

*En guide till att skanna kurslitteratur
på universitet och högskola*

**LADDA NER
OCH DELA MED DIG AV
KURSLITTERATUREN**



Varför kopiera?

Att hjälpas åt!

Kurslitteraturen kostar flera tusen per år för studenter på Sveriges högskolor, komvux och universitet. För många studenter blir dessa kostnader svåra att handskas med när man samtidigt ska betala dyra hyror och höga matpriser. Dessutom är kurslitteraturen i många fall onödig för kursens mål. Inte sällan är listan på kurslitteratur upp emot tio-tolv böcker för ett moment (ca 6 veckors studietid) trots att man oftast inte behöver fler än ett par – tre stycken. Kostnaderna för ett moment kan således hamna på mellan 1000-3000 kr, beroende på hur många böcker man köper. Detta blir på tok för mycket pengar i förhållande till dagens studiemedel och många studenter ekonomiska situation.

Utbildning i Sverige sägs vara gratis men när man börjar kolla på kostnader för kurslitteratur, kompendier och artiklar som ska kopieras så inser man snabbt att så är inte fallet. Studenter med en trygg ekonomisk bakgrund blir förfördelade gentemot andra som får huvudbry över att få pengarna att gå ihop. Många tvingas dessutom jobba extra, trots att lärarna alltid understryker att studierna ska vara en heltidsysselsättning.

För oss som inte är födda med silversked i mun blir det därför naturligt att vi börjar finna andra vägar för att få tag på kunskap. Då har vi allt att vinna på att samarbeta. Om vi hjälps åt att dela kurslitteratur så kan vi se till att alla får tillgång till kunskap, även om pengarna inte räcker till. Det är inte mycket konstigare än att vi har bibliotek, där man gratis får tillgång till all kunskap. Ett sätt att lättare dela med sig av kurslitteraturen och förenkla sin studiegång är att skanna, vilket du i den här guiden kan lära dig det mesta av grunderna.

Skanning och kopiering

Som alla redan vet så är inte kopiering något nytt fenomen. I alla tider har elever stått och svettats i kopiatorrummen för att ändå bara få suddiga och halvdana kopior, som dessutom sällan blir så billiga eftersom det brukar ligga en avgift på ca 1 kr per sida. Och när man äntligen är klar så har man ändå bara en enda kopia. Hur ska vi komma undan billigare och dessutom få ett obegränsat antal kopior som vi kan dela med andra i samma sits? Svaret är skanning!

I mångt och mycket är scanning och kopiering liknande, men med en stor skillnad: istället för en papperskopia så får man en digital kopia i datorn! En digital kopia har många fördelar i jämförelse med en vanlig kopiering: den går att dela hur många gånger man vill, den går att skicka över internet till andra som vill ha den och dessutom kan man fortfarande skriva ut den själv ifall man föredrar att läsa dom på papper och inte i datorn. Att skanna böcker är dessutom inte särskilt svårt och i stort sett är det enda som behövs en vanlig dator, lite programvara och en skanner.

Förberedelser

Att involvera klassen

Att skanna kan vara något som tar en hel del tid om man ska göra det själv. Om man hjälps åt så går det mycket fortare och det kan dessutom vara rätt roligt. En grupp på mellan två till fem personer kan vara bra för att skanna in en bok. Det som krävs för att skanna är givetvis en scanner, en dator med hyfsad prestanda och ett skanningsprogram. Därefter är det bara att sätta igång.

Ibland kan det kännas lite svårt att få med sig kompisar på att skanna då det kan upplevas som att man gör något fel, samtidigt är det inte värre än det vi redan gör, kopierar och delar med oss. Det är ju egentligen helt orimligt att vi alla ska köpa en bok var för 400 kr när vi istället kan låna den på biblioteket, skanna in den och ha den på datorn. Börja snacka med dina närmsta kompisar i klassen och kolla om dom skulle vara intresserade. När ni har skannat in någon bok kan ni snacka med andra, kanske ge dom den bok ni precis skannat in och fråga om dom vill vara med och hjälpa till. Ju fler man är desto fortare går det, bara man kan dela upp arbetet på ett bra sätt.

Varför inte kombinera nytta med nöje och dra ihop en liten klassfest samtidigt som man skannar? Bjud in folk på ett facebook-arrangemang och ät mat, drick öl och ha kul tillsammans. Tänk bara på att försöka göra bort så mycket skanning som möjligt tidigt på kvällen, eftersom man tenderar att bli slarvigare allteftersom klockan går!

Material

Först och främst behövs givetvis en scanner. En bra och snabb scanner är rätt dyr, men om man är några i klassen eller bekantskapskretsen som går ihop så kostar den ungefär lika mycket som EN bok. Du kommer alltså att kunna spara rätt mycket på att skanna in kurslitteratur. För en riktigt state of the art-scanner för böcker är Plustek Optic Book 4600 att rekommendera, den finns att beställa på www.dustin.se. Men smakar det så kostar det, ca 5000 kr får man betala. Men det finns scanners för runt 1000 kr som inte är lika snabba, men som fungerar bra ändå. Det tar bara lite längre tid att skanna in boken.

Det första man gör är alltså att skanna in boken. Beroende på hur snabb scannern är och hur många sidor du vill skanna kan det ta mellan trettio minuter till två timmar. Att skanna kan faktiskt vara ganska avkopplande, men det kan vara skönt att ha någon kompis att byta av med mellan varven. Så länge man är några stycken så kan man sitta och plugga samtidigt och vara lite effektiv, eller så dricker man kaffe och snackar skit!

