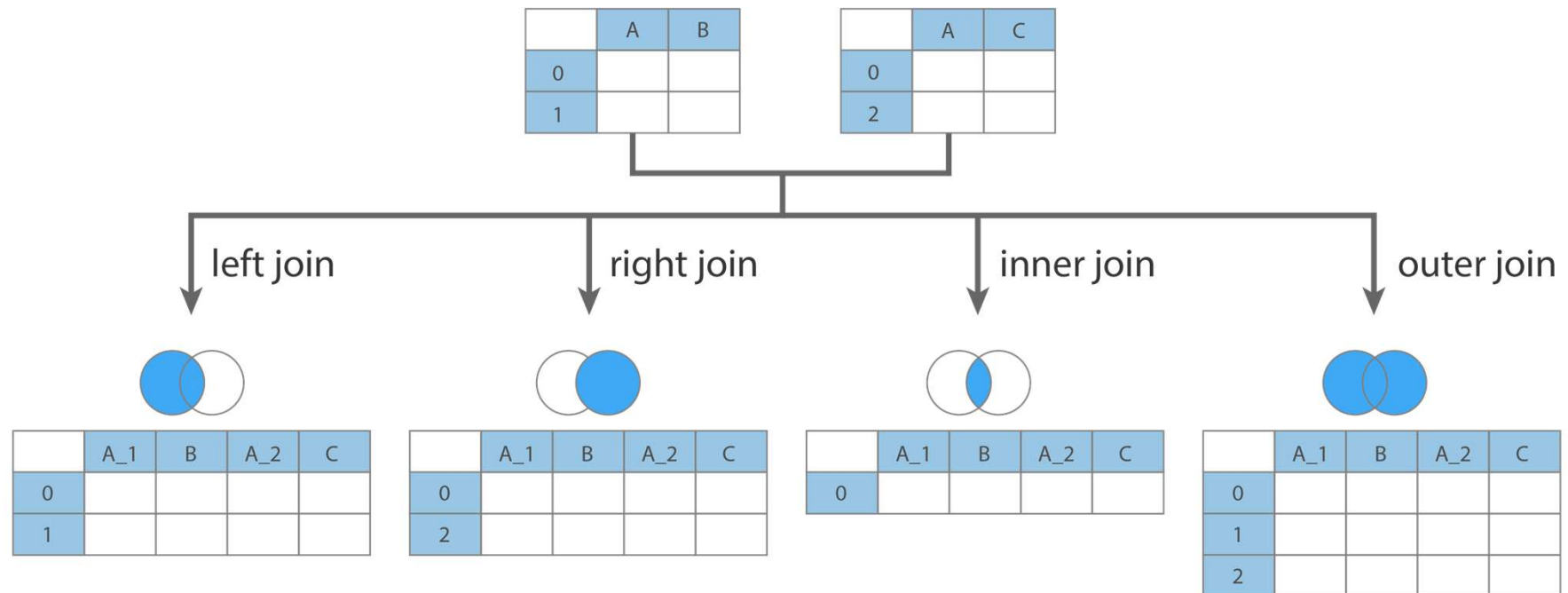
**24 033 shod**



**20 601 aleph | 3 432 discovery**

# spojování dat

merge





# výstup v jupyter notebook



Out[44]:

	isbn	file name	norma	ID	Title	ISBN_orig	PackageName	Publisher	zdroj
0	978-0520055278	bc_287401_t_92462372.txt	052005527	6367994	ALLSEN, Thomas T. <i>Mongol imperialism : the ...	NaN	NaN	NaN	aleph
1	978-1885445131	bc_287401_t_92462372.txt	188544513	NaN	NaN	NaN	NaN	NaN	NaN
2	978-0806137346	bc_287401_t_92462372.txt	080613734	NaN	NaN	NaN	NaN	NaN	NaN
3	978-0520265851	bc_287401_t_92462372.txt	052026585	NaN	NaN	NaN	NaN	NaN	NaN
4	978-1473892682	bc_287401_t_92462372.txt	147389268	NaN	NaN	NaN	NaN	NaN	NaN

```
In [31]: #takže jsme zjistili že máme 24 237 citací zdrojů, které máme, nedohledali  
#jsme 3 432 záznamů  
ISBN_result.loc[ISBN_result["zdroj"] == "discovery"]
```

