

PERO

OCR for old(er) documents

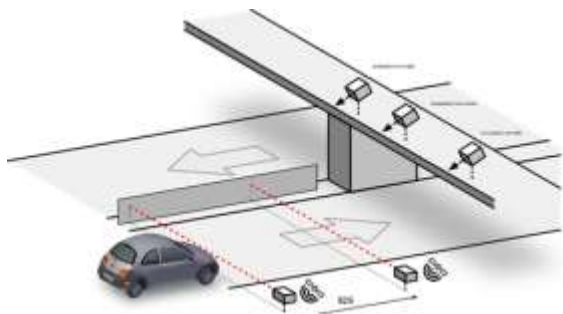
Michal Hradiš, Petr Žabička, Alžbeta Zavřelová







Zkušenosti - Strojové učení v počítačovém vidění



Pavel Svoboda, Michal Hradiš, Lukas Maršík, Pavel Zemčík:
CNN for License Plate Motion Deblurring. ICIP 2016.

Input

CNN output



PERO - Pokročilá extrakce a rozpoznávání obsahu tištěných a rukou psaných digitalizátů pro zvýšení jejich přístupnosti a využitelnosti

Vysoké učení technické v Brně - Michal Hradiš

Moravská zemská knihovna - Petr Žabička

WEB: <http://pero.fit.vutbr.cz>

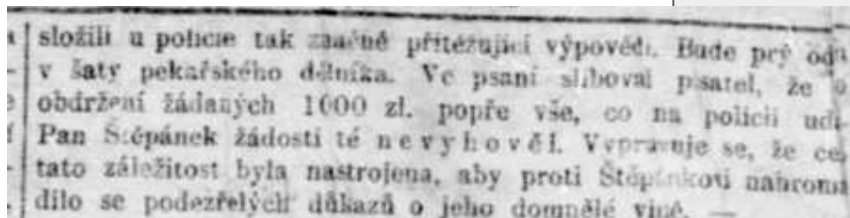


Adam Matthew - Colonial America / East India Comp.

- CA - 70,000 rukopisů – nyní plně indexovaný a dohledatelný obsah díky technologii rozpoznávání ručně psaného písma
- <https://www.amdigital.co.uk/primary-sources/colonial-america>
- <https://www.amdigital.co.uk/primary-sources/east-india-company>



Kvalita OCR



ABBY

složili o policie tak ju. i,é pruériji. i výpověši. ■■■■

v šaty pekařského défiuSa. Vc psaní sliboval p ■■■■

obdržení žáiínýeh 1000 zl. popře vše, co na ■■■■

P*n řččpánek žádosti té uevykovčl. Vvrra-.ej ■■■■

tato záležitost byla na«ro|eча. aby proti Stop ■■■■

dílo se podezřelých ilikazů o jeho doranèlc vii. ■■■■

PERO

složili u policie tak značné přítěžující výpovědi. Bude prý oda

v šaty pekařského děníka. Ve psaní sliboval psatel, že o

obdržení žádaných 1000 zl. popře vše, co na policii udi-

Pan Štěpánek žádosti té nevyhověl. Vypravuje se, že cel

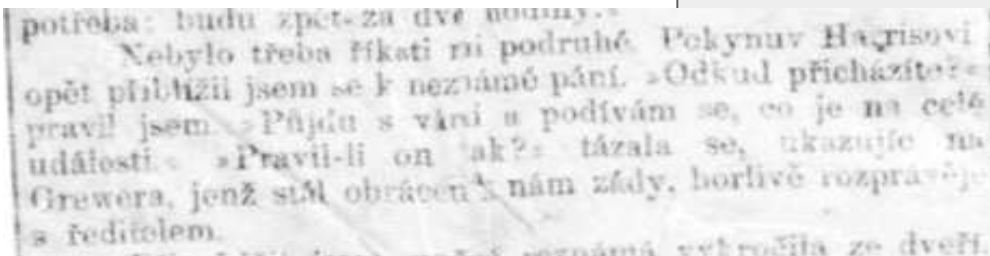
tato záležitost byla nastrojena, aby proti Štěptnkov nahroma

dilo se podezřelýcl důkazů o jeho domnělé vině. —

Kvalita OCR

ABBY

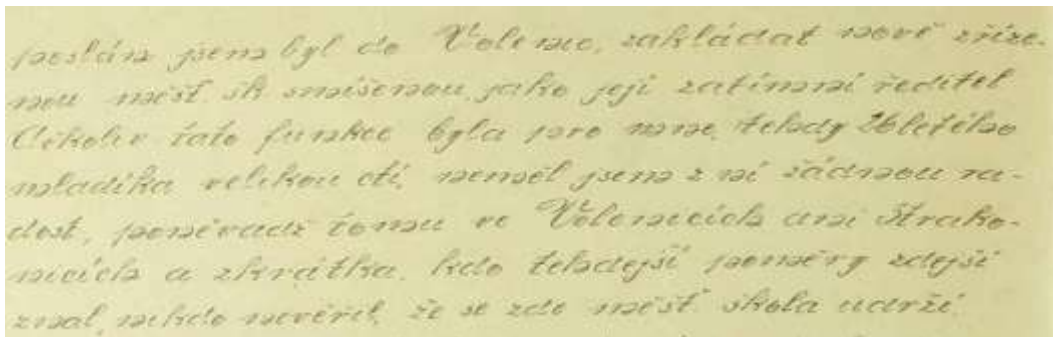
-livlo i ludi i |H>d nl.é l'- V r. Haj-isov; přichazít:- ■ i i- ii«
ceh dál osti i i Pra vil- nž stál , nera l



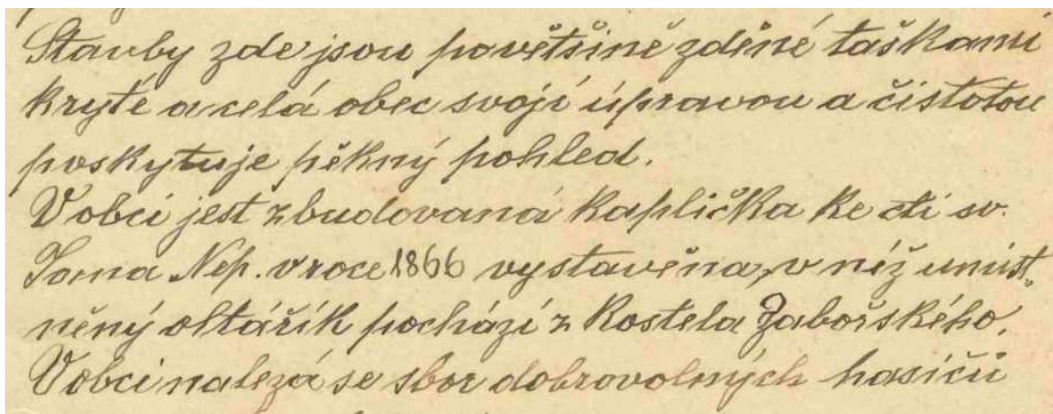
PERO

Nebylo třeba říkati ni podruhé. Pokynuv Hagridovi
opět přiblížil jsem se k nez^támé pání. »Odkud přicházíte?«
pravil jsem. »Půjdu s váⁿi a podívám se, co je na celé
události.« »Pravil-li on ^aak? tázala se, ukazujíc na
Grewera, jenž stál obrácen^k nám zády, horlivě rozprávěje
s ředitelem.