Förutom en scanner så behöver du en dator. En vanlig, hyfsat ny laptop brukar funka. Det är givetvis ännu lättare om man har en större skärm att koppla in den på, men det är inget krav. Se till att ha några gigabyte ledigt på hårddisken då böckerna kan vara så stora som upp till 10 gb innan man har gjort om dem till PDF. De kopiorna kan man dock ta bort så fort man har gjort om dem till PDF, så det är bara under skanningen som du behöver ledigt utrymme på hårddisken.

På datorn behöver du ett OCR-program. OCR står för Optical Character Recognition och innebär att programmet läser av bilderna som du skannar in och känner igen bokstäverna och sätter dem i en textfil. Detta gör att texten går att söka i och minskar storleken på filen. De flesta böckerna brukar blir mellan 4 och 10 megabyte stora vilket gör att ett gäng böcker knappt tar någon plats alls på din hårddisk och enkelt kan skickas via e-post.

Det bästa OCR-programmet enligt oss är ABBY FineReader. Det finns en rad andra OCR-program också som är gratis och ibland skickar scannertillverkarna även med egna program för OCR, men av de program vi provat är ABBY FineReader överlägset bäst. Det finns att ladda ner på [the piratebay](http://www.thepiratebay.com).

Att skanna

Skanningen brukar gå ganska fort och smärtfritt. Det man ska tänka på är att se till att hålla ner boken stadigt mot scannern då den läser in och att inte stressa, då blir det lätt misstag som man måste fixa i efterhand. Första gången man skannar kan det kännas ovant men man hittar snabbt en bra rutin.

I den här guiden så har vi utgått från en Plustek Optic Book 4600 med programmet ABBY FineReader 10 Corporate Edition. Om ni kör med en annan scanner så kan ni behöva fixa med andra inställningar, men det mesta borde vara sig likt. Följ nedanstående tips så borde ni klara er bra:

- Se till att boken är fri från klotter och understrykningar med penna! Dessa streck kan göra att OCR-programmet gör en del fel i avläsningsprocessen vilket kommer leda till mer jobb. Vanlig överstrykningspenna brukar inte vara några problem, men blyerts och bläck kan bli besvärligt. Så gå igenom boken och sudda!
- Se till att scannerbädden (glasrutan som scannern läser av) är ren. Ifall den är dammig och täckt i fingeravtryck så kan det försämra kvalitén på bilderna. Försök att inte ta så mycket med fingrarna på scannerbädden.

Sätt igång scannern.

Starta FineReader10 på datorn.

Finereader kommer ge er olika alternativ på vad ni vill göra.