Out[31]:

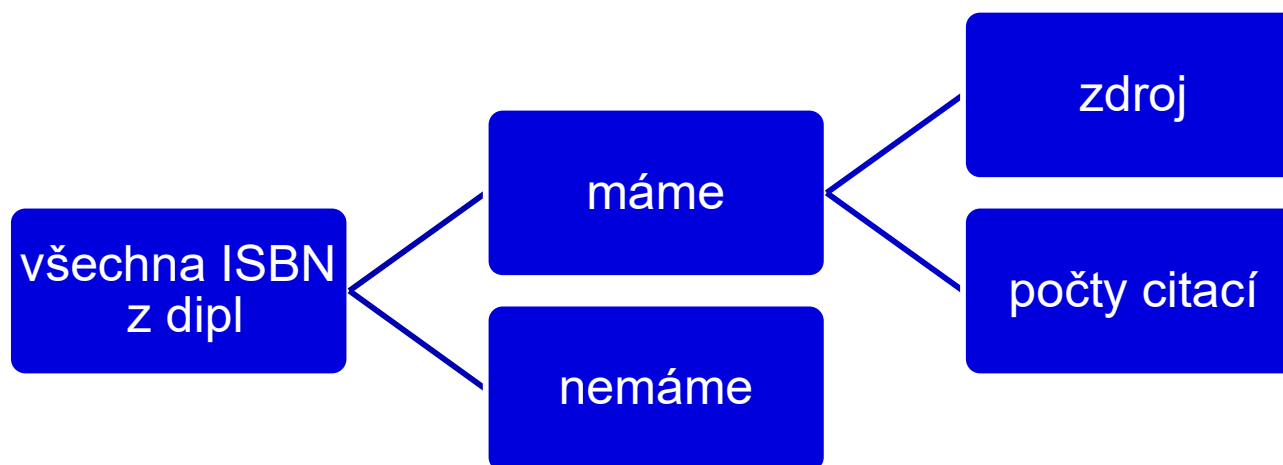
isbn	file name	norma	KBID	Title	ISBN	PackageName	Publisher	ID	zdroj
------	-----------	-------	------	-------	------	-------------	-----------	----	-------

```
In [45]: pokus = ISBN_result.groupby(["zdroj"]).count()  
pokus.head()
```

Out[45]:

	isbn	file name	norma	ID	Title	ISBN_orig	PackageName	Publisher
zdroj								
aleph	20601	20601	20601	20601	20601	0	0	0
discovery	3432	3432	3432	3432	3432	3432	3432	3432


# výstup pro knihovníky



norma	Title	PackageName	Publisher	zdroj	citovanost
2217	807367313 Kvalitativní výzkum v pedagogických vědách	EBSCO eBooks	Portál	discovery	20
1981	802461966 Jak se vyrábí sociologická znalost	eBook Academic Collection (EBSCOhost) - Worldwide	Charles University in Prague	discovery	12
1980	802461966 Jak se vyrábí sociologická znalost	EBSCO eBooks	Charles University in Prague	discovery	12
1989	802462153 Vývojová psychologie. Dětství a dospívání	eBook Academic Collection (EBSCOhost) - Worldwide	Charles University in Prague	discovery	11
1881	802107871 Po divné krajině: Charakteristika a vnitřní čl...	EBSCO eBooks	Masarykova univerzita	discovery	11
1988	802462153 Vývojová psychologie. Dětství a dospívání	EBSCO eBooks	Charles University in Prague	discovery	11
1882	802107871 Po divné krajině: Charakteristika a vnitřní čl...	eBook Academic Collection (EBSCOhost) - Worldwide	Masarykova univerzita	discovery	11
2221	807367698 Dějiny českých médií 20. století	EBSCO eBooks	Portál	discovery	10
1906	802108441 Na rozhraní světů: Fantastická literatura v me...	EBSCO eBooks	Masarykova univerzita	discovery	10
1907	802108441 Na rozhraní světů: Fantastická literatura v me...	eBook Academic Collection (EBSCOhost) - Worldwide	Masarykova univerzita	discovery	10

