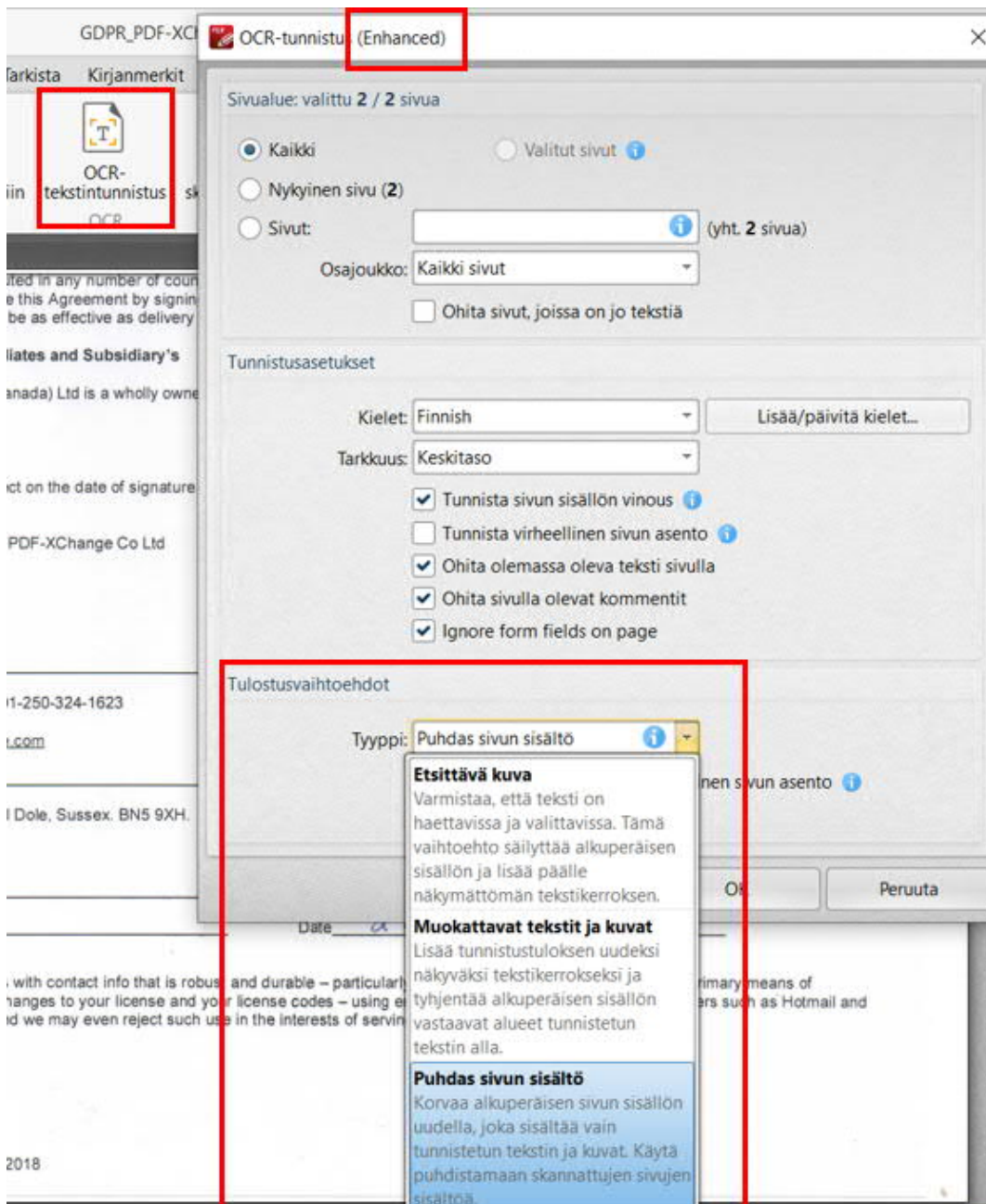


PDF-XCHANGE PRO - ENHANCED OCR PLUGIN (EOCR) PIKAOHJE

Jos sinulla on tarve muokata skannattua dokumenttia, tulee dokumentille tehdä OCR-tekstintunnistus. Tämä sen vuoksi, että skannattu dokumentti on kuva ja sitä ei voi muokata PDF-XChange Editorin Muokkaa-työkalulla suoraan. OCR- tekstintunnistus löytyy PDF-XChange Editorin ylävalikosta Muunna > OCR-tekstintunnistus.

