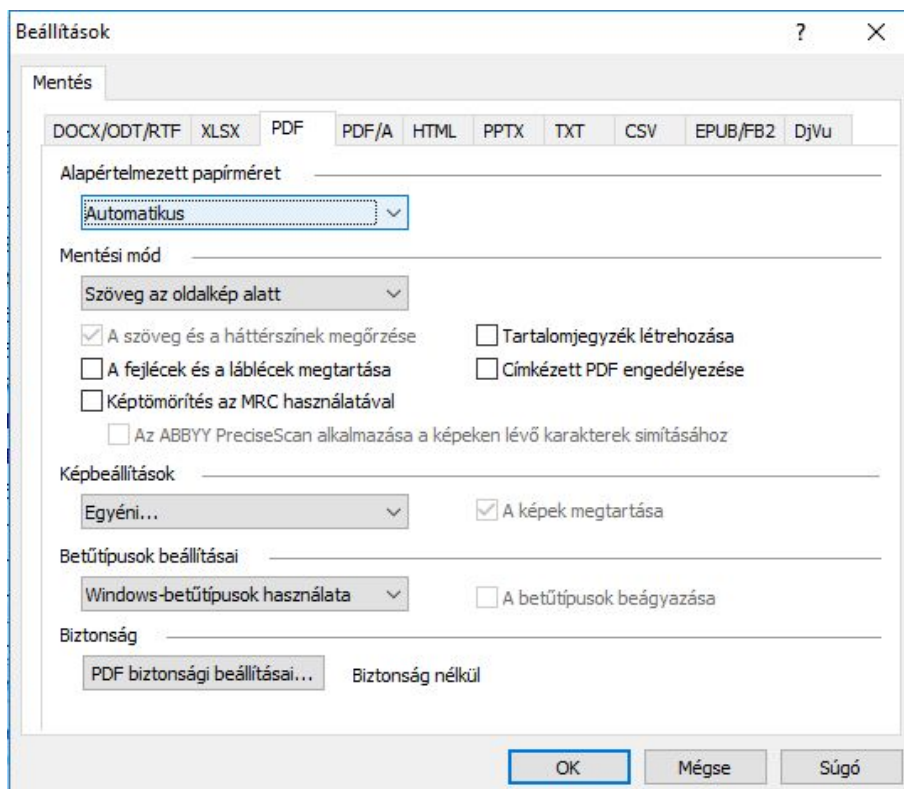


Könyvek digitalizálása a MEK-ben

A digitalizált könyvek elsődleges szolgáltatási formája a MEK-ben a kétrétegű PDF. Ez olyan PDF, amelyben a könyvoldal képe mögött helyezkedik el az OCR-es szöveg, ami kimásolható és kereshető - a szolgáltatott változat esetében mind a Google és más keresőszolgáltatásokon keresztül, mind a megnyitott PDF-ben. Mivel hibátlan OCR a gyakorlatban nem létezik, a nem kereshető, ámde látható és hű képrétegtől lesz hiteles a digitalizált könyv.