norma	isbn	Title	citovanost
2749	156159263	9781561592630	48
5884	807106963	978-80-7106-963-8	21
5226	802444017	978-80-244-4017-0	20
7912	809054801	978-80-905480-1-5	12
5666	807008354	978-80-7008-354-3	11
1719	089042559	9780890425596	11
8512	923001221	978-92-3-001221-2	11
7592	808812311	978-80-88123-11-8	10
5777	807050690	978-80-7050-690-5	9
5665	807008332	978-80-7008-332-1	7

- Overview
- Description
- Reviews and Awards
- From Our Blog



**The ultimate authority on all aspects of music worldwide.**

## Grove Music Online


[www.oxfordmusiconline.com](http://www.oxfordmusiconline.com)

The definitive source for music scholarship

Editor in Chief: **Dr. Deane Root** and

- Features more than 50,000 articles written by over 6,000 leading international experts
- Quickly learn complex terms and content with *The Oxford Companion to Music* and *The Oxford Dictionary of Music*
- Follow links to audio and video recordings as well as score facsimiles, thanks to partnerships with Alexander Street Press and Naxos Music Library
- Enhance research and classroom teaching with learning resources and timelines
- Regular updates keep you up to date on the most current thinking by top scholars

**Institutions**

 How to Subscribe

Contact a Sales Rep

---

**Personal Subscriptions**

Choose a subscription period

**£215.00**

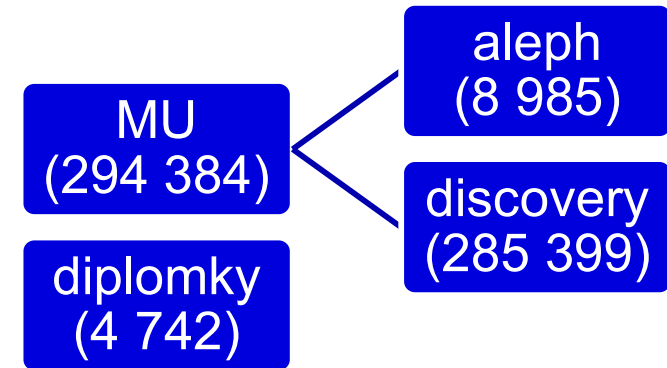
I have a promotion code for this product

Where applicable, tax will be added to the above price prior to payment.

[About Personal Subscriptions](#)

