

Moderní zpřístupnění historických pramenů

Představení projektu

P. Král^{1,2} K. Halla³ R. Široký⁴ L. Lenc² J. Martínek¹

¹Katedra informatiky a výpočetní techniky, FAV ZČU v Plzni

²Nové technologie pro informační společnost - NTIS, FAV ZČU v Plzni
{pkral,llenc,jimar}@kiv.zcu.cz

³Státní oblastní archiv v Plzni, Státní okresní archiv Cheb
halla@soaplzen.cz

⁴Západočeský institut pro ochranu a dokumentaci památek - ZIP o.p.s.
rsiroky@zip-ops.cz

12. října 2018

Vymezení oblasti

- Zpracování historických dokumentů
- Typy dokumentů
- Řada výzkumných problémů

Současný stav

- Dokumenty jen v tištěné podobě
- Digitalizace → záchrana kulturního dědictví
- Několik webových portálů pro zpřístupnění (Porta Fontium¹, apod.)
- Omezené možnosti vyhledávání

Hlavní cíl

- Zpřístupnění archivních pramenů z česko-bavorského příhraničí širokému spektru uživatelů pomocí nejmodernějších informačních technologií

¹<http://www.portafontium.eu>

The screenshot shows the main interface of the Porta Fontium website. At the top, there is a navigation bar with the logo and name 'Porta fontium', a menu icon labeled 'Obsah', a search icon labeled 'Vyhledávání', and utility links for 'O projektu', 'Nápověda', 'Čeština', and 'Deutsch'. Below the navigation bar is a grid of document categories, each represented by a blue circular icon and a text label. The categories are: 'Matriky' (tree icon), 'Kroniky' (open book icon), 'Listiny' (document with pointer icon), 'Fotografie' (photo of a person icon), 'Česko-bavorský archivní průvodce' (blue arrow icon), 'Úřední knihy (pozemkové knihy, urbáře...)' (open book icon), 'Periodika' (newspaper icon), 'Seznamy lázeňských hostů' (mug icon), 'Sčítání lidu' (group of people icon), and 'Přihlášky k pobytu' (box icon).

- Velká škála různých typů dokumentů (tištěných i ručně psaných)
- Potřeba řady odlišných algoritmů pro zpracování dat
- Důraz na zpracování *Kronik a Periodik*

- Program přeshraniční spolupráce Česká republika – Svobodný stát Bavorsko Cíl EÚS 2014 -2020
- Pět partnerů: ZČU (vedoucí partner), SOA, ZIP, GDA, FAU
- Spojení čtyř výzkumných oblastí: informatika, historie, stavebně historický průzkum a archeologie
- Doba trvání: 3 roky
- Celkový objem: 964.140,- € (dotace EU 819.519,- €)
- Pět dílčích cílů (pracovních balíčků)

- První spojení historických map a plánů ke společným česko-bavorským dějinám do roku 1918
- Výběr a případné restaurování relevantních map
- Digitalizace a georeferencování map
- Indexace a doplnění metadaty
- Online prezentace na portálu

Příklad historické mapy



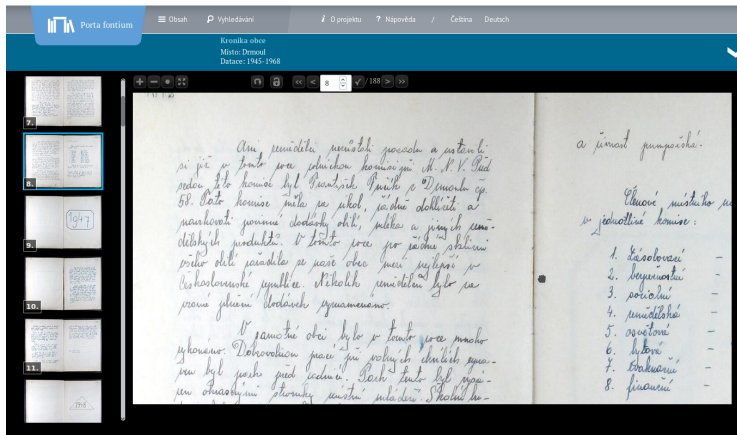
- Mapa obce Tachau

- Automatické doplnění lokalizačních údajů
- Ruční doplnění prázdných, korekce chybných gps souřadnic
- Vytvoření mapového prohlížeče
 - Geografická prezentace dokumentů včetně nově zpřístupněných historických map
 - Různé mapové vrstvy
 - Obousměrná vazba mezi dokumenty a jejich lokalizací