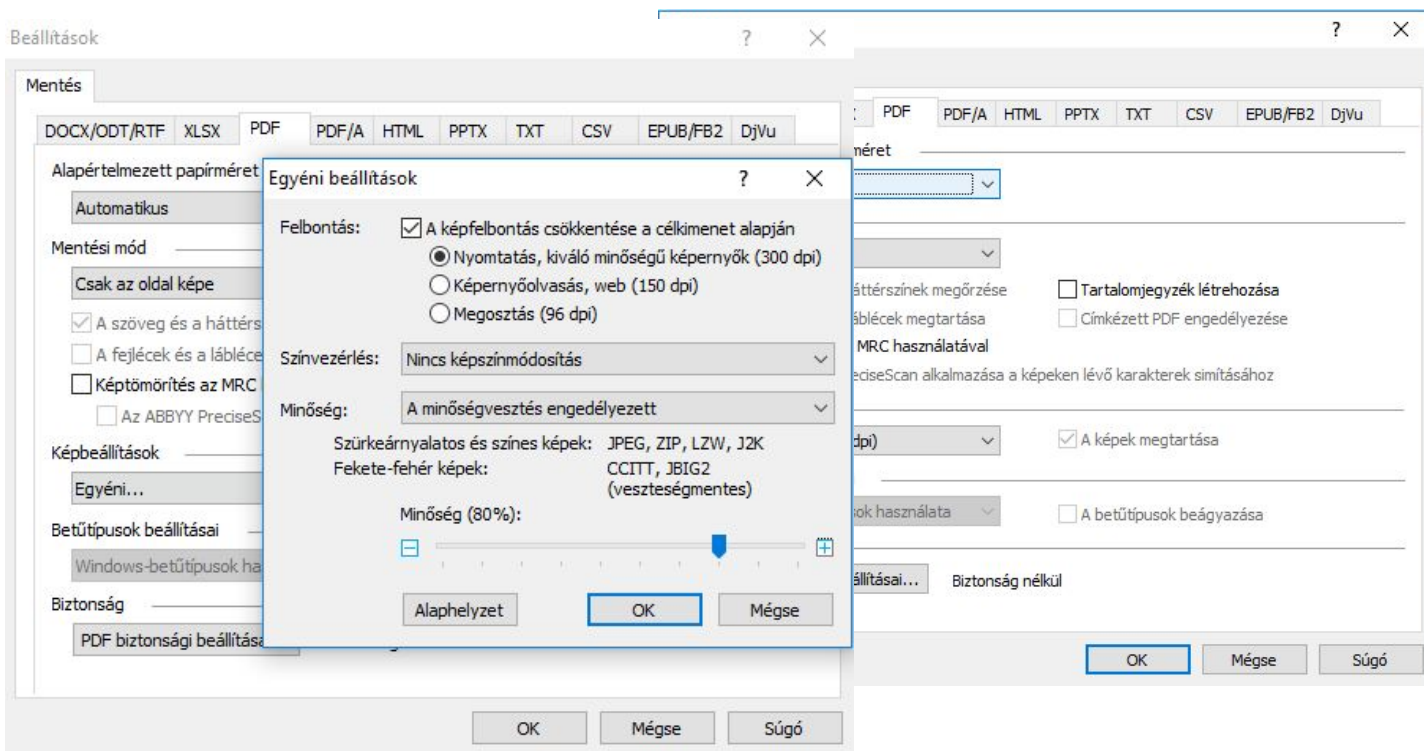
A kétrétegű PDF-hez a mentési beállítás: „Szöveg az oldalkép alatt”.

Az OCR-es PDF elkészítése előtt a következő szempontok figyelembevételével ajánlott megválasztani - szükség esetén néhány oldalon kikísérletezni - a követendő eljárásokat.

A szolgáltatott PDF-ek megengedhető legnagyobb mérete fájlként 40 MB. Ha a fájl méret / szkennelt terület arány nem kifejezetten pazarló, vagyis alatta marad a 3 MB/m² (40 MB/300 B5 oldal vagy 40 MB/215 A4 oldal) aránynak, akkor a terjedelmes kötetek egyben tartása érdekében az 50MB is elfogadható, alaposan indokolt esetben - pl. a kötet nehezen lenne használható, ha szétbontanánk, mert félbeszakadna egy szócikk - a 60 MB is.

Célszerű néhány oldalból próba-PDF-et készíteni - akár OCR nélkül is -, és ennek méretét nézni. A készítendő PDF mérete közelítőleg egyenes arányban növekszik az oldalszámokkal, tehát ha azt látjuk, hogy a 10 oldalas próbafájl 4 MB, akkor ebből látszik, hogy a 200 oldalas könyvből készítendő PDF az adott módszer mellett már túllépné a mérethatárt.

Első lépésben a Finereaderben válasszunk más képtömörítést a kimeneti PDF-be. Az alábbiakban azonos oldalképekből különböző JPG-%-os beállítással kapott PDF-ek méretét hasonlíthatjuk össze. A „Tömör méret - 300 DPI” beállítás méretre is, szemre is körülbelül a 30 %-nak felel meg. A JPG-tömörítés rontja a PDF oldalképeinek minőségét, ezért ajánlatos kikísérletezni azt a %-ot, amivel még a mérethatár alatt maradunk.