# ISSN

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 8017 entries, 6448 to 8016
Data columns (total 7 columns):
#   Column          Non-Null Count  Dtype
---  -
0   norma           8017 non-null   object
1   ID              8017 non-null   object
2   Title           8017 non-null   object
3   Publisher       8017 non-null   object
4   PackageName    8017 non-null   object
5   zdroj          8017 non-null   object
6   citovanost     8017 non-null   int64
dtypes: int64(1), object(6)
memory usage: 501.1+ KB
```



	norma	ID	Title	Publisher	PackageName	zdroj	citovanost
6448	18044255	19017365	Duha	Moravska Zemska Library	Library & Information Science Source	discovery	47
4319	12125075	21968757	Ikaros: Elektronick½ Casopis o Informacn½ Sp...	Ikaros O.s.	EBSCO Open Access Journals	discovery	33
4286	12119938	137836	<i>Host : nezavisly obcasnik pro filozofii, li...			aleph	28
364	00090468	278624	Česká literatura	ASCR Institute of Czech Literature	JSTOR Arts & Sciences XI Archive Collection	discovery	27
363	00090468	278624	Česká literatura	ASCR Institute of Czech Literature	Central & Eastern European Academic Source	discovery	27
362	00090468	278624	Česká literatura	ASCR Institute of Czech Literature	CEEOL (Central and Eastern European Online Lib...	discovery	27
361	00090468	134977	<i>Česká literatura : časopis pro literární vě...			aleph	27
6440	18042406	675158	ProInflow	Masaryk University	EBSCO Open Access Journals	discovery	26
6441	18042406	675158	ProInflow	Masaryk University	Library & Information Science Source	discovery	26
842	00220418	40058	Journal of documentation	Emerald Publishing Limited	Arts & Humanities Database	discovery	24

# DOI z diplomek

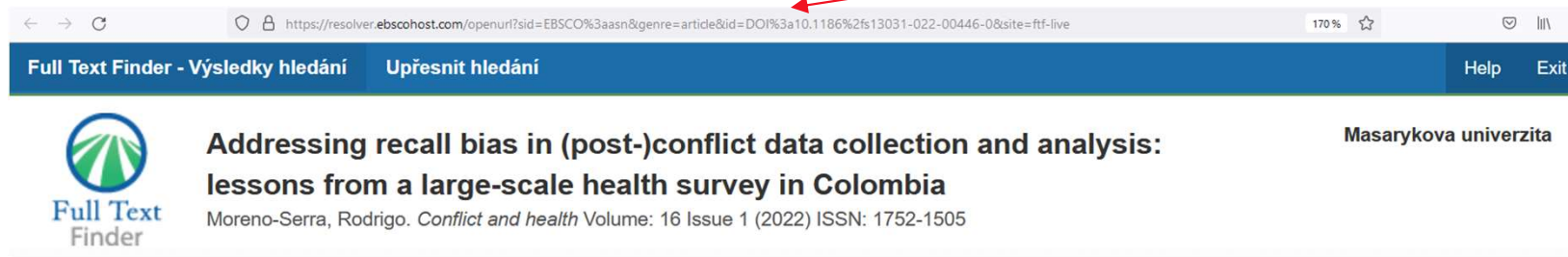
deduplikovaná na id a file name

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 10403 entries, 552 to 278634
Data columns (total 5 columns):
#   Column      Non-Null Count  Dtype
---  -
0   type        10403 non-null  object
1   id          10403 non-null  object
2   file name   10403 non-null  object
3   nazev       10403 non-null  object
4   citace      10403 non-null  object
dtypes: object(5)
memory usage: 487.6+ KB
```


	type	id	file name	nazev	citace
552	doi	10.1016/S0140-6736(15)60755-3	bc_312582_t_82145618.txt		
557	doi	10.1016/S0140-	bc_312582_t_82145618.txt		
1523	doi	10.2147/NDT.S	bc_314435_t_90392361.txt		
1525	doi	10.1037/0033-2909.121.1.65	bc_314435_t_90392361.txt		
1528	doi	10.1097/00004583-199603000-00016	bc_314435_t_90392361.txt		

# Kde to tedy bereme?

<https://resolver.ebscohost.com/openurl?sid=EBSCO%3aasn&genre=article&id=DOI%3a10.1186%2fs13031-022-00446-0&site=ftf-live>



Full Text Finder - Výsledky hledání    Upřesnit hledání    Help    Exit

 **Addressing recall bias in (post-)conflict data collection and analysis:  
lessons from a large-scale health survey in Colombia**    Masarykova univerzita

Moreno-Serra, Rodrigo. *Conflict and health* Volume: 16 Issue 1 (2022) ISSN: 1752-1505

## Full Text Finder - Výsledky hledání

### Nalezené zdroje (plný text)

-  [Check for full text at the publisher's site.](#) 2007-01-01 - present
-  [Check for full text at the publisher's site.](#) 2007-01-01 - present
-  [Plný text dostupný v BioMed Central](#) 2007-01-01 - present
-  [Plný text dostupný v Health & Medical Collection](#) 2009-01-01 - present
-  [Plný text dostupný v ProQuest Central](#) 2009-01-01 - present
-  [Plný text dostupný v PubMed Central](#) 2007-01-01 - present
-  [Plný text dostupný v Springer Verlag](#) 2007-12-01 - present
-  [Plný text dostupný na stránkách vydavatele.](#) 2007-01-01 - present
-  [Plný text dostupný na stránkách vydavatele.](#) 2007-01-01 - present
-  [Plný text dostupný na stránkách vydavatele.](#) 2007-12-01 - present
-  [Procházet tento časopis v Directory of Open Access Journals](#) 2007-01-01 - present

```
requests.get("https://resolver.ebscohost.com/openurl?sid=EBSCO%3aasn&genre=article&site=ftf-live&id=DOI%3a"+doi)
```

