

Kramerius a AI

Martin Lhoták, Knihovna AV ČR

Petr Žabička, Moravská zemská knihovna v Brně

<https://bit.ly/akmkramerius2023>

Archivy, knihovny a muzea v digitálním světě, Praha 29.11.2023

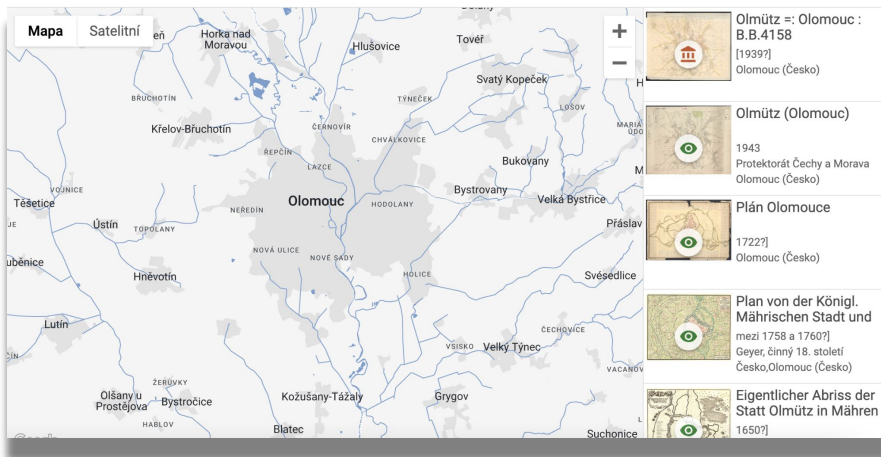
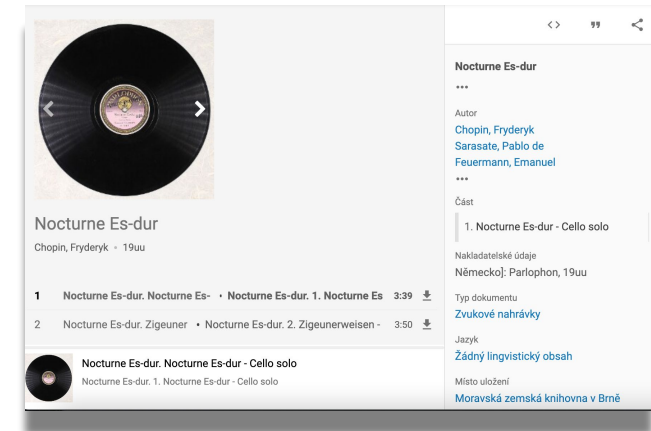
Obsah

- Novinky v Krameriu a ČDK
- AI projekty pro rozvoj digitálních knihoven
- Co bude dál?

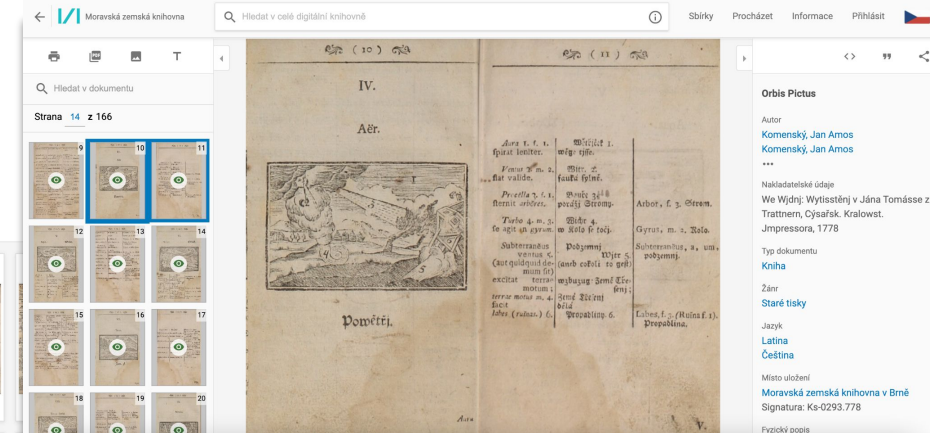
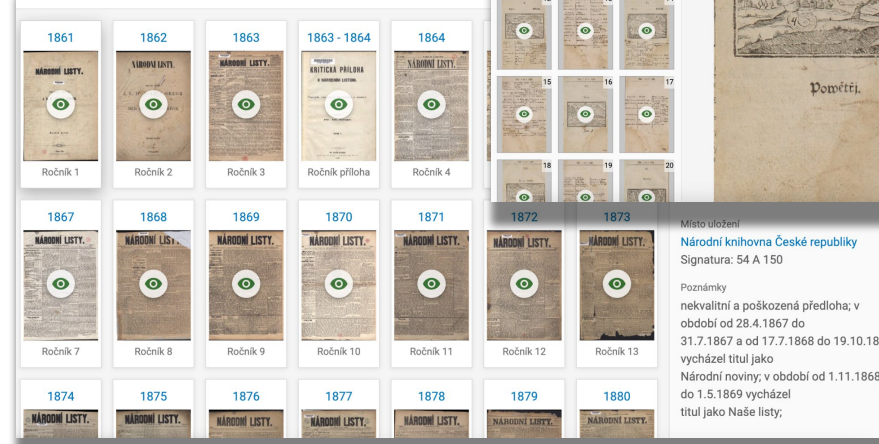


Kramerius

- Open source, volně dostupný, cca 50 instalací
- jpg, jpeg2000, pdf, **epub**, wav
- MODS, ALTO, METS,...
- Přístupová práva (přihlášení, DNNT, speciální požadavky)
- Virtuální sbírky jako dokumenty, uživatelské kolekce/seznamy
- Periodika, konvoluty, vícesvazková díla, mapová díla
- Geo vyhledávání, zobrazení georeferencovaných map
- plně dostupné přes API