České kroniky



poslán jsem byl do Volenic, zhládat nově
zříze-
nou měst. šk. smíšenou, jako její zatínní
ředitel.
Dčkoliv tato funkce byla pro mne, tehdy 4letého
mladíka velikou ctí, neměl jsem z ní žádnou ra-
dost, poněvadž tomu ve Velenicích ani ltraho-
nicích a zkrátka, kdo tehdejší poměry zdejší
znal, nikdo nevěřil, že se zde měst. škola udrží.



Stavby zde jsou pro většinu z nich takzvané
kryté a celá obec svojí úpravou a čistotou
poskytuje pěkný pohled.
V obci jest zbudovaná kaplička ke cti sv.
Joma Nep. v roce 1866 vystavěna, v níž umíst.
něný oltářík pochází z kostela Gosbořského.
V obci nalezá se sbor dobrovolných hošičů

poslán jsem byl do Volenic, zahládat nově
zříze-
nou měst. šk. smíšenou, jako její zatínní
ředitel.

Dčkoliv tato funkce byla pro mne, tehdy 4letého
mladíka velikou ctí, neměl jsem z ní žádnou ra-
dost, poněvadž tomu ve Velenicích ani ltraho-
nicích a zkrátka, kdo tehdejší poměry zdejší
znal, nikdo nevěřil, že se zde měst. škola udrží.

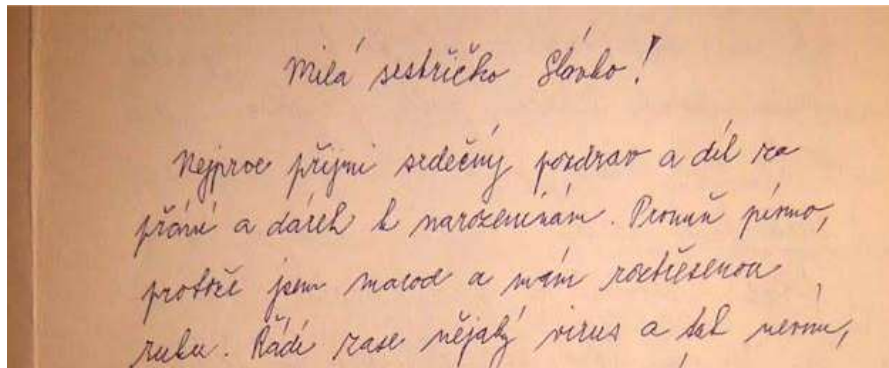
kryté o celá obec svojí úpravou a čistotou
poskytuje pěkný pohled

V obci jest zbudovaná kaplička ke cti sv.

Joma Nep. v roce 1866 vystavěna, v níž umíst.
řěný oltářík pochází z kostela Gosbořského

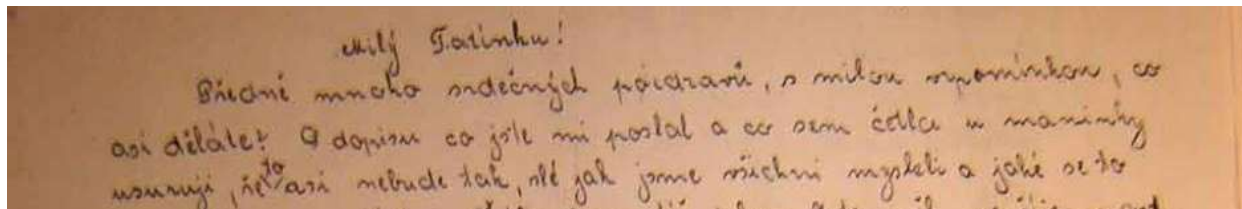
T obci na kazí se sbor dobrovolných hošičů

České písmo



Milá sestríčko Slavko,

Nejprve přijmi srdečný pozdrav a dík za přání a dárek k narozeninám. Promiň písmo, protože jsem marod a mám roztřesenou ruku. Řadí zase nějaký virus a tak nevím,



Milý Tatínku!

Předně mnoho srdečných pozdravů, s milou vzpomínkou, co asi děláte? S dopisu co jste mi poslal a co sem četla u maminky usuzuji, že to asi nebude tak, zlé jak jsme všichni mysleli a jaké se to

Příklady výsledků OCR

Members, at a sum not exceeding 5s. each.

Members, at a sum not exceeding 5s. each.

Previous to the Tragedy **the BAND** will perform, (Third Time)

Previous to the Tragedy the BAND will perform, (Third Time)

Pan Piefczarzycki. W Kitáygrodzie w każd~~ę~~y b~~ę~~ćcie rot~~ę~~ ied~~ę~~

Pan Piefczarzycki. W Kitáygrodzie w każd~~ę~~y b~~ę~~ćcie rot~~ę~~ ied~~ę~~

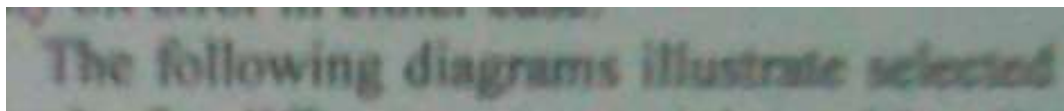
teżnie zno~~ś~~ac. Czar nas Szuyfki po~~ś~~ął nan kilk~~ę~~ razy Kni~~ę~~zi~~ę~~ Sz~~ę~~

teżnie zno~~ś~~ac. Czar nas Szuyfki po~~ś~~ął nan kilk~~ę~~ razy Kni~~ę~~zi~~ę~~ Sz~~ę~~

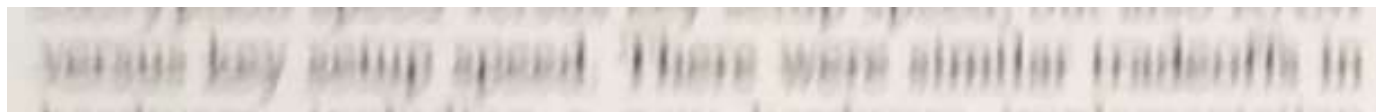
Příklady na fotografiích z mobilních telefonů



ative State Research, Education and Extension



The following diagrams illustrate selected



versus key setup speed. There were similar tradeoffs in



practice of coding hardware-specific

Zlepšování kvality

třídy přeplněny. Zřízení této školy vyžadovalo mnoho práce a kdo se nejvíce proti němu stavěl, byl purkmistr Wieser. (Hlučné provolávání hanby.)

