

Jak vytěžovat data pomocí umělé inteligence?

**Aleš Vejvoda, MPSV
Denis Belucz, KPMG
Dominik Mathauser, Microsoft
ISSS 2023**



Aktuální stav & motivace změny

1

Velká četnost opakujících se rutinních úkolů – přepsání dat do systému

2

Chybí procesní zpracování a rozdílný přístup na jednotlivých pobočkách

3

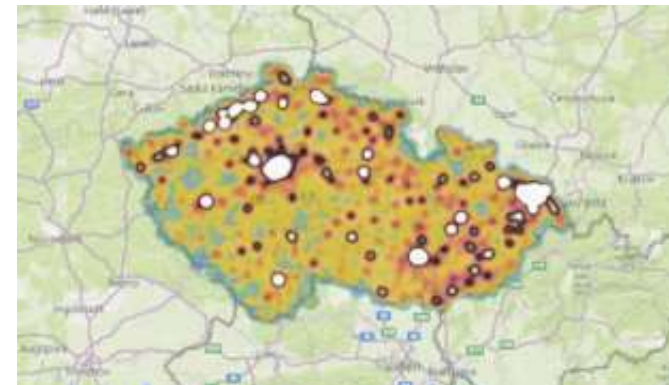
Různě vytížené regiony

4

Zavádíme model Specializovaného pracoviště - BO pro centralizované zpracování rutinních úkolů

5

Specializovaná technická komponenta, time to market



Jaké jsou přínosy?



Digitalizace zpracování ručně vyplněných formulářů

KPMG

Dodavatel řešení digitalizace zpracování



Microsoft

Dodavatel technologie AI



Úspěšnost vytěžení je 91+% (žádost o příspěvek na bydlení)



Identifikace typu podaného formuláře



Kontrola úplnosti podání žádosti (všechny listy hlavního formuláře a všechny přílohy)



Průkazná evidence podání pomocí unikátního identifikátoru (garance integrity dat)



Vytěžení obsahu polí - psacím nebo hůlkovým písmem, v češtině (možné i jazykové mutace)



Kontrola správnosti údajů (možnost manuální korekce nebo doplnění úřední osobou)



Řešení je vytvořeno jako úplná, autonomní, bezpečná služba v cloudu

Vkládání oskenovaných formulářů do vytěžovací linky

The screenshot shows a web application interface titled "Digitalizace formulářů". At the top, there are navigation tabs: "Úvod", "Vkládání oskenovaných formulářů", "Vytěžování údajů z oskenovaných formulářů", and "Komplexní vyřizování žádostí". The main content area is divided into several sections:


- Identifikátor podání zastavil úřední osobou:** A text input field containing "ISSS úřadna 4" and an edit icon.
- Agenda:** A dropdown menu showing "Příspěvek na bydlení".
- Nezaměstnané osoby z hlediska (dávky) sociální pomoci:** A note indicating that the system will select the appropriate category based on the selected category.
- Tabulky:** A table with four columns: "Vložení souboru", "Doposledný číselný kód", "Doposledný vzor formuláře", and "Tabulky".

Two rows of data are visible in the table:

Vložení souboru	Doposledný číselný kód	Doposledný vzor formuláře	Tabulky
00206B3D2C942305091234_0004.jpg		F100-4	[Thumbnail of a form]
00206B3D2C942305091234_0002.jpg		F100-2	[Thumbnail of a form]

At the bottom left, there is a yellow button labeled "Všechny formuláře".

Uložené skeny stránek formulářů k vytěžení

Úvod
Vkládání oskenovaných formulářů ▾
Vytěžování uložených skenů formulářů
Kompletace vytěžených podání
⚙️