Mentés PDF-be 300 DPI-vel adott JPG %-kal, illetve „tömör mérettel”

70%	60%	50%	40%	30%
1138 kB	974 kB	845 kB	755 kB	569 kB

A kapott PDF mérete különböző JPG%-okkal, ugyanabból a 10 oldalból

Az MRC (= mixed raster content, vegyes raszter, lásd a fenti párbeszédpanelen) beállítás differenciál az oldal egyes részei között: amit a Finereaderben valamilyen blokkba (szöveg-, kép- vagy táblázat-) foglaltunk, abból 300 DPI, ami a blokkokon kívül helyezkedik el, abból alacsonyabb felbontású képeket készít, és ezeket helyezi el a PDF-oldalon. Ezzel jelentős megtakarítás érhető el, cserébe gondosan ügyelni kell arra, hogy semmi olyan ne maradjon ki az automatikus blokkozásból, aminek jól kell látszania - tehát például a tulajdonbélyegzők, fejlécek vagy oldalszámok sem. A Finereaderbe beolvasott összes oldalt ellenőrizni kell, és szükség esetén kézzel javítani.

Ha a nyomtatott könyvben a betűk nem elég kontrasztosak - mert pl. erősen egyenetlen felületű papírra vagy egyenetlen festékezéssel nyomták -, akkor a PDF-ben nemcsak a Finereader által elhanyagolt háttér, hanem a betűképek is elfogadhatatlanok lehetnek. Ezen a Finereader karaktersimítási beállítása hivatott javítani (lásd a fenti képeken; csak bekapcsolt MRC-nél választható): a Finereader „arra törekszik”, hogy a szövegblokkok területén a betűképek lehetőleg „vektorosan” nézzenek ki, vonalak és görbék ne szakadjanak meg. De csodákra a betűsimítás sem képes - pár oldalon ezt is ajánlatos kipróbálni. Az 1., illetve 2. **mellékletben** látható ugyanabból az oldalból a helyes és a rossz blokkrendezéssel készült PDF, illetve a karaktersimítással kapott oldal. Egyik sem szép - ilyen oldalaknál nem a vegyes raszter a megfelelő megoldás. (A melléklet oldalait érdemes legalább 200 %-os nagyításban nézni.)

A melléklet 3-5. oldalán egy kellően kontrasztos szövegű, világosabb háttérű oldalon láthatjuk a különbséget - fájl méretben és PDF-oldalkép tekintetében is. Ebben az esetben tehát sikerül a vegyes

raszterrel gyakorlatilag kifogástalan PDF-oldalképet elérni úgy, hogy a homogén oldalképű PDF-hez képest mintegy harmadára csökken a fájl méret. Említésre méltó ugyanakkor, hogy az illusztrációk vegyes rászterrel akkor is romlanak valamelyest, ha a Finereaderben képblokkba helyeztük őket (lásd melléklet 3. vs. 4. vagy 5. oldal).

Természetesen nem kell a könyv minden oldalát ugyanazzal a Finereader-beállítással PDF-be kihozni. Oldaltartományonként differenciálhatunk, például az illusztrációkra való tekintettel, majd a PDF-eket egyesíthetjük a megfelelő sorrendben. Jelentős számú üres oldal (pl. szennylapok vagy fejezetek közötti üres oldal) esetén így lehet takarékoskodni azzal, hogy ezekre alacsonyabb felbontást, illetve durva jpg-tömörítést (pl. 150 dpi és 30 %) alkalmazunk.

Fontos megjegyezni, hogy a fájl méret és a minőség kompromisszuma akkor a legelőnyösebb, ha a PDF eleve a megfelelő beállításokkal készült. A PDF-szerkesztő programok (pl. Acrobat) által kínált „optimalizálási”, „méretcsökkentési” lehetőségekkel utólag már csak rosszabb minőség árán, vagy egyáltalán nem lehet ugyanakkorára csökkenteni a fájl méretet.

1. Vegyes rászteres beállítás, nincs karaktersimítás. A szövegsorok vége kimaradt a blokkból.

majdnem lebukfencezett. Szerencsésebb társai kapták el.

És akkor, ott a vagon tetején, mindhármán csodálkozva néztek egymásra. Ezuttal nem a postakocsi volt az utolsó vagon a szerelvényen, hanem egy szalonkocsi.