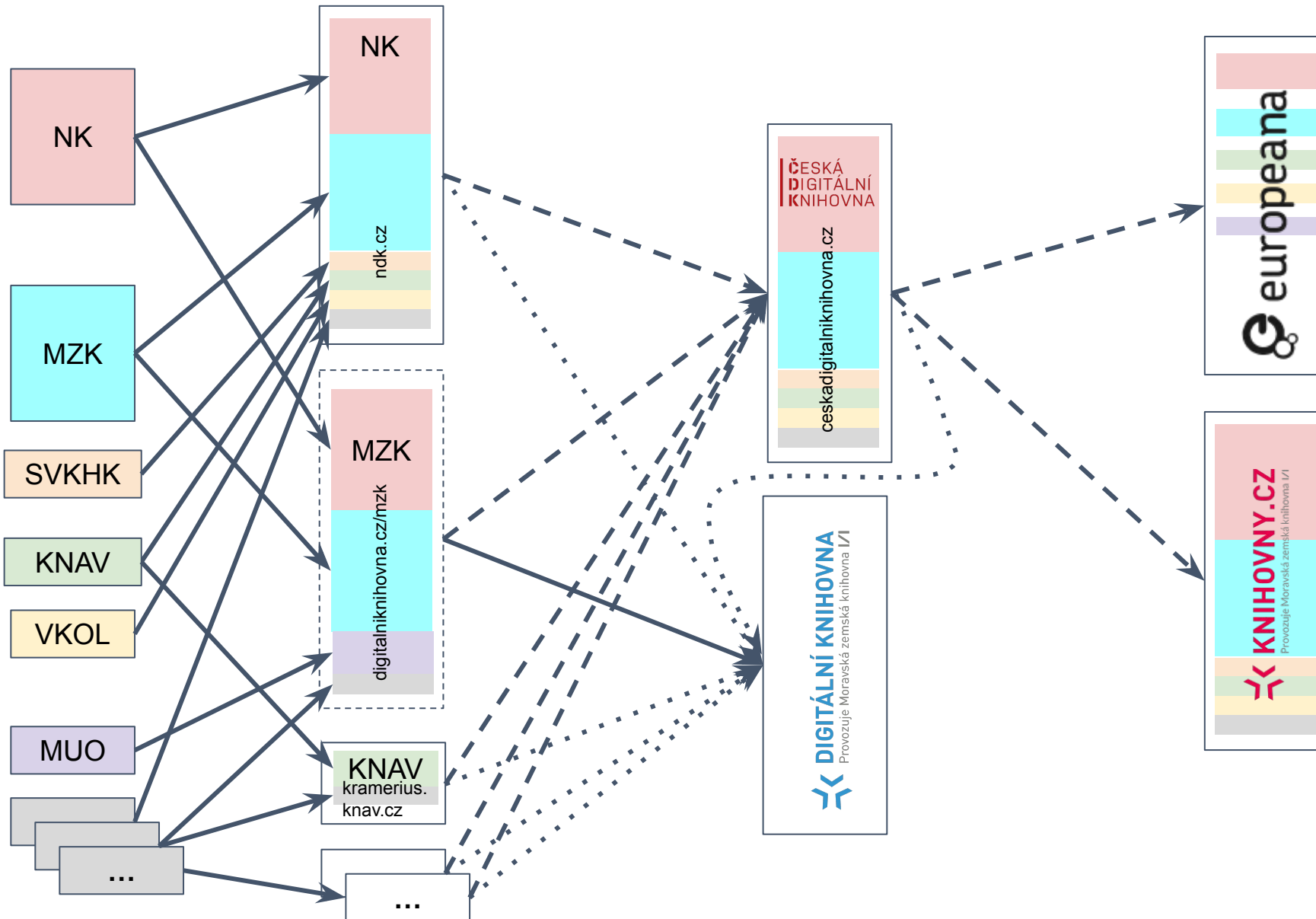
Národní listy



Místo uložení
Národní knihovna České republiky
Signatura: 54 A 150

Poznámky
nevalnitni a poškozéná předloha; v období od 28.4.1867 do 31.7.1867 a od 17.7.1868 do 19.10.1868 vycházel titul jako Národní noviny; v období od 1.11.1868 do 1.5.1869 vycházel titul jako Naše listy;

Digitální knihovny v ČR











Česká digitální knihovna
























Plně provozována v MZK (K7) - [CeskaDigitalniKnihovna.cz](https://ceskadigitalniknihovna.cz)

Plánuje se zrcadlo v NK

Jsou připraveny nové smlouvy se zapojenými knihovnami

Name	Code	Kramerius - API	Version	Kramerius - client	Version	Collections	Documents	Pages ▼
 Czech Digital Library	cdk	 https://api.ceskadigitalniknihovna.cz	7.0.23_...	https://ceskadigitalniknihovna.cz	2.5.0	517	312 581	137 727 417
 National Library of the Czech Republic	nkp	 https://www.ndk.cz	5.8.4.3	https://www.ndk.cz	2.4.9	50	251 774	73 423 126
 Moravian Library in Brno	mzk	 https://api.kramerius.mzk.cz	7.0.29	https://www.digitalniknihovna.cz/mz...	2.6.0	544	279 873	68 662 982

K7

Name	Code		Kramerius - API	Version
 Research Library of South Bohemia in České B...	cbvk		https://kramerius.cbvk.cz	7.0.30-d.
 EoD Open project digital library	eod		https://ds-valve.uibk.ac.at/back	7.0.29
 Moravian Library in Brno	mzk		https://api.kramerius.mzk.cz	7.0.29
 MZK - test installation of K7	k7test		https://api.kramerius.mzk.cz	7.0.29
 Czech Radio	cro		https://kramerius.rozhlas.cz	7.0.27
 František Bartoš Regional Library in Zlín	kfbz		https://kramerius.kfbz.cz	7.0.27
 Slovak Digital Library Catalogue	dikda		https://dikda.snk.sk	7.0.27
 Czech Digital Library	cdk		https://api.ceskadigitalniknihovna.cz	7.0.23_...
 Education and Research Library of Pilsener Re...	svkpk		https://kramerius.svkpk.cz	7.0.23
 Slovak centre of scientific and technical inform...	cvti		https://ekniznice.cvtisr.sk	7.0.23
 Slovak Digital Library Catalogue	snk		https://onk.snk.sk	7.0.23