Jak si radnice počíná proti úřadům, svědčí to, že nechala reklamace české okresní školní rady, jež podány byly na vyloučení 600 českých dětí z německých škol zcela klidně ležet a věc tu nevyřizuje. Radnice je zde podporována vládou a proti takovému nepříteli nezbývá žádný jiný prostředek, než jen jeden — volební lístek. (Bonělivý souhlas.)

Za zřízení českých měšťanských škol v Brně bojují Čechové již od r. 1881, leč marně. Nyní by snad byla radnice ochotna dáti nám jednu měšťanskou školu, kdybychom ustoupili od požadavku rozdě-

třídy přeplněny. Zřízení této školy vyžadovalo mnoho práce a kdo se nejvíce proti němu stavěl, byl purkmistr Wieser. (Hlučné provolávání hanby.)

Jak si radnice počíná proti úřadům, svědčí to, že nechala reklamace české okresní školní rady, jež podány byly na vyloučení 600 českých dětí z německých škol zcela klidně ležet a věc tu nevyřizuje. Radnice je zde podporována vládou a proti takovému nepříteli nezbývá žádný jiný prostředek, než jen jeden — volební lístek. (Bonělivý souhlas.)

Za zřízení českých měšťanských škol v Brně bojují Čechové již od r. 1881, leč marně. Nyní by snad byla radnice ochotna dáti nám jednu měšťanskou školu, kdybychom ustoupili od požadavku rozdě-

Najděte změnu

třídy přeplněny. Zřízení této školy vyžadovalo mnoho práce a kdo se nejvíce proti němu stavěl, byl purkmistr Wieser. (Hlučné provolávání hanby.)

Jak si radnice počíná proti úřadům, svědčí to, že nechala reklamace české okresní školní rady, jež podány byly na vyloučení 600 českých dětí z rakouských škol zcela klidně ležet a věc tu nevyřizuje. Radnice je zde podporována vládou a proti takovému nepříteli nezbývá žádný jiný prostředek, než jen jeden — volební lístek. (Bonělivý souhlas.)

Za zřízení českých městanských škol v Brně bojují Čechové již od r. 1881, leč marně. Nyní by snad byla radnice ochotna dát nám jednu městanskou školu, kdy-

třídy přeplněny. Zřízení této školy vyžadovalo mnoho práce a kdo se nejvíce proti němu stavěl, byl purkmistr Wieser. (Hlučné provolávání hanby.)

Jak si radnice počíná proti úřadům, svědčí to, že nechala reklamace české okresní školní rady, jež podány byly na vyloučení 600 českých dětí z rakouských škol zcela klidně ležet a věc tu nevyřizuje. Radnice je zde podporována vládou a proti takovému nepříteli nezbývá žádný jiný prostředek, než jen jeden — volební lístek. (Bonělivý souhlas.)

Za zřízení českých městanských škol v Brně bojují Čechové již od r. 1881, leč marně. Nyní by snad byla radnice ochotna dát nám jednu městanskou školu, kdy-

PERO - plánované výsledky

- Open source moduly and připravené OCR modely pro rozpoznávání tištěného i ručně psaného písma
- OCR API hostované na VUT v Brně
- Webová aplikace pro OCR a ruční korekce přepisů
- Open source moduly pro odhad kvality dokumentu a pro zlepšení čitelnosti
- Open source nástroje pro extrakci sémantických informací

LIDOVÉ NOVINY

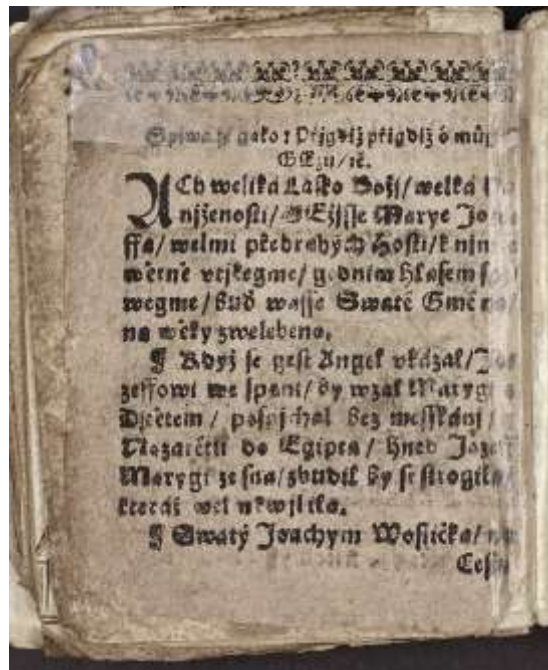
War je o 6. stranách.

• **emulvil.**

Il maschio dell'animale possiede una struttura anatomiche che gli consente di inseguire la femmina. Il maschio possiede un paio di appendici che si chiamano "pedipalpi" e che sono utilizzati per afferrare la femmina durante l'accoppiamento. Inoltre, il maschio possiede un paio di "gnathopodi" che sono utilizzati per afferrare e masticare il cibo.

[illegible]

soodvolyň. Toto tvrzení je leš
to samostatnost, když kněz, chč
ovati, musí si napřed opatřiti
vé svolení své vrchn
kud svolení toho nem
f se mu mandátu ani



1820

Penner

14.
15.

Translatus 214

7

4

St. Johann	1.	1.
------------	----	----

7702

John F. Johnson

Ernst

2d /
9-11-1880

Zeichnung

Indefinite

Figuring.

Humaria

James
L. G. Collins

28. 2

Fig. 11

1871

Fazio.

2. - 1" 4

Randolph
 S. L. L.

