

Indscanning af russiske tekster

Indscanning af russiske tekster

Kort sagt er formålet med OCR-scanning at forvandle papirtekster til elektroniske tekster med alle de kendte fordele de har.

En lang forklaring kan du læse her på Internettet:

http://www.pcwebopedia.com/TERM/o/optical_character_recognition.html

Da vi jo har problemet med de russiske bogstaver, følger her nogle gode råd om ”værktøjet”. Hvad man skal bruge?

- En scanner: Ofte vil en ret billig standardscanner være absolut brugelig
- Et OCR-program.

Det skal straks siges, at de OCR-programmer, der følger med i købet af en scanner, ikke kan bruges til russisk. Det er som regel light-versioner eller ældre versioner af kommercielle programmer. De læser for unøjagtigt.

Kriterierne for anskaffelse er flg.

- Nøjagtighed i karaktergenkendelsen
- Brugervenlighed
- Pris

Af den håndfuld programmer, der kan komme på tale, vil vi kort karakterisere de bedst egnede.

FineReader fra firmaet Abby er nok det bedste valg. Det er som flere andre gode OCR-programmer designet af russiske programmører. Firmaet har til huse i Moskva (www.abby.ru) og har også en repræsentation i USA (www.abby.com). I Danmark kan det købes hos Prosoft i Ringsted (www.prosoft.dk).

Firmaet fører en lidt forvirrende versions- og navngivningspolitik. Derfor en lille vejledning.

- Den næstnyeste version er 4.0. Det er den, vi vil anbefale, pga. kvalitets/prisforholdet. Den findes i en version Standard, som er den, russisklæreren bør anskaffe. Den koster under 1000 kr. og forstår over 150 sprog, deriblandt russisk og dansk. Versionen PRO benytter nøjagtig samme OCR-engine og er altså

ikke bedre. Den har nogle faciliteter, som er uden betydning i vores sammenhæng.

- Den nyeste version hedder 5.0, har en forbedret OCR-engine, men er et problematisk køb. For det første har de smarte Abbyfolk splittet produktet i 2:

En ”EU-udgave”, som ikke kan læse russisk, og ikke bringes til at gøre det – og en Cyrillic Plus, som kan læse slaviske sprog, men som til gengæld er dyrere.

Derudover findes produktet i tre udstyrsvariationer: FR Home, FR PRO og FR Office. Home-udgaven koster \$99, Office-udgaven \$399. Prisen for de kyrillisk-kapable udgaver er ikke oplyst, men skal indhentes hos Abby. (Forklaringen på alle disse narrestreger er vist at firmaet er et klassisk mål for pirateri og crackning.)

Et andet bud er det gammelkendte **Cuneiform**. Det nyeste produkt hedder Cuneiform99 eller Exact99 og er altså tre år gammelt. Det er lidt betænkeligt.

Men programmet er absolut anvendeligt og prisen er sympatisk \$69.

Kan downloades fra webstedet www.ocr.com.

Tastevejledninger til programmerne er overflødige. Brug programmernes Wizards.