← Moravská zemská knihovna

Výsledky: 61 - 120 / 287781 ▼ Pokročilé hledání

Dostupnost ^

Pouze veřejné	58607
Pouze neveřejné	229172
Vše	287779

Typ dokumentu ^


Knihy	243420
Noviny a časopisy	4289
Sbírky	88
Grafiky	19045
Mapy	14025
Hudebniny	5162
Zvukové nahrávky	683
Archiválie	548
Rukopisy	221
Konvoluty	139


Rok vydání ^


Od 0 Do 2023


Klíčové slovo ^

Mapové sbírky	12320
Melova mapová sbírka	12316

Kontrolní závěr: Miloslav Kala
2013  Kniha

Processionale abbatiae s. Erhardi, (fragmentum)
[mezi 1500 a 1550?]  Kniha

Gebet-, Stunden- und Andachtsbuch
[mezi 1525 a 1550?]  Kniha

Vše pro blaho města
Moberg, Vilhelm, 1898-1973
1941  Kniha

← Moravská zemská knihovna

Výsledky: 61 - 120 / 287781 ▼ Pokročilé hledání

Dostupnost ^

Veřejné	61226
Po přihlášení	129161
V knihovně	98289
Vše dostupné	189535
Vše	287781


Licence ^


Díla nedostupná na trhu - online	129793
Studovna	67337
Volná díla	58701
Díla nedostupná na trhu - studovna	32395
Hudebniny Kroměříž	2784
Smluvně zveřejněná díla	362
Studovna - hudebniny	194


[Zobrazit méně](#)

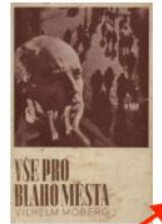
Typ dokumentu ^

Knihy	243420
Noviny a časopisy	4289
Sbírky	88
Grafiky	19045
Mapy	14025
Hudebniny	5162

Kontrolní závěr: Miloslav Kala
2013  V knihovně Kniha

Processionale abbatiae s. Erhardi, (fragmentum)
[mezi 1500 a 1550?]  Veřejné Kniha

Gebet-, Stunden- und Andachtsbuch
[mezi 1525 a 1550?]  Veřejné Kniha

Vše pro blaho města
Moberg, Vilhelm, 1898-1973
1941  Po přihlášení Kniha

Digitalizujeme historii, píšeme budoucnost. Knihy bez hranic. Vzdělání bez bariér.

Hledat ve veřejných dokumentech digitální knihovny

Pouze veřejné

Vstupte

Karel Čapek

Brněnské sochy

Knihy v němčině

Ulice Lidická v 19. století

Mapy ze 16. století

Jan Hus - rukopisy

Tradiční české Vánoce



Vánoce: pro potěšení a poučení všem českým...



Výzdoba vánočního stromku



Eva: časopis moderní ženy



Pohádka o lesním stromku



Cukrovinky a pamlsky na vánoční stromeček...



Letem světem: obrázkový týdeník



Cukrovinky na vánoční stromek

Brněnská nádhera



Železnice



Stará budova čes. Národního divadla v...



[Brno: Stará radnice na Radnické ulici]



[Brno: Zelý trh s odkrytou radniční věží]



[Brno: Jakubská ulice]



[Brno: náměstí Svobody]



Brno - Zelý trh - kašna

Historické kalendáře



Kalendář Pikantního světa na rok



Kalendář české demokracie na rok



Stráž Moravy: velký kalendář Matice...



Kalendář Matice cyrillo-methodějské na rok



Kalendář: Naše posvícení a pěkná...

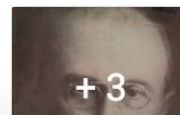
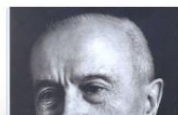


Moravský kalendář národní a obyčejný rok



Kratochvilný slabikář a veselý kalendář obrázkový na rok 1937.

TGM v obrazech



Data v digitálních knihovnách

Obrazová data

- naskenovaný text
 - více sloupců
 - charakter databáze (slovníky, tabulky,...)
 - speciální objekty (vzorci atd.)
- mapy + mapová díla
- grafika
- hudebniny
- kombinované předlohy

Textová data

- OCR (= anotace)
- e-born: pdf, epub, ...

Zvuk, video, 3D objekty, datová média, web

Metadata v digitálních knihovnách

- Deskriptivní (bibliografické záznamy)
 - formát MARC -> MODS
 - až 30 let staré - různá pravidla, konverze, neúplné
 - jazyk záznamu, zkratky
 - nekonzistence uvnitř záznamu, nespolehlivé/nestandardní vzájemné provazby, duplicitní záznamy
 - popis shora vs. zdola; přívazky
 - popis na úrovni titulu vs. exempláře
 - přítomnost / absence věcného popisu
 - Analytické záznamy (články)
 - Autoritní údaje
- Strukturální
- Administrativní (zejména licence, technické údaje)
- Obohacování (např. hodnocení, georeference)

PERO OCR (2018-2022)