May 20 1900

Daß Jedem Gutes wie, im Innern Gemüths
wahrnehmend, und gütigkeitsvoll
als das was es ist, zu thun, das ihm
Jedem im Gutes zu werden, beistehend ist
mühen, im Gutes zu thun, das ihm
zu thun, das ihm zu thun, das ihm zu thun.

Frederick John and S. Dalton.

Größen zugehörig und zugehörig.

Anton Dybner vel' žnigru.

Labradorer Farnes, Längel von 2. 170

17.9

Lynx

Знаменца

July 29

Longmont

Kroniky 20. století

Stavby zde jsou pro většinu zděné taškami
kryté a celá obec svou úpravou a čistotou
poskytuje pěkný pohled.

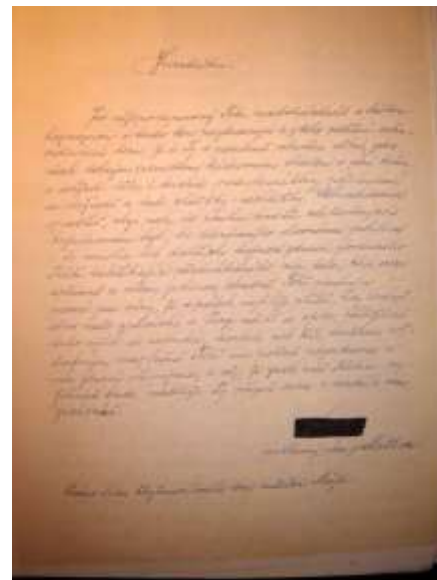
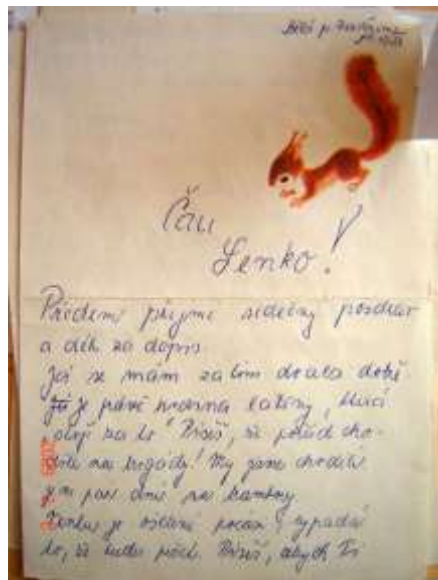
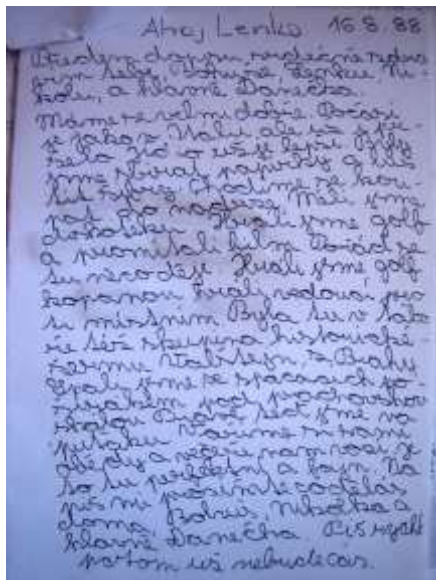
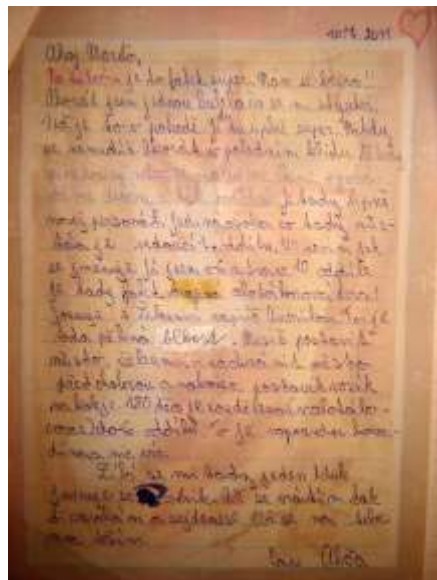
V obci jest zbudovaná kaplička ke ctě sv.
Jana Nep. v roce 1866 vystavěna, v níž umístěný
oltářník pochází z kostela Zdobovského.
V obci nalezá se sbor dobrovolných hasičů

Datové sady

- [IMPACT resources](#) – Evropské starší tisky
 - 1.2M řádků, 9 jazyků, 10 písem
 - 0.5 % chybovost na znak
- Deutsches Textarchiv (~6000 tisků)
 - Většinou německá fraktura
 - 0.3 % chybovost na znak
- České tisky
 - Noviny digitalizované z mikrofilmů
 - Starší tisky
- Bentham
 - Rukopisy několika autorů - 21,403 stránek
 - Potřeba přiřadit text řádkům
- České dopisy 20. stol. - 2000 dopisů

České dopisy

- Sbírka 2000 českých ručně psaných dopisů
- Dopisy z let 1902-2012
- Přepisy bez informací o řádcích
- Hladká, Zdeňka a kol. (2013): 111 let českého dopisu v korpusovém zpracování. Brno: Masarykova univerzita.



Nový detektor řádků

1. What is the purpose of the text?
 The purpose of the text is to provide information about the history and development of the city of London.

2. What is the main idea of the text?
 The main idea of the text is that London has a long and rich history, and it has been a major center of trade and commerce for centuries.

3. What are the key points of the text?
 The key points of the text are:

- London was founded by the Romans in 43 AD.
- It was a major center of trade and commerce during the Middle Ages.
- The city was destroyed by fire in 1666.
- It was rebuilt and became a major center of industry and commerce in the 18th and 19th centuries.
- It is now one of the most important cities in the world.

