

# Čtenář zamčených dokumentů

Před více než rokem byla v Chipu recenzována verze 6 programu FineReader. Nyní jsme vyzkoušeli nejnovější verzi 7.0. Jedním z nejdůležitějších kroků vpřed je schopnost exportovat formát PDF (verze 6.0 ho uměla pouze číst). Znamená to, že dokument PDF, který má uzamčenu možnost kopírování a úprav, lze touto aplikací načíst, integrovaným editorem upravit a exportovat zpět do PDF.

## PŘEHLED FUNKCÍ

### NOVINKY OPROTÍ VERZI 6.0.

Byly přidány specializované anglické a německé slovníky obsahující, právníkou a lékařskou terminologii.

Byla zlepšena přesnost rozpoznávání program nyní lépe analyzuje a rozpoznává dokumenty s komplexním rozložením.

Práce s dokumenty PDF - tento formát se nyní nejen čte, ale výsledek rozpoznávacího procesu (po eventuálních úpravách) se také ukládá.

Podpora XML a integrace s MS Office 2003.

Možnost ukládání ve formátu PowerPoint - prezentace lze nyní připravovat z nejrůznějších naskenovaných nebo načtených podkladů přímo ve FineReaderu.

Zjednodušené ovládání - uživatel zvyklý na MS Office může postupovat navyklym způsobem.

Nástroj pro rozdělení obrázku - umožní rozdělit obrázek do několika částí a uložit každou část jako samostatnou stranu.

Vyhledávání s morfologickou podporou každá dávka vytvořená ve FineReaderu může být použita jako malá databáze s vyhledávacími fulltextovými schopnostmi.

Byla zlepšena možnost ovládání ukládaných obrázků - při ukládání do výstupního souboru lze buď zachovat rozlišení předlohy, nebo lze nastavit rozlišení v dpi nižší, či je možné zvolit kompresi JPEG.

### DALŠÍ DŮLEŽITÉ FUNKCE

Vstupní formáty - vedle přímého skenování jsou jako vstup do OCR k dispozici formáty BMP, PCX a DCX, JPEG, PNG, TIFF a PDF.

Rozpoznávání vícejazyčných dokumentů buď lze pro každý blok textu zadat jiný jazyk, nebo lze celý dokument deklarovat ve více jazycích.

Kontrola a úpravy rozpoznávaného dokumentu - v rozpoznávaném textu jsou nerozpoznaná slova nenalezená ve slovníku daného jazyka vyznačena odlišnými barvami. Při kontrole pracujeme klasickým způsobem - buď korektor nabízí možnosti, nebo lze slovo přidat do slovníku.

Pro úpravy textu před odesláním do cílového formátu (DOC, RTF, HTML a dalších) jsou k dispozici funkce známé z většiny editorů, vyvolávané ikonami obdobnými s MS Office lze slučovat/dělit buňky, sloupce, řádky.

Šablony - při práci s větším množstvím stránek se stejným, ale nestandardním uspořádáním lze vytvořit šablonu a tu uložit pod jménem.

Rozpoznávané jazyky - aplikace ve verzi EU podporuje celkem 122 jazyků, ve verzi Cyrilic 177. Podporu korektoru má 34 jazyků.

FormFiller - zcela samostatná aplikace, která využívá snímání předtištěných formulářů nebo formulářů stažených z internetu (formáty BMP, TIF, JPG, PCX, PNG, DCX). Může sloužit buď k vyplňování a k následnému tisku do předtištěných formulářů, nebo ke kompletnímu tisku včetně rastru formuláře.

## ZKUŠENOSTI Z TESTOVÁNÍ

Testování bylo zaměřeno spíše na problémy při úpravách načteného dokumentu a při jeho exportu. Samotné rozpoznávání probíhá díky dobře zpracovaným algoritmům hladce. Dokonce i barevný tisk na podkladě jiné barvy - tradičně slabé místo OCR aplikací - nedělal potíže.

Načítání tabulek - automatické rozpoznání typu bloku je vhodné jen pro pravidelné tabulky, které mají ve všech sloupcích stejný počet řádek a ve všech řádcích stejný počet sloupců. Jakmile jsou v tabulce použity spojené buňky, je nutné zadat dělení manuálně.

Načítání textů nepravidelně rozmístěných na stránce okolo obrázku je častým případem v technických manuálech. Aplikace některé texty interpretuje jako sloupcový text s proměnnou výškou

řádky, jiné jako textová pole. Výhodnější pro následné zpracování by bylo soustavné použití textových polí.

Načítání obrázků nepravidelně rozmístěných v textové tabulce - funguje většinou bez problémů.

Přidávání slov do slovníku - velmi dobře zpracováno pro angličtinu - program se táže na slovní druh, způsob psaní velkého písmene. Do českého slovníku přidává pouze tvary, bez jakýchkoliv dotazů - pro tvorbu slovníku používá zřejmě odlišných algoritmů.

## **ZÁVĚR**

FineReader 7.0 není už pouze OCR aplikace. Široká paleta vstupních a výstupních formátů spolu s integrovaným editorem z něj dělají univerzální nástroj na přípravu dokumentu. Českého uživatele bude produkt lákat především širokou paletou korektorů pro všechny jazyky států obklopujících ČR a pro mnoho dalších jazyků. Také funkce pro rozeznávání znaků psacího stroje a mozaikové tiskárny (lze použít i na kvalitnější faxy) jsou velmi příjemné.

*Ing. Miroslav Herold, CSc.*

### **FINEREADER 7.0 PROFESSIONAL**

OCR software.

Vyrábí ABBYY (Rusko)

Poskytl Nupseso CZ, [www.nupseso.cz](http://www.nupseso.cz)

Cena (bez DPH)

FineReader 7.0 Professional 4980 Kč,

FineReader 7.0 Corporate Edition 10 100 Kč