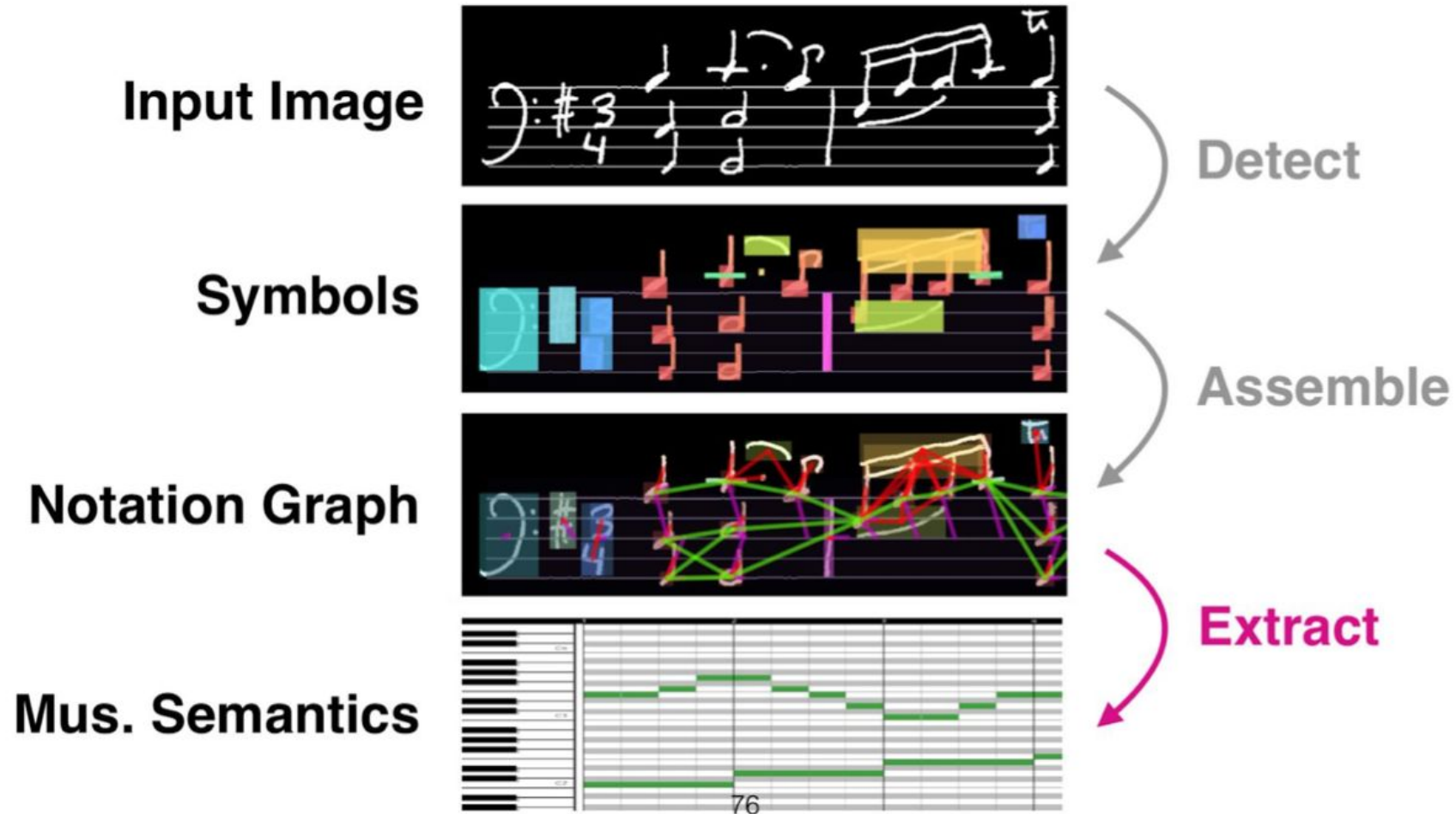
- Projekt **VUT + MZK**
- Cíl: vytvořit OCR nástroj pro převod skenovaných dokumentů (včetně rukopisů)
- Jádro OCR - pero-ocr python balíček <https://github.com/DCGM/pero-ocr>
- Webová aplikace pro kontrolu a opravy - pero_ocr_web
 - Běží na <https://pero-ocr.fit.vutbr.cz>
 - Zdrojové kódy https://github.com/DCGM/pero_ocr_web
- OCR API pro hromadné zpracování
 - <https://pero-ocr.fit.vutbr.cz/api>
- Informace o projektu - <https://pero.fit.vutbr.cz/>

OmniOMR – rozpoznávání hudebního záznamu pomocí strojového učení pro digitální knihovny

Hlavní řešitel **MFF UK**, partner **MZK**

- **Vyhledávání hudební notace,**
- **Vyhledávání v hudební notaci a pomocí hudební notace**
 - **Detekovat části stran** obsahujících hudební notaci a určit její **typ**
 - **Extrahovat** z detekovaných regionů hudebně-sémantické informace (**výšky, délky, pořadí tónů**),
 - **Indexovat výsledky** v databázi a **vyvinout** uživatelsky přívětivé flexibilní **vyhledávání** v notačních datech.

OmniOMR – optical music recognition using machine learning for digital libraries



OmniOMR – rozpoznávání hudebního záznamu pomocí strojového učení pro digitální knihovny

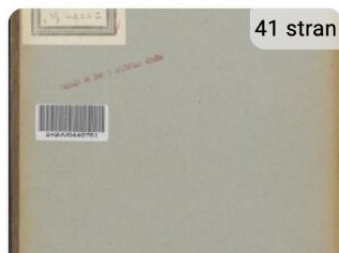
MAKARIUS

Hledej

Skladatelé ▾ Další Autoři ▾ Žánry ▾ Klíčová slova ▾ Zdroje ▾ Jazyky ▾ Rok vydání ▾

Výsledky: 1 - 50 / 404

Seřadit: Podle Relevance



Městská knihovna v Pr...

Duety

Lysenko, Mykola Vitalijovyč

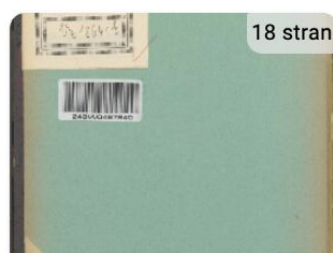
1927



Městská knihovna v Pr...

**Schule für Bandurria
(Spanische Mandoline)**
Leonhardt, Emil

[1894]

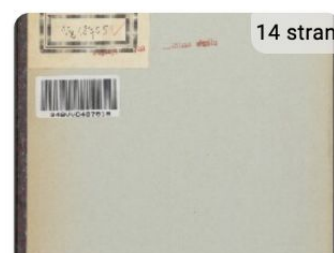


Městská knihovna v Pr...

Křížáci na moři

Bazin, Francois

1906-1928

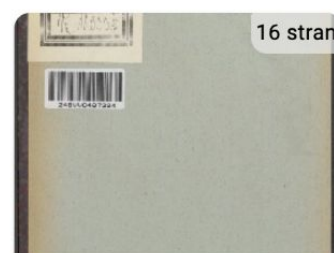


Městská knihovna v Pr...

Šťastné manželství

Bendl, Karel

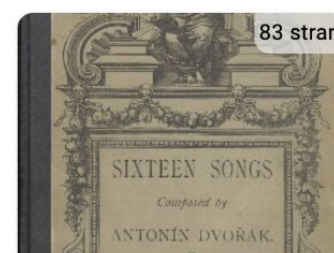
1885



Městská knihovna v Pr...