4. What are the details of the text?
 The details of the text include:

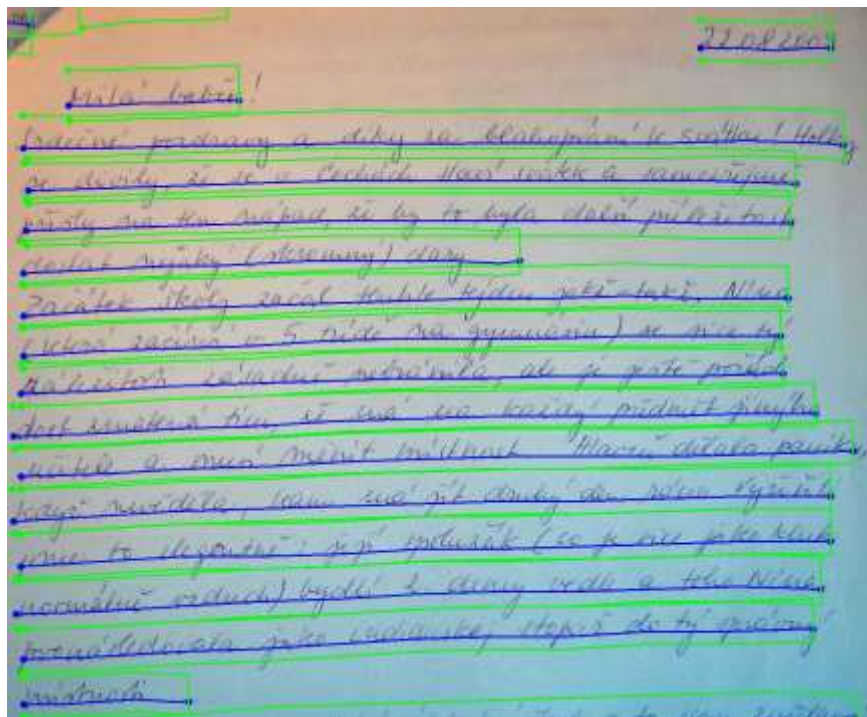
- The city was founded by the Romans in 43 AD.
- It was a major center of trade and commerce during the Middle Ages.
- The city was destroyed by fire in 1666.
- It was rebuilt and became a major center of industry and commerce in the 18th and 19th centuries.
- It is now one of the most important cities in the world.

5. What are the conclusions of the text?
 The conclusions of the text are:

- London has a long and rich history.
- It has been a major center of trade and commerce for centuries.
- The city was destroyed by fire in 1666.
- It was rebuilt and became a major center of industry and commerce in the 18th and 19th centuries.
- It is now one of the most important cities in the world.

[illegible]

Zarovnání textu s řádky



22. 08. 2009

Milá babčo!

Srdečné pozdravy a díky za blahopřání k svátku! Holky se divily, že se v Čechách slaví svátek a samozřejmě přišly na ten nápad, že by to byla další příležitost dostat nějaký (skromný) dary.

Začátek školy začal tenhle tejden jakž-takž. Nína (která začíná v 5. třídě na gymnáziu) se sice tý záležitosti zásadně nebránila, ale je ještě pořád dost zmatená tím, že má na každý předmět jinýho učitele a musí měnit místnost. Hlavně dělala paniku, když nevěděla, kam má jít druhý den ráno. Vyřešili jsme to elegantně: její spolužák (co je sice jako kluk normálně vzduch) bydlí v 2. domy vedle a toho Nína pronásledovala jako indianskej stopař do té správný místnosti...

Přepsáním následujícího textu přispějete k vývoji automatického přepisu českých historických dokumentů. Naším cílem je, aby bylo v budoucnu možné online vyhledávat například v obsahu kronik, matričních a pozemkových knih i historické korespondence. Prosím, přesně přepište vytištěný text kamkoliv na tuto stránku se zachováním řádků i případných chyb. Vyplněním tohoto listu souhlasíte s jeho zveřejněním a volným využitím pro libovolné účely. Vyplněný list prosím zašlete na adresu: Michal Hradiš, Fakulta informačních technologií, VUT v Brně, Božetěchova 1/2, 612 66 Brno. Případně list naskenujte na stolním skeneru a zašlete na adresu: ihradis@fit.vutbr.cz s předmětem „Brno HWR dataset“. —

6253419

„Těhle loutně se chce hrát jenom
v moll,“ usmál se na něho Arren z
okna. „Chce se jí plakat. A co
byste chtěli slyšet, vážení
hostitelé?“

„Já jim to řeknu, Zebbie.
Normální hladinu.“ Hilda počkala,
dokud jí Zeb nedal signál, a pak
zvolala: „Chlapci! Plukovník

Brumby potřebuje pomoc. Pojd'te si
pro něj. Už vás nebudeme strašit
řevem.“

„Tím vám vyhrožuje?“

13.

„Ne, pane.“

„To dělám i dnes,“ usmál se
Alexijev. „Předně - tihle tu jsou
déle, měli by se uzdravit dřív. A
pokud jste slyšeli, co jsem říkal

6253419

„Těhle loutně se chce hrát jenom
v moll,“ usmál se na něho Arren z
okna. „Chce se jí plakat. A co
byste chtěli slyšet, vážení
hostitelé?“

Brumby potřebuje pomoc. Pojd'te si
pro něj. Už vás nebudeme strašit
řevem.“

„Tím vám vyhrožuje?“

13.

„Ne, pane.“

„To dělám i dnes,“ usmál se

volebního — budějovického — prorazí.

Už dosti dlouho nrskalo se zde pověst, že literární-
řečnický spolek Slavia vst. rozpuštěn. Dávody jež s-
nvádí jsou tak za... nřtazeny. ... nich nedoměšlím
ani do našich výminečných úřadů Slavia. ... spolek m-
odpovídati za to, co př. jednotlivci, snad i jen ... slovo
provedli při instalaci rektor Křesťotka a na technice. Tak
to budí úřední kyselost, že e surávě Slavia převládá strana
„Neodvislosti“ a „Nových bouř.“ Starosta Slavia je nyní
MUC. Dřtka, jehož bratr je také jako „omladinista“
zatčen.

Když nř jsem při tom — denním thematu pražském
při zatýkání strašných zločinnů kteří se na Praze vyrojili
podotýkám, že zatčen zde byl př. předevčírem MUC. La-
Grosman z Lužce, známý přítel zatčených Sokola, Haima
Rašina atd. „Zasvědčení“ oviní jeho zatčení s jakými-
nočními pouličními demonstracemi které se nedávno udály
př. na Vinohradech a při nř př. neznámými mladíky pro-
znévovány „protidynastické“ řísně.