1.1 Ignorera dessa och tryck istället på "Läs in"-alternativet till vänster.

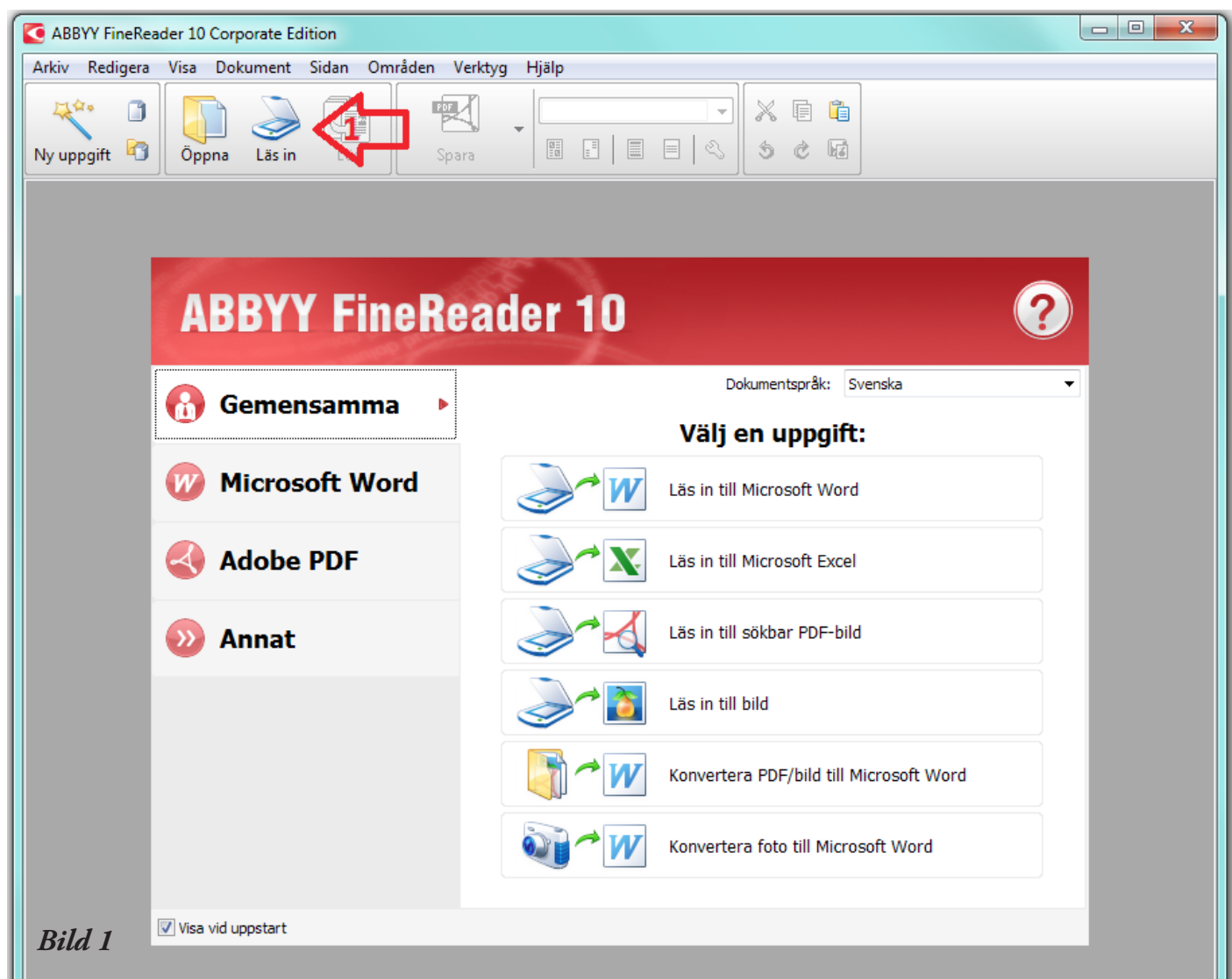


Bild 1

2.1. Ni får nu upp ett nytt tomt dokument och Inläsnings-rutan öppnas. Lägg på boken på scannern och välj valfri sida (gärna någonstans i mitten av boken) och tryck på "Preview"-knappen. Scannern kommer antagligen värma upp och kalibreras innan den kör igång så stressa inte, bara vänta.

2.2. När scannern har läst sidan ser ni till att markera med hjälp av den blåa ramen hur mycket av sidan ni vill ska läsas in.

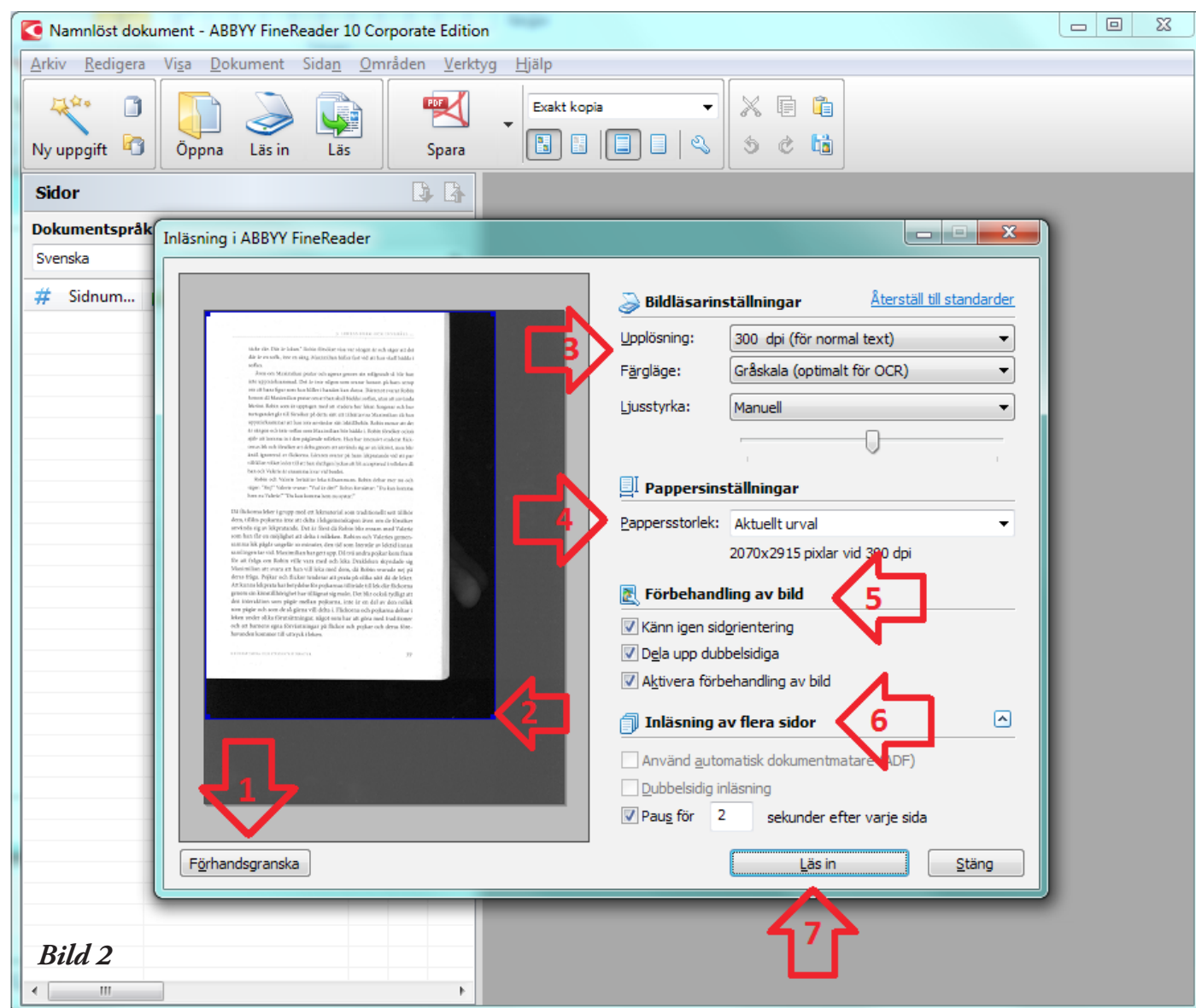
2.3. Se till att scannern är inställd på gråskala och 300 dpi. Om du har en bok med extremt mycket liten text och kursiv text så kan du höja dpi:n men då kommer boken ta upp ett större filutrymme och längre tid att skanna.

2.4. Se till att det står "Aktuellt urvall" under Pappersinställningar efter att du har förhandsgranskat bilden du ska läsa in.