Uložené skeny stránek

↻ Obnovit

Úřad/Pracoviště	Agenda	ID naskenované dávky	ID přidělený	Kód formuláře	Čarový kód	Datum a čas	ID uživatele
ÚP P1	Příspěvek na bydlení	d99baa02-b897-45b2-ac9b-a6df234e4f90	ISSS ukázka 002	F100-2		2023-04-19T14:24	099e307b-8726-4236-b7d4-a21a7992fc0e
ÚP P1	Příspěvek na bydlení	d99baa02-b897-45b2-ac9b-a6df234e4f90	ISSS ukázka 002	F100-4		2023-04-19T14:24	099e307b-8726-4236-b7d4-a21a7992fc0e
ÚP P1	Příspěvek na bydlení	238d8257-a2e5-44c0-9eab-5a6fb35a42c7	ISSS ukázka 4	F100-4		2023-04-19T17:32	099e307b-8726-4236-b7d4-a21a7992fc0e
ÚP P1	Příspěvek na bydlení	238d8257-a2e5-44c0-9eab-5a6fb35a42c7	ISSS ukázka 4	F100-3		2023-04-19T17:32	099e307b-8726-4236-b7d4-a21a7992fc0e
ÚP P1	Příspěvek na bydlení	238d8257-a2e5-44c0-9eab-5a6fb35a42c7	ISSS ukázka 4	F100-2		2023-04-19T17:32	099e307b-8726-4236-b7d4-a21a7992fc0e
ÚP P1	Příspěvek na bydlení	238d8257-a2e5-44c0-9eab-5a6fb35a42c7	ISSS ukázka 4	F100-1		2023-04-19T17:32	099e307b-8726-4236-b7d4-a21a7992fc0e

Zobrazeno 1 až 10 z/ze 35 položek

<
1
2
3
...
5
>

Náhled oskenované stránky formuláře

Úvod Vkládání oskenovaných formulářů Vytěžování uložených skenů formulářů Kompletace vytěžených podání

Vybraná vytěžená stránka formuláře

Identifikátor podání: **c401eb2e-3af7-4e80-8948-3adcf76e6889**
Čarový kód nalepený na stránce formuláře: **885KL904HJ**
QR kód s identifikátorem formuláře a číslem stránky: **F100-1**

Oskenoaná stránka formuláře Vytěžené údaje připravené ke kontrole a opravě Náhled na sken a na vytěžené údaje

 **ÚŘAD PRÁCE ČR**
SSP
Tiskopis prosím vyplňte čitelně

podací razítko OSÚ S 15

Žádost o příspěvek na bydlení

A. Žadatel o příspěvek:

Kontrola vytěžených dat

Úvod Vkládání oskenovaných formulářů Vytěžování uložených skenů formulářů Kompletace vytěžených podání

Vybraná vytěžená stránka formuláře

Identifikátor podání: **c401eb2e-3af7-4e80-8948-3adcf76e6889**
 Čárový kód nalepený na stránce formuláře: **885KL904HJ**
 QR kód s identifikátorem formuláře a číslem stránky: **F100-1**

Oskanovaná stránka formuláře **Vytěžené údaje připravené ke kontrole a opravě** Náhled na sken a na vytěžené údaje

Název pole na formuláři	Vytěžené údaje	Vytěžené údaje	Sken vytěžovaného pole
Příjmení	NOVÁČEK	● Ověřeno	NOVÁČEK
Jméno	JIŘÍ	● Ověřeno	JIRI
Titul před		● Ověřeno	
Titul za		● Ověřeno	
Rodné číslo v ČR	490622 1014	● Chybné	490622 1014
Trvalý pobyt - Část obce	VYSOČANY	● Ověřeno	VYSOCANY
Trvalý pobyt - Část obce	KBELSKÁ	● Upozornění	KBELSKA

Případná manuální korekce

Úvod Vkládání oskenovaných formulářů Vytěžování uložených skenů formulářů Kompletace vytěžených podání

< Vybraná vytěžená stránka formuláře

Identifikátor podání: c401eb2e-3af7-4e50-8948-3adcf75e6889
 Čárový kód nalepený na stránce formuláře: 835KL904HJ
 QR kód s identifikátorem formuláře a číslem stránky: F100-1

Oskenovaná stránka formuláře **Vytěžené údaje připravené ke kontrole a opravě** Náhled na sken a na vytěžené údaje