WP3 - Přehledná prezentace a efektivní vyhledávání v rukopisných textech

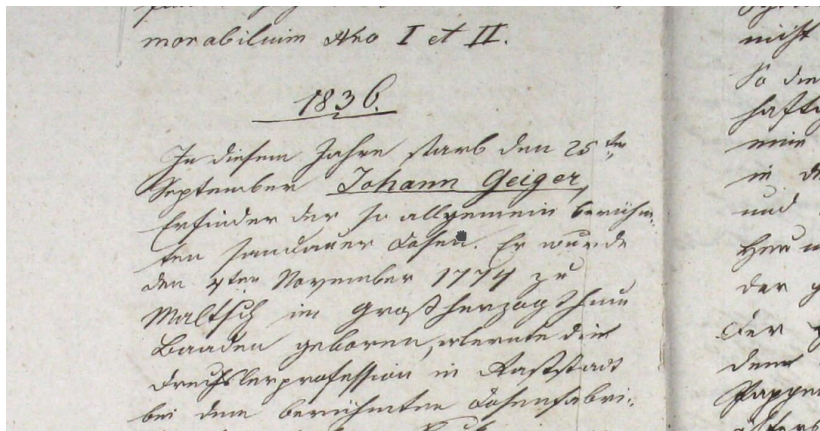
- Segmentace stránek (**text**, obrázky, pozadí, atd.)
- Související anotace
- Vyhledávání podle ručně extrahovaných částí psaného textu (QBE)
- Automatické určení pisatele a dalších metadat
- → Nový moderní prohlížeč obrazových dokumentů

Příklad rukopisného textu v češtině



- Kronika obce Dmouš z let 1945 - 1968

Příklad rukopisného textu v němčině



- Kronika fary v Dolním Žandově z let 1836 - 1946

WP4 - Inteligentní full-textové zpřístupnění dvojjazyčných tištěných dokumentů

- Segmentace tištěných dokumentů
- OCR - převod do textové podoby
- Zpracování / analýza metodami automatického zpracování přirozeného jazyka (NLP)
 - Vícejazyčné zpřístupnění textu
 - Normalizace slov
 - Sémantická analýza (určení významu textu) → přesnější vyhledávání
 - Automatické určení pojmenovaných entit a klíčových slov → nová metadata pro vyhledávání

Příklad jednoduchého tištěného dokumentu

Porta fontium

Obsah Vyhledávání O projektu Nápověda / Čeština Deutsch

Kronika města
Místo: Aš
Datace: 1878-1948

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

Kronik Asch 1.

Streik 1878. Als Beweis für die revolutionäre Tätigkeit der Ascher Arbeiter, folgendes. Bereits im Jahre 1878 erhob sich die Belegschaft der ersten Firma Weberei Christian Geipel, heutige Ohara, gegen die Hungerlöhne durch die einzige Möglichkeit Streik versuchten sie ihre Lage zu verbessern wurden jedoch durch Einsatz von Militär blutig unterdrückt. Todesfälle und einige Verwundete war das Ergebnis.

Vorkämpfer. In den 90ziger Jahren wurden unsere alten Vorkämpfer auf Grund ihrer politischen Tätigkeit von der Gendarmerie in Haft genommen, in Ketten gelegt und nach Pankraz überführt.

Streik Pa Klaubert. Im Jahre 1903 stellte die Belegschaft der Pa J.O.Klaubert Weberei Lohnforderungen. Die gestellten Lohnforderungen wurden von d. Pa rundweg abgelehnt. Die Belegschaft trat in Streik. Daraufhin wurde sie von der Pa Monate ausgesperrt. Die Streikleitung wurde gemässregelt und war gezwungen im Ausland Arbeit zu suchen.

Maifeier 1906. 1906 war die Ascher Arbeiterschaft bereits soweit, dass sie bereits rein politisch auftrat. In allen Betrieben wurde zur 1 Maifeier aufgerufen. Allen jenen die den Ruf Folge leisteten, wurden an folgenden Tag, von seiten der Unternehmer mit den Argument begegnet, dass sie dort bleiben können wo sie gestern waren, nämlich auf der Strasse.

Streik Pa Lindemann. Im gleichen Jahre konnte die Belegschaft der Firma Lindemann durch Streik ihre Lohnforderung bis zu 80% durchsetzen. Ein Beispiel dass sofort auf die

- Kronika města Aš z let 1878 - 1948

- Seznámení maximálního počtu potenciálních uživatelů s připravovanými možnostmi portálu „Porta Fontium“
- Získání zpětné vazby a požadavků pro jeho efektivní provoz a další rozšíření
- Zohlednění požadavků uživatelů během řešení projektu
- *Diskuzní workshop - 8. až 9. listopadu 2018, Klášter Teplá²*

²více viz <http://www.portafontium.eu/sites/default/files/workshop-2018-11-pozvanka.pdf>

- Základní výsledky
 - Segmentace
 - OCR
 - Normalizace slov
 - Dvojjazyčné reprezentace
- **Poděkování:** Tento článek vznikl díky podpoře projektu č. 211: Moderní zpřístupnění historických pramenů z programu přeshraniční spolupráce Česká republika – Svobodný stát Bavorsko Cíl EÚS 2014 -2020.