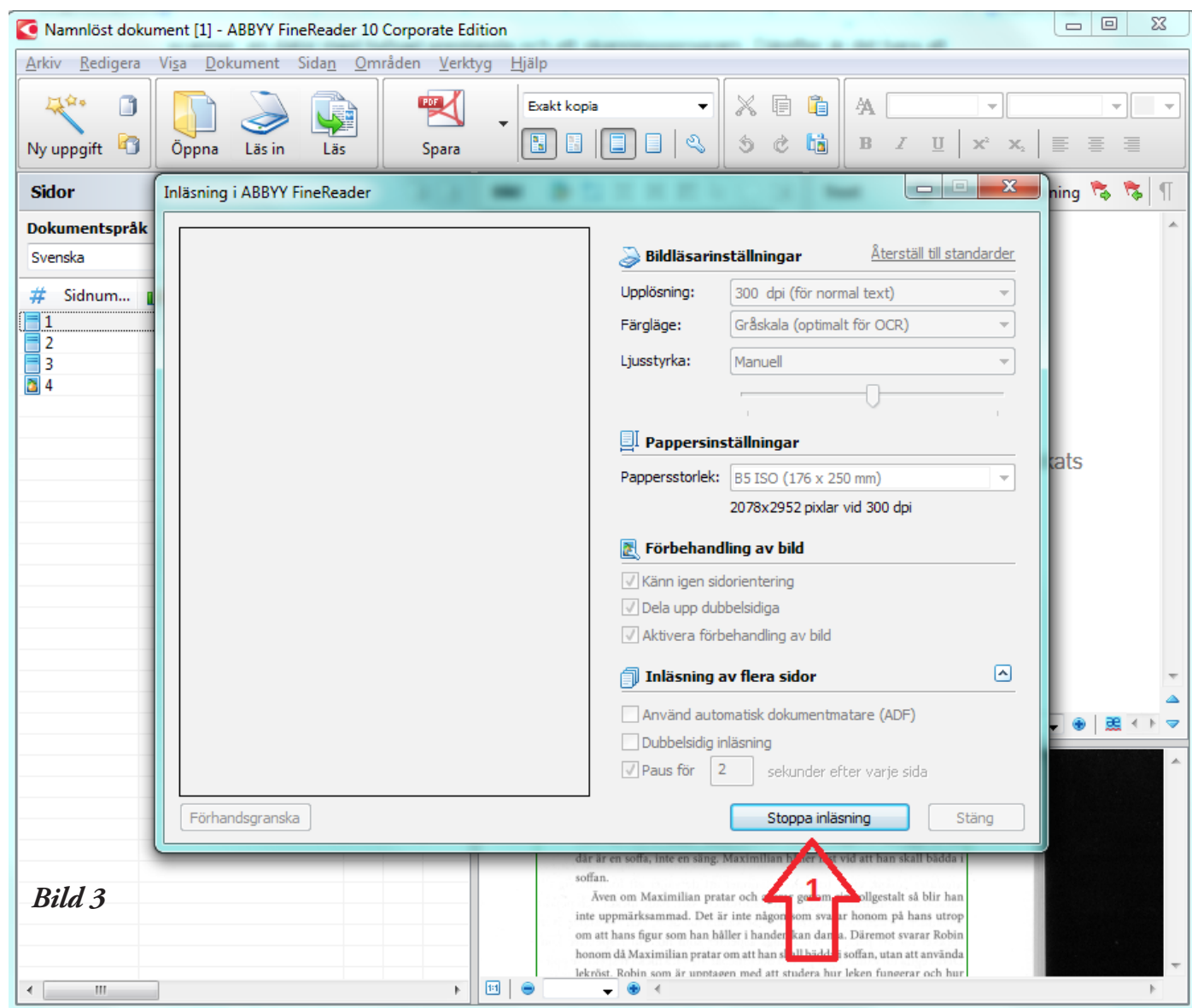
2.5. Under "Förbehandling av bild" kan du ha alla rutorna förbockade då det underlättar OCR:ningen.

2.6. Klicka i om ni vill ha manuell inläsning (tryck då bara på Läs in). Om du vill att scannern kontinuerligt ska läsa in sidor utan att behöva klicka någonstans så går du till "Inläsning av flera sidor" och fyller i hur många sekunder du vill ha mellan varje inläsning. (3-4 sekunder brukar de flesta hinna med på).

2.7. Tryck på "Läs in". Om du valt att läsa in manuellt så kommer dialogrutan att försvinna för att sedan komma tillbaka när scannern läst sidan färdigt. Om du har valt att ha automatisk inläsning med några sekunders mellanrum så försvinner dialogrutan och du måste själv klicka upp den.



3.1. OBS! Om du har valt kontinuerlig inläsning med några sekunder mellanrum och vill ta en paus, vill avbryta eller är klar så måste du trycka upp dialogrutan (som minimerats och lagts i verktygsfältet) och trycka på "Stoppa inläsning". Tryck INTE på Avbryt i inläsningsrutan eftersom programmet då avbryter inläsningen vilket blir jobbigt när du ska börja OCR:a senare.



4.1. Detta är en miniatyr av sidan du läst in. Genom att klicka på en miniatyr kommer du till den sidan i det aktuella dokumentet.