Uusi tehostettu OCR tunnistus Enhanced OCR Plugin (lyhennettynä EOCR) mahdollistaa tulostusvaihtoehdoissa valinnan 'Etsittävä kuva', 'Muokattavat tekstit ja kuvat' sekä 'Puhdas sivun sisältö' välillä



The screenshot shows the 'OCR-tunnistus (Enhanced)' dialog box in PDF-XChange Pro. The window title is 'GDPR_PDF-XC... OCR-tunnistus (Enhanced)'. The 'Sivualue: valittu 2 / 2 sivua' (Page range: selected 2 / 2 pages) is shown. The 'Kaikki' (All) radio button is selected. The 'Osajoukko: Kaikki sivut' (Subset: All pages) dropdown is set to 'Kaikki sivut'. The 'Tunnistusasetukset' (Recognition Settings) section includes: 'Kielet: Finnish', 'Tarkkuus: Keskitaso', and several checked options: 'Tunnista sivun sisällön vinous', 'Ohita olemassa oleva teksti sivulla', 'Ohita sivulla olevat kommentit', and 'Ignore form fields on page'. The 'Tulostusvaihtoehdot' (Output Options) section is highlighted with a red box, showing 'Tyypit: Puhdas sivun sisältö'. A tooltip is displayed over the 'Puhdas sivun sisältö' option, containing the following text:

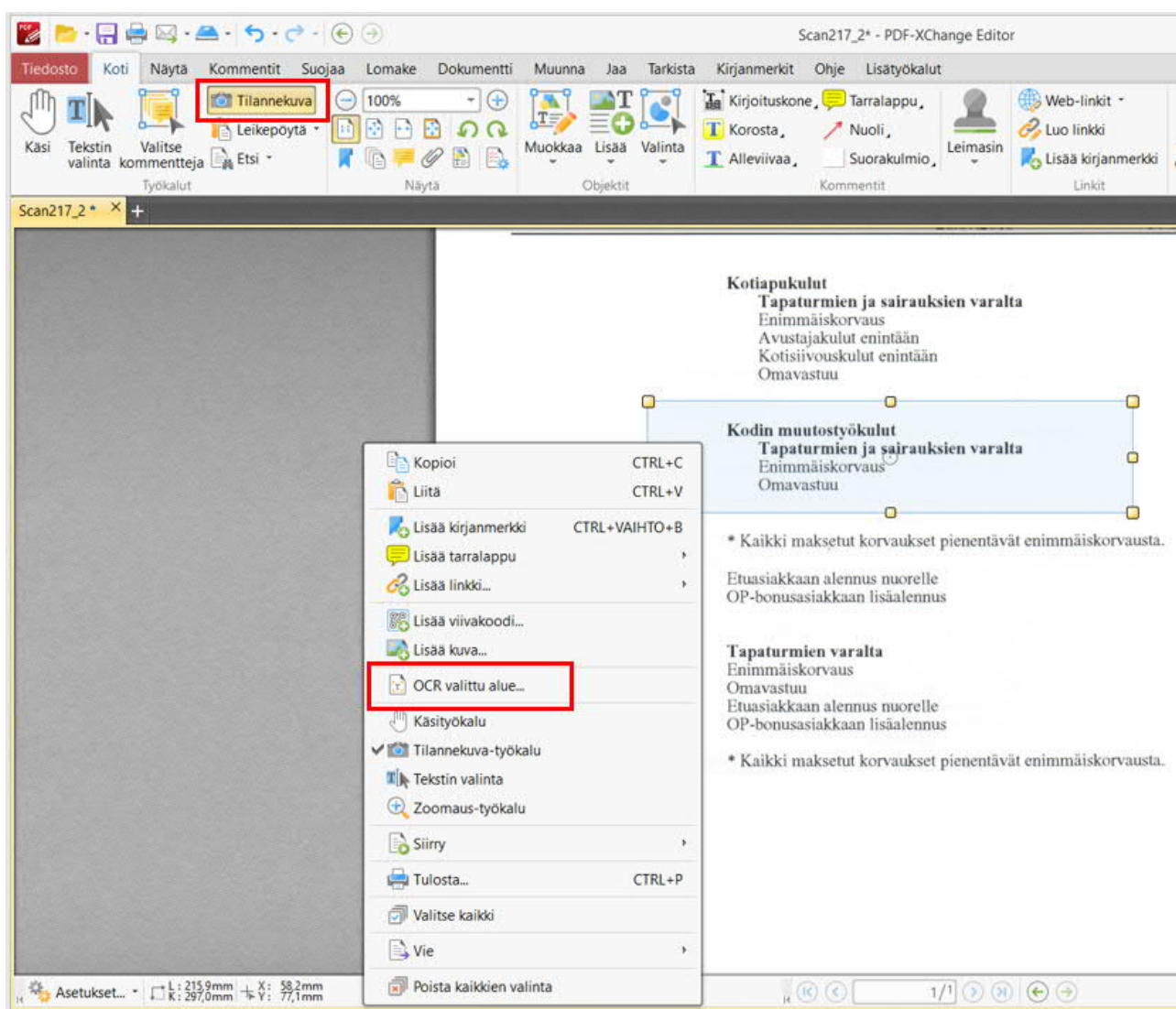
- Etsittävä kuva**
Varmistaa, että teksti on haettavissa ja valittavissa. Tämä vaihtoehto säilyttää alkuperäisen sisällön ja lisää päälle näkymättömän tekstikerroksen.
- Muokattavat tekstit ja kuvat**
Lisää tunnistustuloksen uudeksi näkyväksi tekstikerrokseksi ja tyhjentää alkuperäisen sisällön vastaavat alueet tunnistetun tekstin alla.
- Puhdas sivun sisältö**
Korvaa alkuperäisen sivun sisällön uudella, joka sisältää vain tunnistetun tekstin ja kuvat. Käytä puhdistamaan skannattujen sivujen sisältöä.