A szalonkocsi nyitott ablakán keresztül mindent láttak. Egy káprázatosan szép, vöröshajú leány ült a rádiókészülék előtt és cigarettázott.

— Most mi lesz? — kérdezte az egyik cowboy.

— Mi lenne, mafla. . . — hangzott a válasz. Lepakcsoljuk! Csak nem képzeled, hogy Tom megijed ettől a vöröshajú kis dögtől!

Ketten leereszkedtek a postakocsi elejére. Úgyesen, zajtalanul. Egyikük még meg is állt ütközben és bekémlelt a postakocsi belsejébe. Előgödten vigyorgott. A két detektív kártyázott, a postaaltisztek bóbiskoltak.

— Gyerünk, — sugta a harmadik a kocsi elejéről. Már el is hagytuk a megjelölt helyet. Seb- helyes szereti, ha betartjuk az utasításait! Vigyázatok, boys! Úgyesen.

Azokat nem kellett biztatni.

Félperccel később a postakocsi leszakadt a vonat- ról. Ezuttal a mozdonyvezető észre sem vette, mert a dolog közvetlenül a kanyar előtt történt. De odabent, a postakocsi belsejében is csak akkor vették észre, amikor már lépésben gördültek. Na- gyon elfoglalta őket a kártya. Igaz, ha előbb re- szik észre, akkor sem segíthetnek. . . .

... a kocsik megálltak, már csatoltak a

2. Vegyes raszteres beállítás karaktersimítással

majdnem lebukfencezett. Szerencsésebb társai kapták el.

És akkor, ott a vagon tetején, mindhárman csodálkozva néztek egymásra. Ezuttal nem a postakocsi volt az utolsó vagon a szerelvényen, hanem egy szalonkocsi.

A szalonkocsi nyitott ablakán keresztül mindent láttak. Egy káprázatosan szép, vöröshaju leány ült a rádiókészülék előtt és cigarettázott.

— Most mi lesz? — kérdezte az egyik cowboy.

— Mi lenne, mafla. . . — hangzott a válasz. Lepakcsoljuk! Csak nem képzeled, hogy Tom megijed ettől a vöröshaju kis dögtől!

Ketten leereszkedtek a postakocsi elejére. Úgyesen, zajtalanul. Egyikük még meg is állt útközben és bekémlelt a postakocsi belsejébe. Elégedetten vigyorgott. A két detektív kártyázott, a postaaltisztek bóbiskoltak.

— Gyerünk, — sugta a harmadik a kocsi tetejéről. Már el is hagytuk a megjelölt helyet. Sebhelyes szereti, ha betartjuk az utasításait! Vigyázatok, boys! Úgyesen.

Azokat nem kellett biztatni.

Félperccel később a postakocsi leszakadt a vonatról. Ezuttal a mozdonyvezető észre sem vehette, mert a dolog közvetlenül a kanyar előtt történt. De odabent, a postakocsi belsejében is csak akkor vették észre, amikor már lépésben gördültek. Nagyon elfoglalta őket a kártya. Igaz, ha előbb veszik észre, akkor sem segíthetnek. . . .

a kocsik megálltak, már csattogtak a

3. "Tömör méret", nincs vegyes rászter. Fájl méret 104 kB

Jókai Mór:

A koldus veréb



Egyszer, a budai hegyek közé menve, valami pákosz siheder elém áll egy madárral, még pedig verébbel, hogy vegyem meg. Olyan nagyon kérte, hogy vegyem meg, hiszen csak egy krajcár kell érte. Hogy lehetne egy verébért többet kérni egy krajcárnál.

Nekem ugyan a világon akármire másra volt szükségem, csak verébre nem, hanem azt gondoltam, hogy ennek a fiúnak nagyon kell az a krajcár, ennek a verébre pedig nagyon kellene a szabadság, hát csak megvettem a jámbort.

A sváb kölyök szaladt a krajcárral, hogy vissza ne híhassam vele, én pedig a tenyeremre tevém a madarat, hogy repüljön, ahova akar.

De az csak nem repült sehova, a jámbor, hanem ült a tenyeremen nagy gubbasztva, felborzolt tollakkal, s csak fejét emelgette felém.