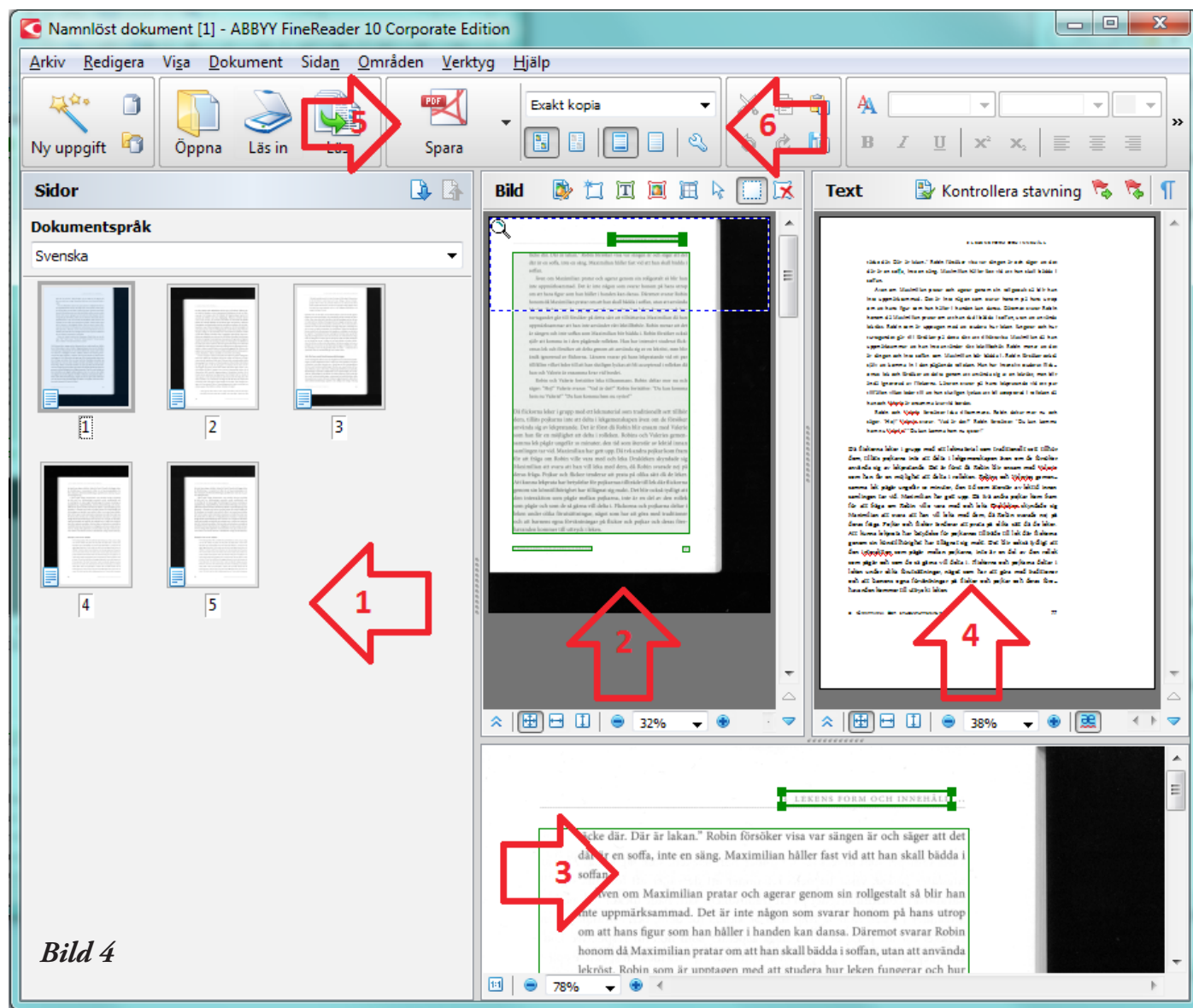
4.2. Detta är hur bilden du skannat in ser ut. Som ni kan se har vi inte valt en bra markering.

4.3. Bilden som du skannat in i förstoring.

4.4. Så här kommer sidan se ut när den är inläst och du sparat den som PDF.

4.5. Se till att "Spara-rutan" är inställd på PDF och att det står (6) "Exakt kopia" i rutan bredvid.

Gå in på Verktyg > Alternativ ... > PDF-fliken. Under alternativet "Standardpappersstorlek" ändrar du från "automatisk" till "Behåll ursprunglig bildstorlek". På det sättet undviker du att du får för mycket vit kant runt texten. Denna inställning brukar inte sparas ifall du påbörjar ett nytt scanningsprojekt, så se till att göra detta varje gång.



Nu är det bara att sätta igång att skanna, ett tips för att minska antalet sidor är att hoppa över register-sidorna i slutet av boken då man ändå lätt kan söka i texten senare. Ta gärna med fram-och baksida för en snyggare layout.

När hela boken är inskannad går du på Arkiv > Spara som FineReaderdokument. Då sparar du en råkopia så att du kan fortsätta OCR:a senare. Om du har sparat filen en gång så sparas den automatiskt eftersom. För säkerhets skull så spara alltid när du läst in alla sidor.

OCR

När boken är färdigskannad ska du ge dig på OCR:ningen. OCR står för Optical Character Recognition och innebär att programmet läser av bilderna som du skannar in och känner igen bokstäverna och sätter dem i en textfil. Detta gör att texten går att söka i och minskar storleken på filen. De flesta böckerna brukar bli mellan två och tio megabyte stora vilket gör att ett gäng böcker knappt tar någon plats alls på din hårddisk och kan enkelt skickas som e-post.

När programmet läser in texten till ett nytt dokument så kan det ibland bli fel. Oftast är det sid- och kapitelnummer som strular, och ibland ser den inte skillnad på t.ex. "i" och "l". Ibland tror den också att en bild är text och tvärtom. Därför kan det vara bra att gå igenom dokumentet och ändra detta så att allt ser bra ut. Man kan själv välja hur noggrann man vill vara. Om man ser till att fixa alla sidnummer och bilder och ändrar de mest uppenbara stavfelen så brukar det räcka gott!

Det är enklare än man tror, man märker rätt snabbt vad som blir fel och så kan man korrigera det, men det är tidskrävande och tar några timmar, beroende på hur noga man vill vara. Följ nedanstående tips så borde ni klara er fint!

Vi går först och främst igenom de grundläggande funktionerna i ABBY FineReader när man OCR:ar.

I FineReader finns tre olika områdestyper som man behöver hålla koll på. De grönmarkerade områdena är textområden, de röda är bildområde och de blå är tabeller.

5.1-2. Ibland så kommer programmet feltolka områden, och i det här exemplet har ett bildområde tolkast som en text, vilket visas som en massa konstiga symboler i textfönster.