'Etsittävä kuva' on vaihtoehto, joka löytyy oletus OCR-versiossa, eli se luo alkuperäisen kuvan päälle näkymättömän tekstikerroksen. Mikäli tarve on vain maalata ja kopioida dokumentista tekstiä, tämä vaihtoehto riittää.

'Muokattavat tekstit ja kuvat' lisää tunnistetut objektit näkyväksi kerrokseksi ja poistaa niitä vastaavat alkuperäiset kohteet tunnistetun tekstin alta.

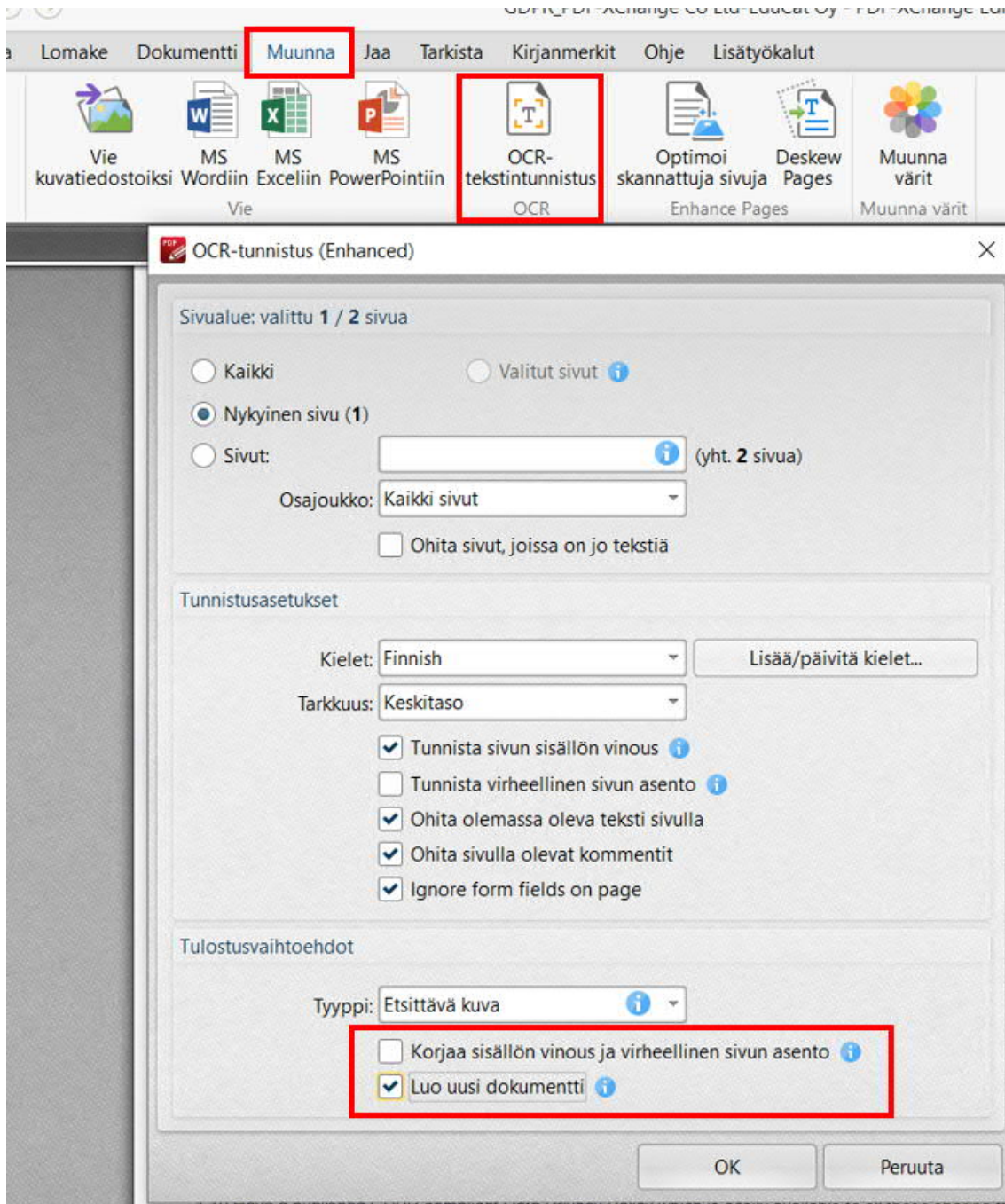
'Puhdas sivun sisältö' korvaa alkuperäisen sisällön tunnistetulla sisällöllä. Se sisältää pelkistettynä vain tunnistetut tekstit ja kuvat. Jos luot tuloksesta uuden dokumentin, säilyy alkuperäinen kuvadokumentti ennallaan ja saat ilman lisätoimenpiteitä uuden tekstityökaluilla muotoiltavan ja muokkaustyökalulla muokattavissa olevan tekstidokumentin.