**Loučení: pro 3 anebo 4
hlasy s průvodem piana (a...)**
Bendl, Karel

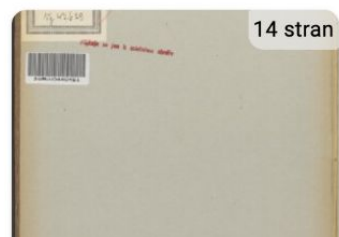
1925



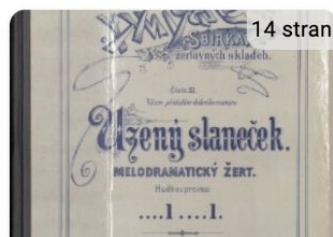
Městská knihovna v Pr...

**Sixteen songs: with
pianoforte accompanimen...**
Dvořák, Antonín

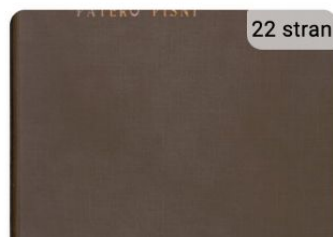
[1913?]



14 stran



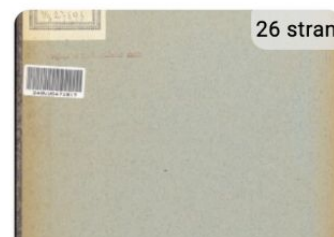
14 stran



22 stran



14 stran



26 stran



25 stran

Orbis Pictus – oživení knihy pro kulturní a kreativní odvětví

Hlavní řešitel **KNAV**, partneři **VUT FIT, NK, MZK**

Otevřít veřejnosti grafický obsah digitálních knihoven

- **identifikovat grafické prvky** obsažené v digitalizovaných dokumentech,
- typově je **kategorizovat**, doplnit o kontextové údaje
- **vyhledávání** grafických prvků
- nástroj pro nalezení různých **vyobrazení stejných osob** a databáze takto nalezených osob, které se podaří identifikovat.

Vyhledávání podle textu

Image Search

tlustý muž kouřící douti

RANDOM

MATCH NON-MATCH FAIL

About the demo

Privacy · Terms

Retrieved images

The interface displays a search for "tlustý muž kouřící douti" (fat man smoking a pipe). The search results are presented in a grid of 24 images. The first row contains six images: a sketch of a man's face with a pipe, a man in a suit sitting and smoking, a man in a suit standing and smoking, a silhouette of a man's head with a pipe, a man in a coat smoking, and a silhouette of a man's head with a pipe. The second row contains seven images: a man in a tuxedo, a man in a suit sitting and smoking, a man in a checkered coat smoking, a silhouette of a man's head with a pipe, a man in a hat smoking, and a hand holding a pipe. The third row contains seven images: a man in a suit smoking, two men in suits, a blurry image of a man's face, a man's face with a pipe, and a man in a hat smoking. The image in the second row, seventh column, which shows a hand holding a pipe, is highlighted with an orange border, indicating it is a match.

Vyhledávání osoby



Image Search



About the demo

[Privacy](#) · [Terms](#)

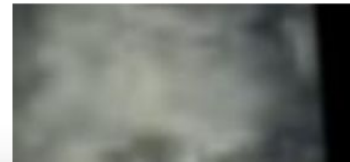
RANDOM

MATCH

NON-MATCH

FAIL

Retrieved images



semANT - Sémantický průzkumník textového kulturního dědictví

Hlavní řešitel **VUT FIT**, partneři **MZK, FSS MU**

Zlepšit možnosti vyhledávání ve fulltextu

- vyhledávat podle významu slov
- vyhledávání podle částí textu, témat
- zlepšit možnosti analýzy textových dat - vývoj témat v čase, vazby mezi tématy apod.

semANT - Sémantický průzkumník textového kulturního dědictví

Výstupy:

- Automatické rozčlenění dokumentu na části (encyklopedie, slovníky, kroniky...)
- Automatické generování popisů textů (titulky, klíčová slova, důležité části textu)
- Nalezení a analýza témat v dokumentech
- Uživatelská aplikace pro pokročilé vyhledávání v digitální knihovně (rozšíření Krameria)

semANT - Sémantický průzkumník textového kulturního dědictví

Stav:

- testování existujících jazykových modelů
- trénování vlastního jazykového modelu pro češtinu (data z webu + data z knihoven)
- dokončuje se demo webová aplikace pro textové vyhledávání (hledání textu podobného jinému textu)

semANT - Sémantický průzkumník textového kulturního dědictví

The screenshot displays the Label Studio interface. At the top, the 'Label Studio' logo is on the left, and navigation buttons for 'Instructions', 'Settings', and a user profile 'ZA' are on the right. Below the header, the project path 'Projects / textbite / Labeling' is shown. The main workspace contains a document titled '#3160' by 'michal.hradis #816' from 'about 1 month ago'. The document text is displayed in two columns, with names 'Masaryk T. G.' and 'Mauthner' at the top. Several paragraphs are highlighted in blue, and a red line connects the two columns. A sidebar on the right shows 'Info' and 'History' tabs, followed by 'Selection Details' with a list of five regions, each labeled '1 title'. At the bottom of the sidebar, there are 'Regions' and 'Relations' tabs, and a legend with 'Manual' and 'By Time' options.

Masaryk T. G. (1850–1937) považoval psychologii za nejdůležitější filosofickou vědu. Imponovaly mu fyziologické pokusy (Lipsko) a vysvětlení ps. pochodů jako vědecký korektiv čisté spekulativní psychologie. Více však než teoretická zajímala ho praktická ps., poznávání lidí, pronikání všech jiných otázek a oborů ps. zřetelem. Jinak bral za základ „Psychologii“ Brentanovou.

Spisy: pro psychologii důležité: O hypnotismu (1880), Der Selbstmord als soziale Erscheinung der modernen Zivilisation (1881, čes. 1904, 3. v. 1930), Počet pravděpodobnosti a Humova skepse (1883), Základové konkrétné logiky (1885), O nynější filosofii evoluční (1896), V boji o náboženství (1994, 2. v. 1932).