5.3. För att rätta till detta finns det två alternativ: Antingen så högerklickar man på området, väljer Ta bort område och markerar sedan området med bildverktyget.

5.4. Det andra alternativet är att man högerklickar på området och väljer Byt typ på område och sedan väljer Bild.

Samma sak gäller givetvis ifall programmet tolkat någon annan områdestyp fel, t.ex. ett tabellområde som ett textområde. Om ett område är svårtolkat och det aldrig ser riktigt bra ut kan du alltid göra om det till ett bildområde, då ser sidan ut som i originalet. Tänk dock på att bilder gör den slutgiltiga filen större och att det inte går att söka i text som är markerad som bildområden.

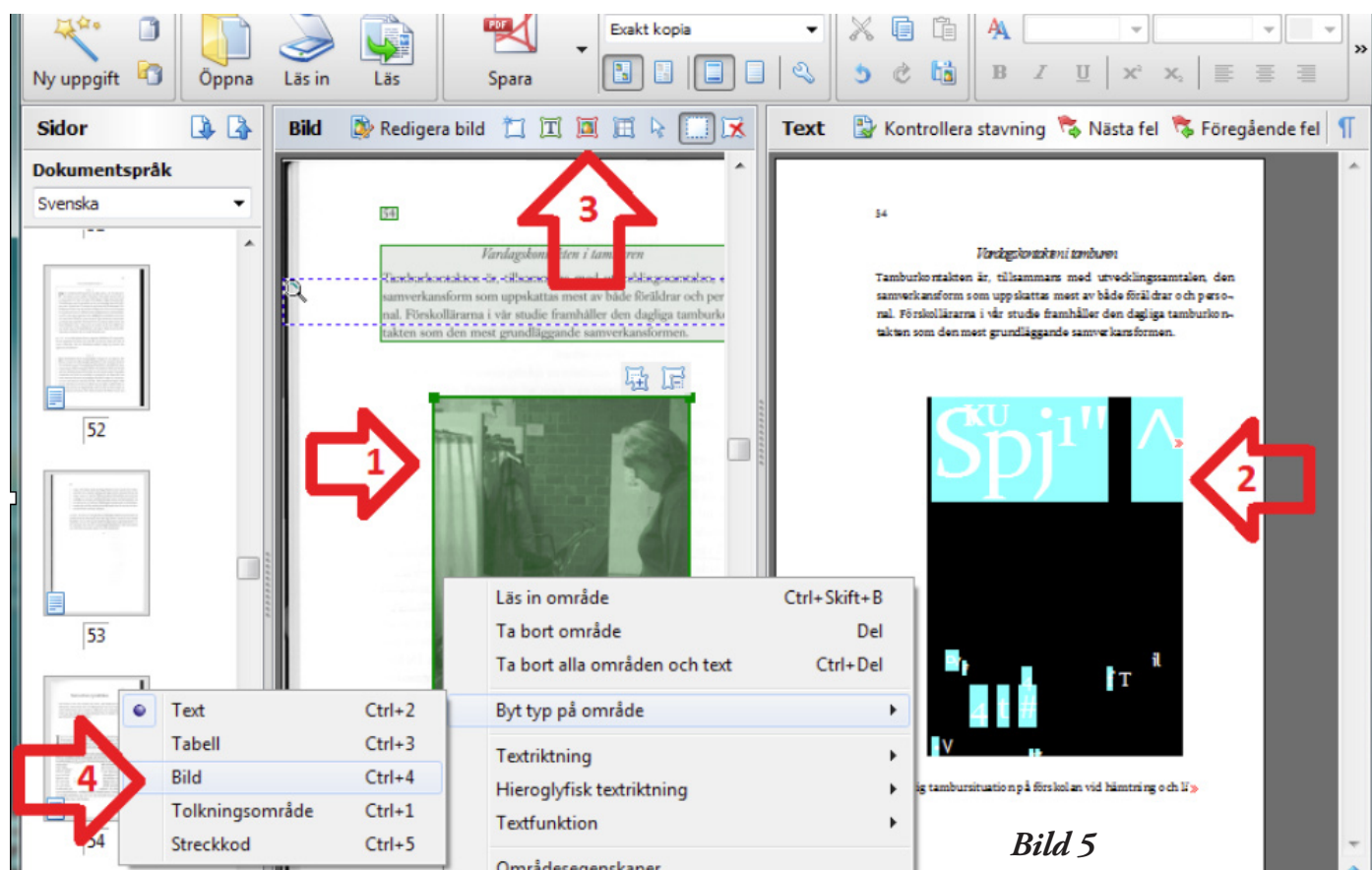


Bild 5

6.1. När texten är grönmarkerad i bildfönstret (det vänstra fönstret) så innebär det att programmet anser att det är text. I det här fallet stämmer det vilket också syns på den högra sidan (textfönstret) där sidan har blivit inläst. Här har dock programmet missat den översta rubriken vilket man ser i det vänstra textfönstret. Längst ner kan man också se på den förstorade bilden att texten inte är markerad.

Vad man måste göra då är markera den missade rubriken, antingen genom att dra upp den gröna rutan så att den täcker även den missade texten eller genom att skapa en till textruta. Sen högerklickar man på det grönmarkerade området > Läs in område. Då läser programmet även in den texten. Då kan man se texten även i det högra textfältet.

6.2. Det rödmarkerade fältet i bildfönstret innebär att programmet analyserat detta som en bild. I det här fallet stämmer det, men den har inte fått med hela bilden vilket syns i textfönstret. Man gör då på samma sätt som med texten, antingen så drar man ut den röda rutan så den täcker hela bilden eller så gör man en ny bildruta genom att markera bildverktyget. Observera då att man inte behöver trycka på "Läs in område", bilder visas direkt i det högra textfönstret.