Politika.

nil, stegně nám učinj. V trůnu věčného stoji

stela gdaui/ tázaný slova Bo

nedbávagi. K zachovávanj pak Církyve musjt

žijho bedlivě poslauchagj/

Země Lužytánská. Anobř ať Katalog celého Swěta po straně

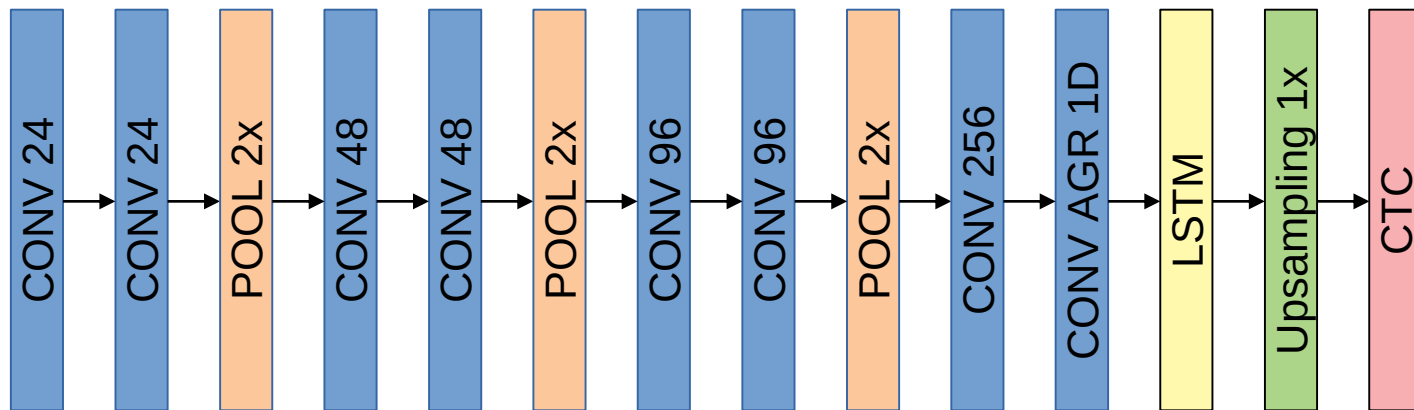
stutků Božich, a čteť pilně Písino swaté,

Předmluva.

Lečna.

Řádky textu nám přepisují neuronové sítě

fer tan delicado que se deshaze en la boca, no se senti-



Jazykové modely

Která věta je pravděpodobnější?

Unicorns from France like to visit libraries.

or

Unicorns from France likes to visit libraries.

Language models

Co může být další slovo ve větě?

People like to go to ???????

library

cinema

pubs

restaurants

....

work

OpenAI GTP-2: Příklad generovaného textu

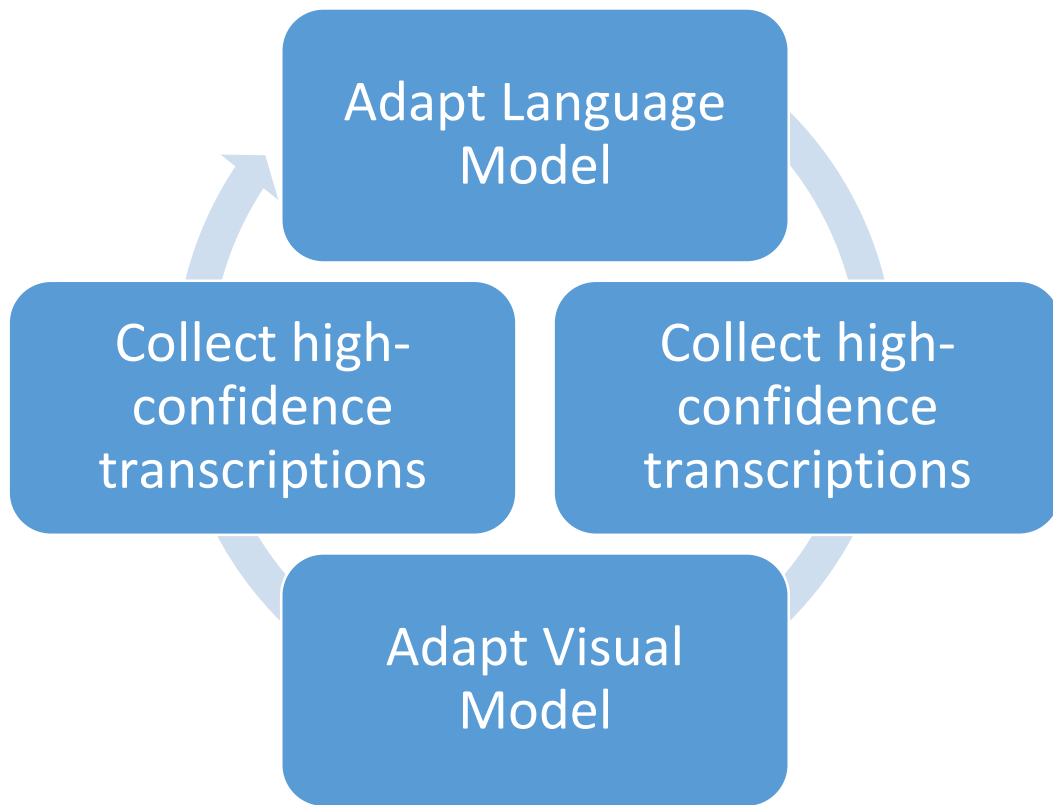
***Text seed:** In a shocking finding, scientist discovered a herd of unicorns living in a remote, previously unexplored valley, in the Andes Mountains. Even more surprising to the researchers was the fact that the unicorns spoke perfect English.*

Generated text: The scientist named the population, after their distinctive horn, Ovid's Unicorn. These four-horned, silver-white unicorns were previously unknown to science.

Now, after almost two centuries, the mystery of what sparked this odd phenomenon is finally solved.