Közelebről nézve észrevettem aztán, hogy a kis madárnak az egyik lába eltörött, hihetőleg a törben vergődése közben, s csak a bőrfoszlánynál fogva függ azon.

Már ezen majd segítünk, mondtam a jámbor szenvedőnek, s elővéve oltókésemet, szépen lemetszém vele az összezúzott csonkot. Én bizony szívesen ojtottam volna szegénynek mindjárt másik lábat is be, de ámbár ojtott lábú emberekről már hallottam, ojtott lábú madarakat nem ismerek; s ehelyett zsebemben hordott viasszal és gylcs-szatyínggal, mint afféle kertészember, lekötém az amputált lábat. Nem tudom, helyesen cselekedtem-e, mert sebész nem vagyok.

4. Vegyes raszter, nincs karaktersimítás. Fájl méret 22,5 kB

Jókai Mór:

A koldus veréb



Egyszer, a budai hegyek közé menve, valami pákosz siheder elém áll egy madárral, még pedig verébbel, hogy vegyem meg. Olyan nagyon kérte, hogy vegyem meg, hiszen csak egy krajcár kell érte. Hogy lehetne egy verébért többet kérni egy krajcárnál.

Nekem ugyan a világon akármire másra volt szükségem, csak verébre nem, hanem azt gondoltam, hogy ennek a fiúnak nagyon kell az a krajcár, ennek a verébre pedig nagyon kellene a szabadság, hát csak megvettem a jámbort.

A sváb kölyök szaladt a krajcárral, hogy vissza ne híhassam vele, én pedig a tenyeremre tevém a madarat, hogy repüljön, ahova akar.

De az csak nem repült sehova, a jámbor, hanem ült a tenyeremen nagy gubbasztva, felborzolt tollakkal, s csak fejét emelgette felém.

Közelebről nézve észrevettem aztán, hogy a kis madárnak az egyik lába eltörött, hihetőleg a törben vergődése közben, s csak a bőrfoszlánynál fogva függ azon.

Már ezen majd segítünk, mondám a jámbor szenvedőnek, s elővéve oltókésemet, szépen lemetszém vele az összezúzott csonkot. Én bizony szívesen ojtottam volna szegénynek mindjárt másik lábat is be, de ámbár ojtott lábú emberekről már hallottam, ojtott lábú madarakat nem ismerem; s ehelyett zsebemben hordott viasszal és gylcs-szatyínggal, mint afféle kertészember, lekötém az amputált lábat. Nem tudom, helyesen cselekedtem-e, mert sebész nem vagyok.

5. Vegyes rászter karaktersimítással. Fájl méret 23 kB

Jókai Mór:

A koldus veréb



Egyszer, a budai hegyek közé menve, valami pákosz siheder elém áll egy madárral, még pedig verébbel, hogy vegyem meg. Olyan nagyon kérte, hogy vegyem meg, hiszen csak egy krajcár kell érte. Hogy lehetne egy verébért többet kérni egy krajcárnál.

Nekem ugyan a világon akármire másra volt szükségem, csak verébre nem, hanem azt gondoltam, hogy ennek a fiúnak nagyon kell az a krajcár, ennek a verébre pedig nagyon kellene a szabadság, hát csak megvettem a jámbort.

A sváb kölyök szaladt a krajcárral, hogy vissza ne híhassam vele, én pedig a tenyeremre tevém a madarat, hogy repüljön, ahova akar.

De az csak nem repült sehova, a jámbor, hanem ült a tenyeremen nagy gubbasztva, felborzolt tollakkal, s csak fejét emelgette felém.

Közelebről nézve észrevettem aztán, hogy a kis madárnak az egyik lába eltörött, hihetőleg a törben vergődése közben, s csak a bőrfoszlánynál fogva függ azon.

Már ezen majd segítünk, mondám a jámbor szenvedőnek, s elővéve oltókésemet, szépen lemetszém vele az összezúzott csonkot. Én bizony szívesen ojtottam volna szegénynek mindjárt másik lábat is be, de ámbár ojtott lábú emberekről már hallottam, ojtott lábú madarakat nem ismerem; s ehelyett zsebemben hordott viasszal és gylcs-szatyínggal, mint afféle kertészember, lekötém az amputált lábat. Nem tudom, helyesen cselekedtem-e, mert sebész nem vagyok.