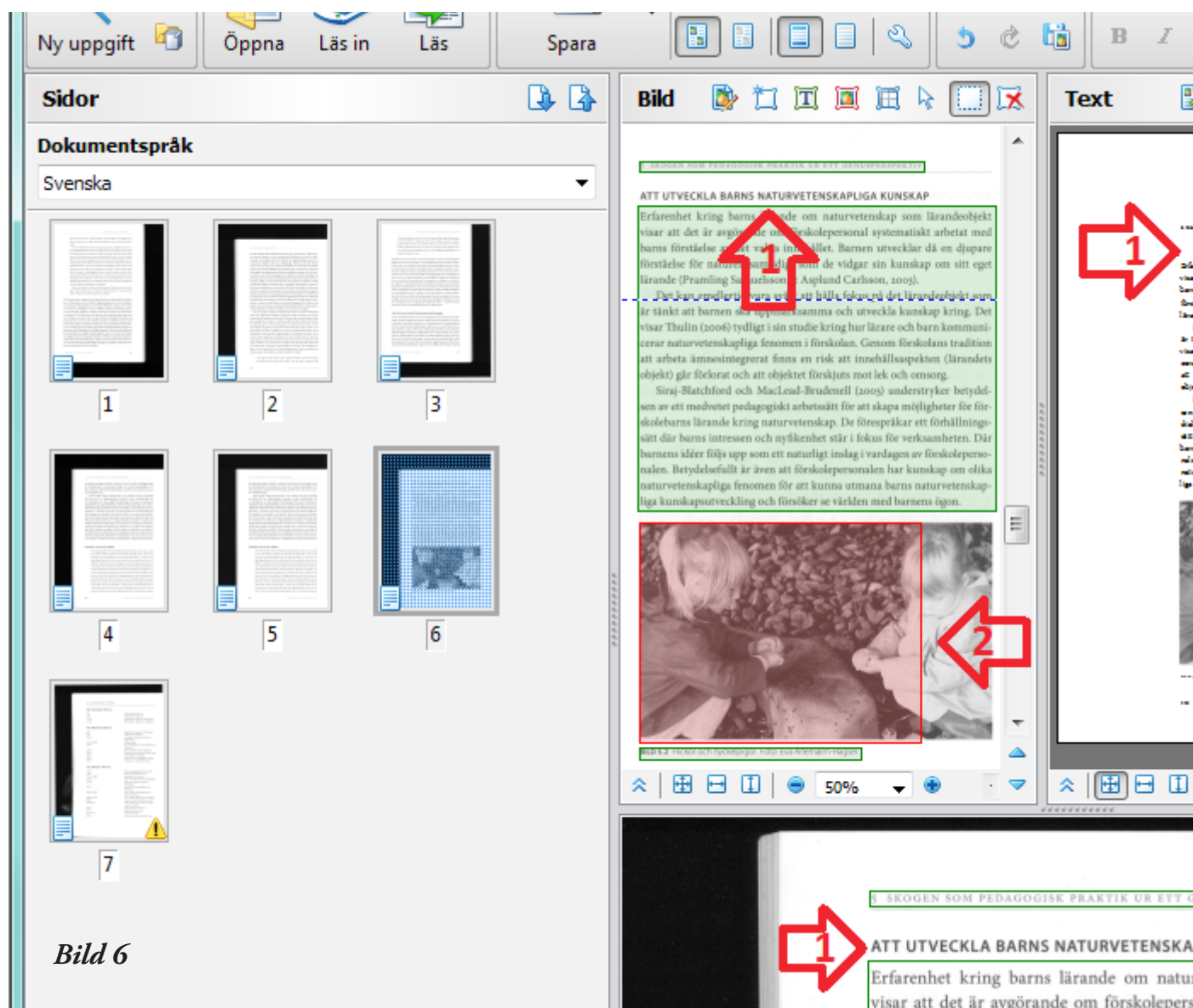
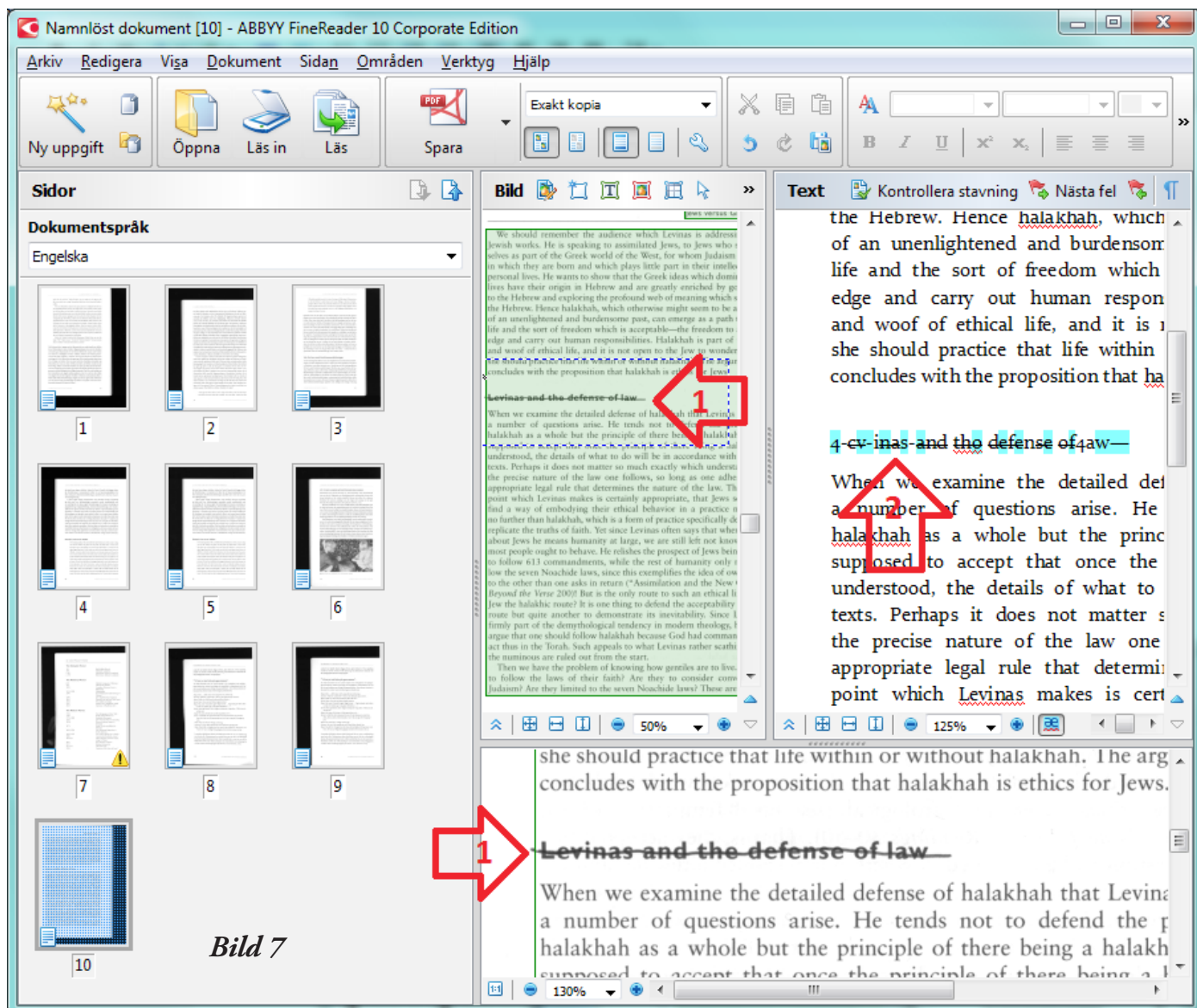