# DOI výstup

C	D	E	
id	file name	nazev	
10.1016/S0140-67	bc_312582_t_82145618.txt	What is a resilient health system? Lessons from Ebola	Kruk, Margaret. <em>The Lancet</em> Volume: 385 Issue 9980 (2015) ISSN: 014
10.1016/S0140-	bc_312582_t_82145618.txt	[nenalezeno]	[nenalezeno]
10.2147/NDT.S	bc_314435_t_90392361.txt	[nenalezeno]	[nenalezeno]
10.1037/0033-290	bc_314435_t_90392361.txt	Behavioral inhibition, sustained attention, and executive functions: C	Barkley, Russell. <em>Psychological bulletin</em> Volume: 121 Issue 1 (1997) I
10.1097/00004583	bc_314435_t_90392361.txt	Predictors of Persistence and Remission of ADHD into Adolescence: F	BIEDERMAN, JOSEPH. <em>Journal of the American Academy of Child & Adoles
10.3758/BF032093	bc_314435_t_90392361.txt	Prospective and retrospective duration judgments: A meta-analytic re	Block, Richard. <em>Psychonomic Bulletin & Review</em> Volume: 4 Issue 2 (:
10.3758/APP.72.3	bc_314435_t_90392361.txt	Timing and time perception: A review of recent behavioral and neuro	Grondin, Simon. <em>Attention, Perception & Psychophysics</em> Volume: 7
10.1080/09297045	bc_314435_t_90392361.txt	Retrospective and Prospective Time Deficits in Childhood ADHD: The	Hurks, Petra. <em>Child Neuropsychology : A journal on normal and abnormal
10.1177/00332941	bc_314435_t_90392361.txt	Exploring the Effects of Working Memory on Time Perception in Atten	Lee, Hom-Yi. <em>Psychological reports</em> Volume: 122 Issue 1 (2019) ISSN
10.1097/00004583	bc_314435_t_90392361.txt	Variability in Time Reproduction: Difference in ADHD Combined and	Mullins, Celine. <em>Journal of the American Academy of Child & Adolescent I
10.1016/S0010-44	bc_314435_t_90392361.txt	Attention deficit hyperactivity disorder adults: Comorbidities and ad	Murphy, Kevin. <em>Comprehensive Psychiatry</em> Volume: 37 Issue 6 (199
10.1080/09297040	bc_314435_t_90392361.txt	Time Perception in Children with ADHD: The Effects of Task Modality	Plummer, Carol. <em>Child Neuropsychology : A journal on normal and abnorm
10.1177/10870547	bc_314435_t_90392361.txt	Time Estimation Abilities of College Students With ADHD	Prevatt, Frances. <em>Journal of Attention Disorders: A Journal of Theoretical
10.1037/14043-00	bc_314435_t_90392361.txt	Role of dopamine in the pathophysiology of attention-deficit/hypera	Vaidya, Chandan. <em>Cognition and brain development: Converging evidenc
10.1080/17470210	bc_314435_t_90392361.txt	Subjective shortening with filled and unfilled auditory and visual inte	Wearden, J.. <em>The Quarterly Journal of Experimental Psychology</em> Vol

# DOI

tam a zase zpátky

```
In [21]: DOI_all = analiza_doi.groupby(["id", "nazev", "citace"])['file name'].count()
DOI_all = DOI_all.to_frame()
DOI_all = DOI_all.reset_index()
DOI_all.rename(columns={"file name": "citovanost"}, inplace=True)
DOI_all = DOI_all.sort_values('citovanost', ascending=False)
DOI_all.head(20)
```