Literatura: pro psychologii směrodatná: Nejedlý Zd., T. G. Masaryk (5 dílů 1930–1936), Utitz E. M., Als Volkserzieher (1935), Drtina F., Masarykova filosofie a osvobozovací dílo (1921), Fajfr F., Masaryk a Comte (1925), Kozák J. B., Masaryk filosof (1925), Masaryk jako etik a náboženský mys-

Mauthner

lovým. Někdy objevuje se snaha u osob vlastní nedostatečné schopnosti hraditi tyto nedostatky tím, že snášeji tělesná nebo mravní příkoří od osob, na kterých jim záleží. Při trvalejším a nápadnějším objevování se lze mluvit o masochistickém chování (Hynie). Nelze však podezírat z m. různé způsoby mazlení, které podobně jako u zvířat děje se také někdy méně uhlazeným počínáním (kousáním, lehtáním, šmráním a p.).

Masselonův test stanovil původně E. Sharp. Masselon použil ho v psychiatrii. Nazýváno bývá také methodou dvou a tří slov, která mají býti podkladem věty, vyšetřovaným tvořené. Moede a Piorkowski použili k vyšetření 14letých pozmeněného t. M. Z daných slov mělo býti utvořeno více vět. Stanovena měla býti takto představová pružnost. Schüssler a Schwarzhaupt nechali dokonce ze tří slov tvořiti celá vypravování neb dávali tři slova k vypravování děje, úchylným mladistvým (od 15. do 21. roku).

Massonův kotouč, přístroj ke zkoumání zrakové vnímavosti. Bílý kotouč s ba-

semANT - Sémantický průzkumník textového kulturního dědictví

semAnt annotation

News

Annotation tasks

Annotation results

Time tracking

Annotation statistics

User management

Privacy · Terms · About PERO

← Anotace textů -- ostrá **DONE**

TASK INFO **REJECT**

Text

Když jsme pili u lékaře čaj, zašel k nim. Neočekávaně uviděl nás, a napiatě se zamračil, mlčky si sedl do koutka a počal listovat roztrhaný ročník „Nivy“.

Klíčová slova **i**

Témata **i**

Otázky **i**

semANT - Sémantický průzkumník textového kulturního dědictví



Text search demo

obézní muž dýmající doutník



Národní politika



Source: <https://www.digitalniknihovna.cz>

Vy sahnete zas po cigaru, jak rádi, kdyby bylo ještě dvakrát dražší! Proto, myslím, že by měli tito kuřáci chodit dříve do učení k našim kuřákům ze staré dobré školy. Pak by mohli státi se nebezpečnými. Ohledněme se ještě po kuřácích doutníků! Probereme jen nejvíce na odbyt jdoucí druhy. Pan starý Břicháček batolí se po ulici jako soudeček, v ústech má viržinku. Viržinka nejlépe prý slouží „tloušťíkům“, neboť tráví a odstraňuje přebytek sádla. Viržinka mu uhasíná, pán starý bafá a tahá, vzdor tomu ji ale musí znova zapalovati, ač po druhé zapálená viržinka chutná jako ohříváná pečeně Kouří a bafá až do konce. Toť kouření praktické, materiální, využitkující tabákový smotek do posledního vzdechu čili bafu! Jak jinak činí viržinkář od cechu. Toť ideal kuřáka. Avšak dále. — Ševcovský učeň s hlavou ježatých vlasů stojí za rohem domu a zapaluje si tak zvané „kuba d., za tři dvě“. Zapáliv kouří a prská jako kocour na psa. „Tohle cigaro je mizerný šmejd, já nevím, jak jen to pan mistr může snést!“ Kdybychom ho sledovali dále, stali bychom se svědky katastrofy, avšak zároveň

Využití externích AI nástrojů v Krameriu

Využití existujících online služeb (DeepL, ChatGPT, Google,...)

- sumarizace, překlad části textu, čtení textu
- volba cílového jazyka dle uživ. rozhraní
- sumarizace = překlad + sumarizace
- čtení = překlad + čtení

Cíl: otestovat využití na vícejazyčné digitální knihovně (portál EoD Open)

- poplatky za tokeny (míra využití)
- spolehlivost a rychlost externích služeb
- hledání uživatelského rozhraní pochopitelného pro uživatele



netvořily hrudky. Kvásek roztrháme
 středka na všechny strany. Když
 lepena, dáme ji schnouti pod lis.
 lepenou šachovnici pod několik k
 větších než je šachovnice!) a
 závažím nebo čistými kameny. Př
 nici před sebe vždy tak, aby kaž
 proti sobě, měl bílé pole po

Figurky.

Ku hře šachové potřebujeme
 2 střelce (běžce), 2 jezdce (koně),
 lovnu a jednoho krále. V celku
 násobný, protože každý druh fig
 Zhotovujete si je z fazolí nebo bob
 by) čerstvé a pracuješ-li s fazo
 je, než začneš. Chcete-li mít
 vyřežte si je z lipového dřeva v
 rech podle obrázků. Budete-li fig
 (bobů), vezměte pro bílé figurky
 černé fazole (boby) černé. Využit
 když o ně budete hrát a při t
 Budete-li vyřezávat ze dřeva, c
 na černo, bílým ponechejte barvu



A digital library, also called an online library, an internet library, a digital repository, a library without walls, or a digital collection, is an online database of digital objects that can include text, still images, audio, video, digital documents, or other digital media formats or a library accessible through the internet. Objects can consist of digitized content like print or photographs, as well as originally produced digital content like word processor files or social media posts. In addition to storing content, digital libraries provide means for organizing, searching, and retrieving the content contained in the collection. Digital libraries can vary immensely in size and scope, and can be maintained by individuals or organizations.[1] The digital content may be stored locally, or accessed remotely via computer networks. These information retrieval systems are able to exchange information with each other through interoperability and sustainability.

Vysvětlí Přelož Sumarizuj Přečti





netvořily hrudky. Kvásek roztrh
 středka na všechny strany. Kd
 lepena, dáme ji schnouti pod lis.
 lepenou šachovnici pod několik k
 větších než je šachovnice!) a
 závažím nebo čistými kameny. P
 nici před sebe vždy tak, aby kaž
 proti sobě, měl bílé pole po

Figurky.