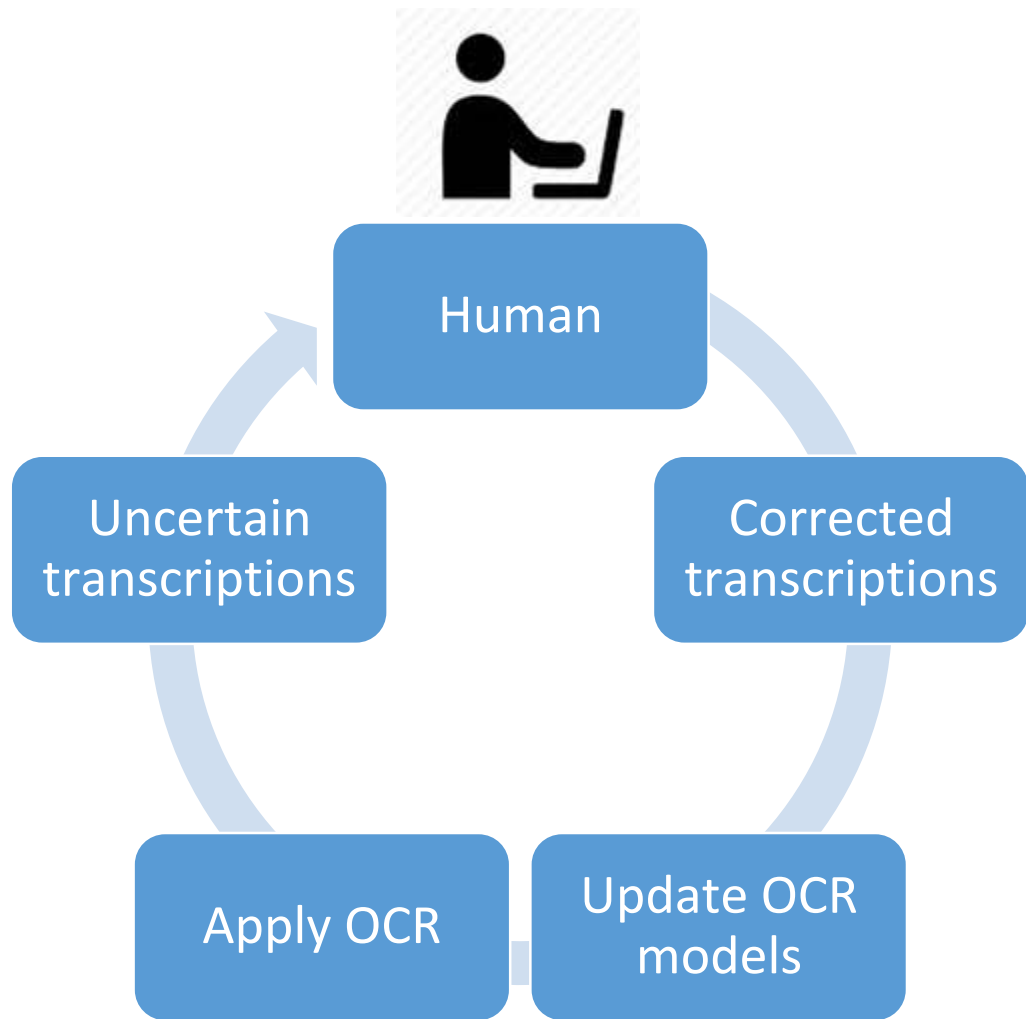
Dr. Jorge Pérez, an evolutionary biologist from the University of La Paz, and several companions, were exploring the Andes Mountains when they found a small valley, with no other animals or humans. Pérez noticed that the valley had what appeared to be a natural fountain, surrounded by two peaks of rock and silver snow.

Adaptace na specifický dokument



Bez lidí se neobejdeme

f horrors and desolation.	29s to 31s,
f horrors and desolation.	29s to 31s
en Heugouwen woonach	ИМОТЬ СЕ
en Heugouwen woonach	ИМОТЬ СЕ
kse-tiende	пошавнувало,
kse-tiende	пошавнувало
are the working	this Evening Tuesday
are the working	this Evening Tuesday
weet men too veel	AMINER will
weet men too veel	AMINER will



Web pro OCR a ruční korekce přepisů

from doing to suffering, tastes the griefs as well as glories of fame, and is well thrashed by Mr Curll's chairman as he leaves the player's house,—the act closed with a roar. There should have been no second. There is nothing but stale repetition in the second act. The ladies retort with a practical

Leaflet

glories of fame, and is well thrashed by Mr Curll's chairman as he leaves the player's house,—the act closed with a roar. There should have been no second. There is nothing but stale repetition in the second act. The ladies pett with a pical

Coarse styles
($4^2 - 8^2$)



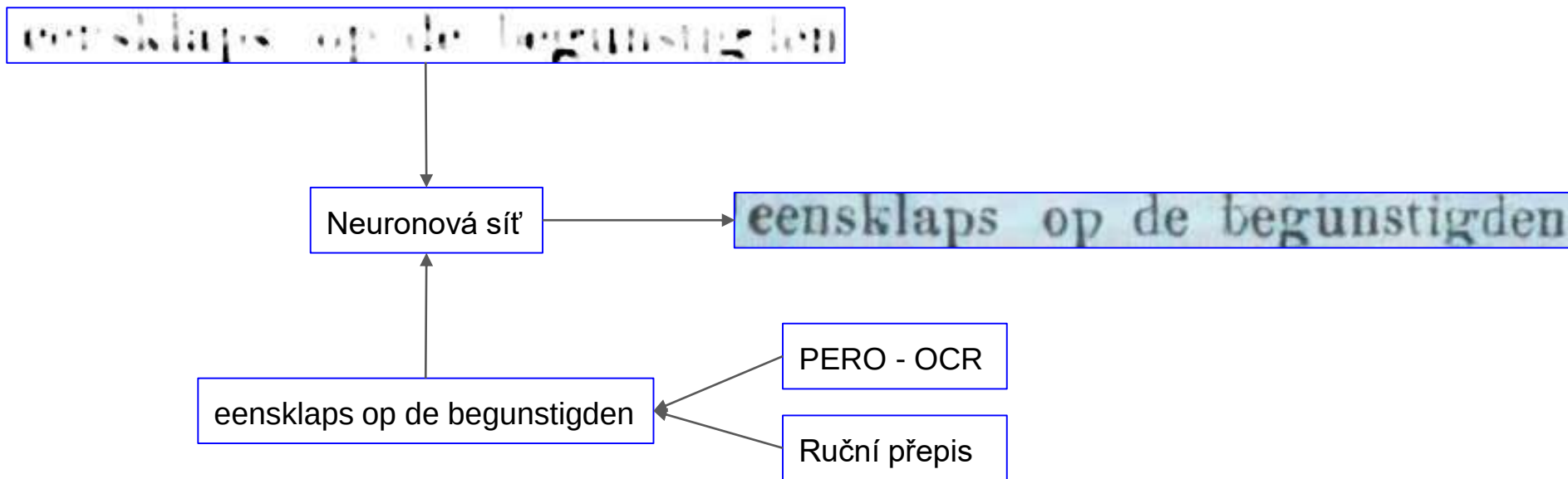
Middle styles
($16^2 - 32^2$)



Fine styles
($64^2 - 1024^2$)



Zlepšování kvality s referencí



Příklady zlepšování kvality textu

eensklaps op de begunstigden over te brengen; want
eensklaps op de begunstigden over te brengen; want
eensklaps op de begunstigden over te brengen; want
eensklaps op de begunstigden over te brengen; want

pieczęć zapieczętować. Tak Obrzezanie było znane
pieczęć zapieczętować. Tak Obrzezanie było znane

Powierzyć nie dano / o kruku nam piosnkę gróć kazał
Powierzyć nie dano / o kruku nam piosnkę gróć kazał