Out[21]:

	id	nazev	citace	citovanost
7236	10.5817/P	[nenalezeno]	[nenalezeno]	53
1445	10.1037/0022-	[nenalezeno]	[nenalezeno]	36
119	10.1007/978-	[nenalezeno]	[nenalezeno]	31
1203	10.1023/A:	[nenalezeno]	[nenalezeno]	29
1699	10.1037/0033-295	[nenalezeno]	[nenalezeno]	16
3823	10.1097/01	Sleep Hygiene and Melatonin Treatment for Chil...	WEISS, MARGARET. <em>Journal of the American A...	15
1646	10.1037/0033-	[nenalezeno]	[nenalezeno]	14
5902	10.1300/J082	[nenalezeno]	[nenalezeno]	11
1344	10.1037/0003-	[nenalezeno]	[nenalezeno]	11
1379	10.1037/0012-	[nenalezeno]	[nenalezeno]	9
1907	10.1056/NEJM	[nenalezeno]	[nenalezeno]	8
1399	10.1037/0021-	[nenalezeno]	[nenalezeno]	8
5807	10.1191/1478088706	[nenalezeno]	[nenalezeno]	8
1365	10.1037/0003-066X.55.5.469	Emerging adulthood: A theory of development fr...	Arnett, Jeffrey. <em>American Psychologist</em>...	7
7258	10.5817/SP2016-	[nenalezeno]	[nenalezeno]	7
156	10.1007/978-3-319-	[nenalezeno]	[nenalezeno]	7
7270	10.5817/SP2019-	[nenalezeno]	[nenalezeno]	7



# co dál

u doi vydolovat nakladatele



## Addressing recall bias in (post-)conflict data collection and analysis: lessons from a large-scale health survey in Colombia

Moreno-Serra, Rodrigo. *Conflict and health* Volume: 16 Issue 1 (2022) ISSN: 1752-1505

### Full Text Finder - Výsledky hledání

#### Nalezené zdroje (plný text)

- [Check for full text at the publisher's site.](#) 2007-01-01 - present
  - [Check for full text at the publisher's site.](#) 2007-01-01 - present
  - [Plný text dostupný v BioMed Central](#) 2007-01-01 - present
  - [Plný text dostupný v Health & Medical Collection](#) 2009-01-01 - present
  - [Plný text dostupný v ProQuest Central](#) 2009-01-01 - present
  - [Plný text dostupný v PubMed Central](#) 2007-01-01 - present
  - [Plný text dostupný v Springer Verlag](#) 2007-12-01 - present
  - [Plný text dostupný na stránkách vydavatele.](#) 2007-01-01 - present
  - [Plný text dostupný na stránkách vydavatele.](#) 2007-01-01 - present
  - [Plný text dostupný na stránkách vydavatele.](#) 2007-12-01 - present
  - [Procházet tento časopis v Directory of Open Access Journals](#) 2007-01-01 - present
- OPEN ACCESS

# URL

tady to taky zavání čištění

Out[42]:

	type	id	file name
17215	url	https://doi.org	9897
40437	url	https://www.youtube.com	3274
17206	url	https://doi.Org	3010
19862	url	https://is.muni.cz	2936
17724	url	https://elpais.com	1903
7928	url	http://www.digitalniknihovna.ez	1359
37255	url	https://www.researchgate.net	1113
7010	url	http://www.beazley.ox.ac.uk	1110
31129	url	https://www.facebook.com	1001
33262	url	https://www.jstor.org	978
29897	url	https://www.czso.cz	964
16228	url	https://cs.wikipedia.org	942
29868	url	https://www.czechency.org	906
17096	url	https://dnnt.mzk.ez	841
32571	url	https://www.idnes.cz	804
40496	url	https://www.zakonyprolidi.cz	727
28790	url	https://www.britannica.com	671
25874	url	https://theses.cz	638
28399	url	https://www.beazley.ox.ac.uk	581
34815	url	https://www.mkcr.cz	560

# Co dál

- návrat ke kořenům – extrakce identifikátorů
- vyčistit url, doi
- předat data na akvizici a info služby
- vymyslet jak se dostat k plným textům akademiků
- vyčistit a zjednodušit kód
- ...

MUNI  
ARTS

**Děkuji za pozornost.**

[mbockova@phil.muni.cz](mailto:mbockova@phil.muni.cz)

<https://github.com/LibraryFFMU/text-analyzy>