Ku hře šachové potřebujeme
 2 střelce (běžce), 2 jezdce (koně),
 lovnu a jednoho krále. V celku
 násobný, protože každý druh fig
 Zhotovujete si je z fazolí nebo bob
 by) čerstvé a pracuješ-li s fazo
 je, než začneš. Chcete-li mít
 vyřežte si je z lipového dřeva v
 rech podle obrázků. Budete-li fig
 (bobů), vezměte pro bílé figurky
 černé fazole (boby) černé. Využit
 když o ně budete hrát a při
 Budete-li vyřezávat ze dřeva, c
 na černo, bílým ponechejte barvu



A digital library, also called an online library, an internet library, a digital repository, a library without walls, or a digital collection, is an online database of digital objects that can include text, still images, audio, video, digital documents, or other digital media formats or a library accessible through the internet. Objects can consist of digitized content like print or photographs, as well as originally produced digital content like word processor files or social media posts. In addition to storing content, digital libraries provide means for organizing, searching, and retrieving the content contained in the collection. Digital libraries can vary immensely in size and scope, and can be maintained by individuals or organizations.[1] The digital content may be stored locally, or accessed remotely via computer networks. These information retrieval systems are able to exchange information with each other through interoperability and sustainability.

Přelož

Generativní AI je experimentální

Digitální knihovna, nazývaná také online knihovna, internetová knihovna, digitální úložiště, knihovna beze stěn nebo digitální sbírka, je online databáze digitálních objektů, která může obsahovat text, statické obrázky, zvuk, video, digitální dokumenty nebo jiné formáty digitálních médií, nebo knihovna přístupná prostřednictvím internetu. Objekty se mohou skládat z digitalizovaného obsahu, jako je tisk nebo fotografie, i z původně vytvořeného digitálního obsahu, jako jsou soubory textového procesoru nebo příspěvky na sociálních sítích. Kromě ukládání obsahu poskytují digitální knihovny prostředky pro organizaci, vyhledávání a získávání obsahu obsaženého ve sbírce. Digitální knihovny se mohou nesmírně lišit velikostí a rozsahem a mohou je spravovat jednotlivci nebo organizace[1]. digitální obsah může být uložen lokálně nebo k němu lze přistupovat vzdáleně prostřednictvím počítačových sítí. Tyto systémy pro vyhledávání informací jsou schopny vzájemné výměny informací díky interoperabilitě a udržitelnosti.

↶ ↷ Vysvětlí Přelož Sumarizuj Přečti 🇨🇪



netvořily hrudky. Kvásek roztrá
středka na všechny strany. Kd
lepena, dáme ji schnouti pod lis.
lepenou šachovnici pod několik k
větších než je šachovnice!) a
závažím nebo čistými kameny. P
nici před sebe vždy tak, aby kaž
proti sobě, měl bílé pole po

Figurky.

Ku hře šachové potřebujem
2 střelce (běžce), 2 jezdece (koně),
lovnu a jednoho krále. V celku
násobný, protože každý druh fig
Zhotovujte si je z fazolí nebo bob
by) čerstvé a pracuješ-li s fazo
je, než začneš. Chcete-li mít
vyřežte si je z lipového dřeva
rech podle obrázků. Budete-li fig
(bobů), vezměte pro bílé figurky
černé fazole (boby) černé. Využit
když o ně budete hrát a při
Budete-li vyřezávat ze dřeva, c
na černo, bílým ponechejte barv



A digital library, also called an online library, an internet library, a digital repository, a library without walls, or a digital collection, is an online database of digital objects that can include text, still images, audio, video, digital documents, or other digital media formats or a library accessible through the internet. Objects can consist of digitized content like print or photographs, as well as originally produced digital content like word processor files or social media posts. In addition to storing content, digital libraries provide means for organizing, searching, and retrieving the content contained in the collection. Digital libraries can vary immensely in size and scope, and can be maintained by individuals or organizations.[1] The digital content may be stored locally, or accessed remotely via computer networks. These information retrieval systems are able to exchange information with each other through interoperability and sustainability.

Přelož



Generativní AI je experimentální

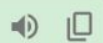


Digitální knihovna, nazývaná také online knihovna, internetová knihovna, digitální úložiště, knihovna beze stěn nebo digitální sbírka, je online databáze digitálních objektů, která může obsahovat text, statické obrázky, zvuk, video, digitální dokumenty nebo jiné formáty digitálních médií, nebo knihovna přístupná prostřednictvím internetu. Objekty se mohou skládat z digitalizovaného obsahu, jako je tisk nebo fotografie, i z původně vytvořeného digitálního obsahu, jako jsou soubory textového procesoru nebo příspěvky na sociálních sítích. Kromě ukládání obsahu poskytují digitální knihovny prostředky pro organizaci, vyhledávání a získávání obsahu obsaženého ve sbírce. Digitální knihovny se mohou nesmírně lišit velikostí a rozsahem a mohou je spravovat jednotlivci nebo organizace[1]. digitální obsah může být uložen lokálně nebo k němu lze přistupovat vzdáleně prostřednictvím počítačových sítí. Tyto systémy pro vyhledávání informací jsou schopny vzájemné výměny informací díky interoperabilitě a udržitelnosti.

Sumarizuj



Generativní AI je experimentální



- Digitální knihovna je online databáze digitálních objektů, která je přístupná prostřednictvím internetu.
- Obsahuje různé formáty digitálních médií, jako jsou text, obrázky, zvuk nebo video.
- Systémy digitálních knihoven umožňují vzájemnou výměnu informací díky interoperabilitě a udržitelnosti.