Tehostetussa OCR-tekstintunnistuksessa voit tehdä tunnistuksen myös pelkästään osalle dokumentista. Valitse ensin tunnistettava alue dokumentista Tilannekuva-työkalulla. Valinnan päällä klikkaa hiiren kakkospainiketta ja valitse 'OCR valittu alue..'



Tällöin varmistut siitä, että kaikki muu alkuperäisen dokumentin kuvat yms. säilyvät koskemattomina.

Tallenna OCR tunnistettu dokumentti 'Tallenna nimellä...' valinnalla jolloin alkuperäinen skannattu dokumentti jää talteen myös. Vaihtoehtoisesti voit OCR tekstintunnistussikkunassa ruksata valmiiksi kohdan 'Luo uusi dokumentti' jolloin OCR tekstintunnistus luo uuden dokumentin kopiosta ja alkuperäinen jää muuttamattomana avoimeksi PDF-XChange Editoriin:



'Korjaa sisällön vinous ja virheellinen sivun asento' korjaa automaattisesti dokumentin suoraan mikäli se on skannattu hieman vinoon.



Tässä alla vielä vertailuja mitä Enhanced OCR Plugin tuo tullessaan verrattuna ohjelmaan kuuluvaan perus OCR toimintoon:

Enhanced OCR Plugin:in merkkien tunnistus on tarkempaa ja merkittävästi **nopeampaa** kuin ohjelmaan kuuluvalla oletus OCR toiminnolla – yleensä kahdesta kolmeen kertaan nopeampaa.

Enhanced OCR Plugin tunnistaa tekstin tyylielementit, kuten **kirjasintyytin, värin, lihavoinnin, kursivoinnin, yliviivauksen** sekä **alleviivauksen**.

Enhanced OCR Plugin tunnistaa lähdedokumentista **taulukot** ja lisää ne tekstin lisäksi uuteen dokumenttiin.

Enhanced OCR Plugin voi luoda PDF tiedostoja valinnalla "**Muokattavat tekstit ja kuvat**" - joka korvaa kuvapohjaisen tekstin lähdedokumentissa muokattavalla tekstillä.

Enhanced OCR Plugin voi luoda PDF tiedostoja valinnalla "**Puhdas sivun sisältö**" – joka korvaa alkuperäisen sivun sisällön uudella sisällöllä, sisältäen vain OCR prosessin aikana tunnistetun tekstin ja kuvat.

Enhanced OCR Plugin mahdollistaa useille sivuille samanaikaisia OCR tehtäviä. Käyttäjät, joilla on koneita useammalla prosessorilla/ytimellä, säästävät aikaa jonka OCR ajo vie. Oletus OCR:ään nähden kaksin verroin nopeampi.