Podvrhy dokumentů

BEATA IASPAS NAFRYCY PU	Dobrupospol. temu af. ft.	Dobrze b
BEATA IASPAS NAFRYCY PU	 temu af. ft.	Dobrze b
BEATA IASPAS NAFRYCY PU	Bratislava temu af. ft.	Dobrze b
BEATA IASPAS NAFRYCY PU	Knihovna temu af. ft.	Dobrze b
BEATA IASPAS NAFRYCY PU	Polovium temu af. ft.	Dobrze b
BEATA IASPAS NAFRYCY PU	lenochod temu af. ft.	Dobrze b
BEATA IASPAS NAFRYCY PU	monitor temu af. ft.	Dobrze b
BEATA IASPAS NAFRYCY PU	bashtara temu af. ft.	Dobrze b
BEATA IASPAS NAFRYCY PU	gentleman temu af. ft.	Dobrze b

Podvrhy dokumentů

ad miseros & orphanos vngaros . qui ferè ad flagellum nati
ad miseros & [REDACTED] vngaros . qui ferè ad flagellum nati
ad miseros & Bratislava vngaros . qui ferè ad flagellum nati
ad miseros & Knibovna vngaros . qui ferè ad flagellum nati
ad miseros & Eoloſvium vngaros . qui ferè ad flagellum nati
ad miseros & lenochob vngaros . qui ferè ad flagellum nati
ad miseros & monitor vngaros . qui ferè ad flagellum nati
ad miseros & baſtyara vngaros . qui ferè ad flagellum nati
ad miseros & gentlenan vngaros . qui ferè ad flagellum nati

Hodnocení kvality

tických bojích nepočítá
měrodatnými jen povu
nost. Podáním návrhu
právo hlasovací nechtě
ký nijak zavděčiti něm
iratickým vůdcům. S je
u nás nikdo nepočítá,
e výhradně zájem a p
o lidu. Také nám ner
si hleděli získati důvěr
a Eldeceho Nizem
voinější dychání. S ne
tali přeměnu v povětrnost
ia, odsouzení k práci d
sných kobách, ale hlavně
teří na vyprahlé pole, i
ueí fepy pohlíželi již s
v náladě teskné. Dou
včetrnosti vyhoví přáním
pož tolik osvěžení a vlá
sdárnému vývoji letního
Leti Václavě náhlému

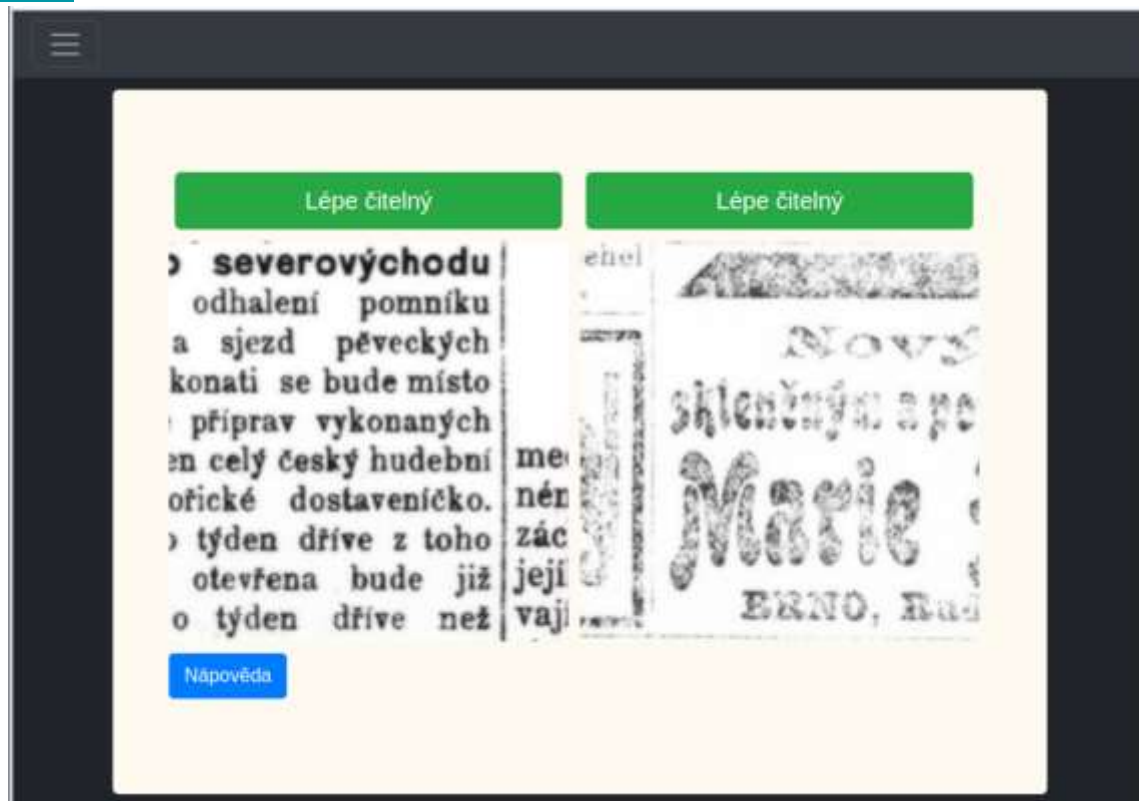
opatili si obě nohy. Čle
se poskytl neopatrnému
bí.
— Hra o dynamitem
si 30letý dělník Emil Háj
namitovou patronou a
ní v ní kloboukem. Z no
ebia a neopatrného mladí
na levé ruce. Nejen že mu
potrhala, ale zlomila m
a dno vyprázdnila. Pivo
ajednou se vleje do hrdla
stkou chvíli pak počala o
eli s touto okolností ná
sepějeme k diagnóze, že
vé smylem se asi da
odivnější je působení tét
přece byl zabrán: slič
postele a polehtaví pávis

se poskytl. Na kom
a sít však vlnul ka
reálna došlo zachycup
se k pronášení čle
od dne 15. srpna v ho
a jeho duška, aby vaka
s Mla. Odravě s
e zařazení českých mužů
it se zručuje s důvod
dující čtvrtrohodině již
gné neřeklo mu dříve
dného pramene, že si d
fisahá na spasení své d
týrá jej při každé příle
divost jest umělé vypě
z počátku mrzut, pak
poměr, a konečně —
která již nyní bez dův
taky, nemůže býti šča
a o sebevraždu. — St

Hodnocení kvality

<https://pero.fit.vutbr.cz/post/2019-09->

[11 1 Quality annotation](#)



Hyperspectral imaging



403.09 nm



469.68 nm