Vysvětlí

Přelož

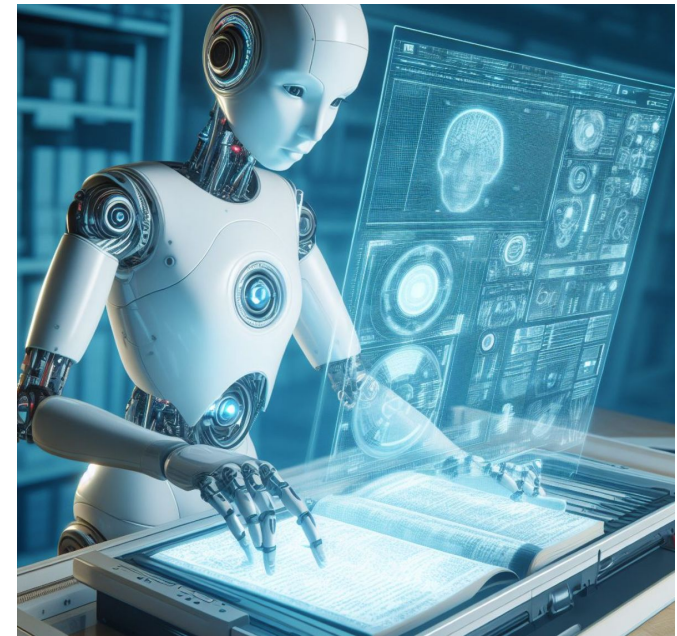
Sumarizuj

Přečti



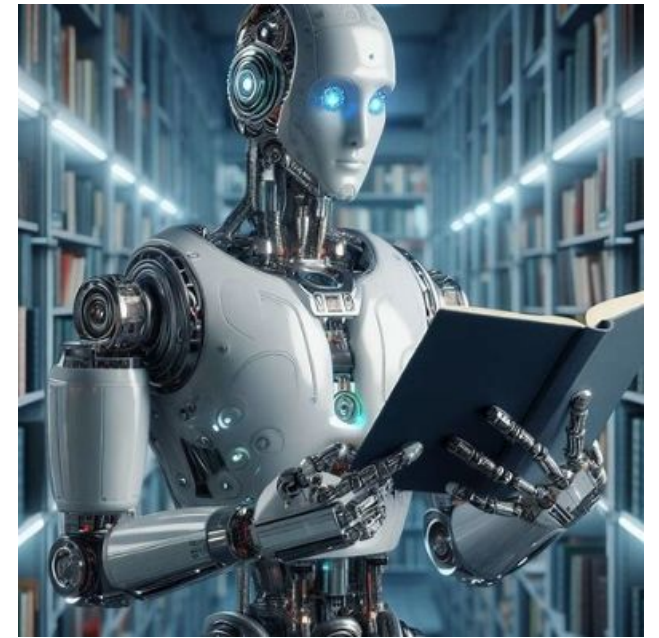
Využití strojového učení

- **Během digitalizace (jednorázově)**
 - Řízení automatického skeneru
 - Scelování velkých dokumentů
 - Tvorba bibliografických metadat
 - Ořez, úprava obrazu (zvuku, videa...)
 - Tvorba strukturálních metadat
- **Před zpřístupněním (opakovaně; anotace)**
 - Obrázky: lokalizace na stránce, kategorizace, anotace, ...
 - OCR, OMR, identifikace osob na obrázcích
 - Speech to text
 - Identifikace článků a kapitol, segmentace textu
 - Sémantické zpracování textu (identifikace míst, osob apod.)
 - Indexace obrazu či textu pomocí vektorů
 - Trénování velkých jazykových modelů



Využití strojového učení

- **Během zpřístupnění (na vyžádání)**
 - Ústní dotazy (speech to text)
 - Překlady
 - Sumarizace
 - Čtení (text to speech)
 - Hledání, doporučování, podobnost (textu, obrázků)
 - Našeptávač
 - Chatbot (dotazy na obsah dokumentů)
- **Speciální úlohy**
 - generování databází (ze slovníků, tabulek, soupisů)
 - zpracování matematických nebo chemických vzorců
 - georeferencování map
 - ...



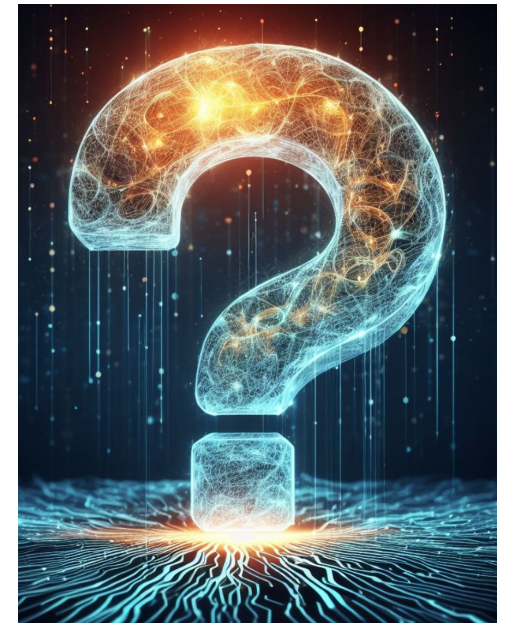
Kramerius a AI do budoucna

Rozšíření o výstupy NAKI projektů, jak?

- Co je nutné předpočítat, co je možné dělat v reálném čase na vyžádání
- Hledání textu a obrázků jedním dotazem
- Jak rozlišit přesné hledání od AI hledání podobnosti
- Hledat podobný text, podobný obrázek, stejnou osobu...
- Povídej si s knihou, skupinou dokumentů...
- Přehrát noty na stránce jako midi? Jak to midi uložit?

Výzvy a omezení

- Etické otázky
 - autorská práva k trénovacím datům, jejich ochrana
 - práva k výstupům AI
 - ochrana osobních/důvěrných údajů v trénovacích datech
 - předsudky zesilované nevhodnou volbou trénovacích dat
 - učení AI lhát
- Kvalita trénovacích dat
- Náklady (předzpracování dat vs. úlohy v reálném čase)
 - V DK MZK je nyní cca 22 miliard slov
- Halucinace jazykových modelů



Budoucnost

Mírný pokrok v mezích zákona

vs.

budeme ještě potřební?



Komunita AI4LAM



- [AI4LAM.org](https://ai4lam.org) (AI pro knihovny, archivy a muzea)
- [slack](#), [X](#), [google groups](#), [github](#)
- otevřená platforma pro spolupráce paměťových organizací v oblasti AI
- zlepšuje povědomí a sdílí zkušenosti ve využití nástrojů umělé inteligence

Pravidelná setkání

- Zoom online jednou za ca. 6 týdnů, obvykle v 17 nebo 18 hod. našeho času, k dispozici zápisy a nahrávky jednání
- Konference Fantastic Futures

vzniká [IIIF AI + Machine Learning Community Group](#)



Děkujeme za pozornost!

Dotazy



Školení knihovnických personálních asistentů, ca 2033, kolorováno