Název pole na formuláři	Vytěžené údaje	Vytěžené údaje	Sken vytěžovaného pole
Příjmení	NOVÁČEK	● Ověřeno	NOVÁČEK
Jméno	JIRÍ	● Ověřeno	JIRÍ
Titul před		● Ověřeno	
Titul za		● Ověřeno	
Rodné číslo v ČR	490622 1014	● Chybně	490622 1014
Trvalý pobyt - Část obce	VYSOČANY	● Ověřeno	VYSOČANY
Trvalý pobyt - Část obce	KBELSKÁ	● Upozornění	KBELSKA

Dokončení manuální korekce

Úvod Vkládání okenovaných formulářů Vyřizování uložených skenů formulářů Kompletace vytěžených podání

< Vybraná vytěžená stránka formuláře

Identifikátor podání: c401eb2e-3af7-4e80-8948-3adcf76e6889
 Čárový kód nalepený na stránce formuláře: 885KL504HJ
 QR kód a identifikátorem formuláře a číslem stránky: F100-1

Oskanovaná stránka formuláře **Vytěžené údaje připravené ke kontrole a opravě** Náhled na sken a na vytěžené údaje

Název pole na formuláři	Vytěžené údaje	Vytěžené údaje	Sken vytěžovaného pole
Vlastník bytu - Příjmení	KADRNOŽKOVÁ	Ověřeno	KADRNOŽKOVA
Vlastník bytu - Jméno	VLADIMÍRA	Ověřeno	VLADIMI' RA
Vlastník bytu - Rodné číslo v ČR	540203/0721	Ověřeno	540203/0721
Vlastník bytu - Rodné příjmení	VYCHODILOVÁ	Ověřeno	VYCHODILOVA
Vlastník bytu - Titul před		Ověřeno	
Vlastník bytu - Titul za		Ověřeno	

Uložit skenované údaje

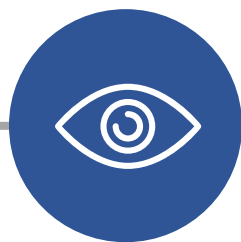
Kontrola uložených dat a předání do agendové aplikace

The screenshot displays a web application interface with a dark grey header. The header contains a breadcrumb trail: 'Úvod' (highlighted in yellow), 'Vkládání oskenovaných formulářů', 'Vytěžování uložených skenů formulářů', and 'Kompletace vytěžených podání'. On the right side of the header, there is a gear icon for settings and a circular profile picture of a man. The main content area is light grey and features a white card with the title 'Výsledný stav'. Below the title, the text reads: 'Zkontrolovaná a případně opravená data vytěžené stránky formuláře byla uložena do úložiště a je možné je použít k dalšímu zpracování.' At the bottom of the card is a prominent yellow button with the text 'Zpět na seznam'.

Azure AI/Kognitivní služby



Předem natrénované a přizpůsobitelné modely dle vašich dat



VISION

Identifikace a analýza obsahu v obrázcích, videích a textu



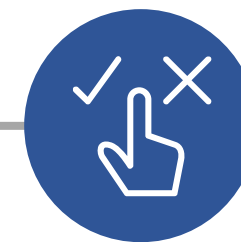
SPEECH

Integrace zpracování řeči do aplikací a služeb



LANGUAGE

Extrahování významu z nestrukturovaného textu



DECISION

Přijímejte rozhodnutí rychleji

Form Recognizer - vlastní formulář

Natrénovaný model vytvořený na míru tak, aby extrahoval všechna požadovaná pole formuláře



Stačí 5 variant dokumentu

Trénování modelu během několika minut

Trénování modelu k extrakci:

- Pole formuláře (včetně zaškrťovacích políček)
- Tabulky
- Pole podpisu

The image shows a 'W-9 Request for Taxpayer Identification Number and Certification' form. A red bounding box highlights the 'Name' field, which contains 'John Doe'. To the right of the form, a list of extracted fields is shown with corresponding colored circles:

- Name: John Doe
- BusinessName: Arctex Inc.
- TaxClassInd: unselected
- TaxClassCCorp: unselected
- TaxClassSCorp: unselected
- TaxClassPart: unselected
- TaxClassTrust: unselected
- LLC: selected
- Other: unselected
- LLCType: S
- 4-Exempt: S
- 4-FATCA: unselected