Bild 6

7.1. I det här fallet har någon kluddat med en bläckpenna i texten vilket gör att texten inte lästs in ordentligt (se pil 2). För att åtgärda detta får man helt enkelt gå in i det vänstra textfönstret och, markera biten som är fel och skriva in den korrekt. Lättast är då att titta i det inzoomade fönstret och skriva "Levinas and the defense of law" eller vad det nu ska stå så blir det korrekt.

7.2. Observera att texten i det högra textfönstret blivit bitvis blåmarkerad, detta innebär att programmet är osäker på vad som ska stå där. När man letar efter fel i texten är det alltså väldigt enkelt att snabbt titta efter blåmarkerade avsnitt och jämföra dem med bildfönstret och se om det överrensstämmer. Oftast är det dock bara programmet som inte känner igen ord eller bokstäver men ändå skrivit in det rätt. Man kan alltså snabbt bara se över detta när man OCR:ar.



Vanliga fel

Sidnummer och kapitelnummer

Den version som vi använder oss av har en liten ovana i att läsa in fel sidnummer. Oftast stämmer sidnumren, men i vissa partier så kommer den lägga in samma sidnummer på varje/varannan sida. Om man ska referera till boken med sidhänvisning så blir detta givetvis ett problem. För att åtgärda detta får man gå igenom hela dokumentet sida för sida och ändra på de ställen där programmet gjort fel. Det görs antingen genom att man manuellt skriver in sidnumret i textfönstret på varje sida, eller genom att högerklicka på textområdet och välja *Läs in område*.

ÅÄÖ och andra okända bokstäver och symboler

Programmet känner inte alltid igen bokstäver som t.ex. ÅÄÖ, punkter och asterixer, fotnoter och annat. Dessa blir då oftast blåmarkerade i textfönstret. Om man vill vara petig så kan man ändra alla dessa, men oftast så är texten ändå läsbar så det är absolut inget måste.

Varningstrianglar

Ibland blir vissa av de inlästa sidorna i Sidor-fönstret markerade med en liten varningstriangel. För att se vad felet är så kan man byta till Detaljvy i Sidor-fönstret (se nedan). Denna kan bero på olika saker. Det absolut vanligaste är att programmet är osäker på vilket språk som gäller på sidan. Detta är oftast inget man behöver bry sig om.

En annan vanlig orsak är att programmet inte tolkat sidan. Detta kan åtgärdas genom att högerklicka på sidan i Sidor-fönstret och sedan välja *Läs in sida*.

Sidfönstervy > Detaljer

Ett bra verktyg för att snabba på OCR-processen är att använda sig av Detaljvyn i Sidor-fönstret. För att växla till den så högerklickar man varsomhelst i Sidorfönstret och klickar på Sidfönstervy > Detaljer.

Genom att sortera sidorna efter Otydliga tecken, så kan man snabbt se vilka sidor som har flest fel på sig och rätta dessa direkt. Ifall man inte känner att man vill göra en extremt noggrann OCRning så kan man se till att i första hand gå igenom alla sidor som har en felprocent över 2%.

Spara

När du känner att du är färdig med dokumentet är det dags att göra om den till en PDF-fil. Men först måste du kolla att programmets PDF-inställningar är rätt. Den hittar du på Verktyg > Alternativ... och sedan i fliken PDF.

8.1. Standardpappersstorleken ska vara inställd på “Behåll ursprunglig storlek”, annars kommer du få en stor vit ram eller bortklippt text på din slutgiltiga PDF.

8.2. Sparläge ska vara inställd på “Endast text och bilder”, för att hålla nere storleken på filen.

8.3. Det viktigaste bland dessa alternativ är att du markerar “Behåll sidhuvud och sidfot”, annars riskerar sidnumreringen att försvinna.

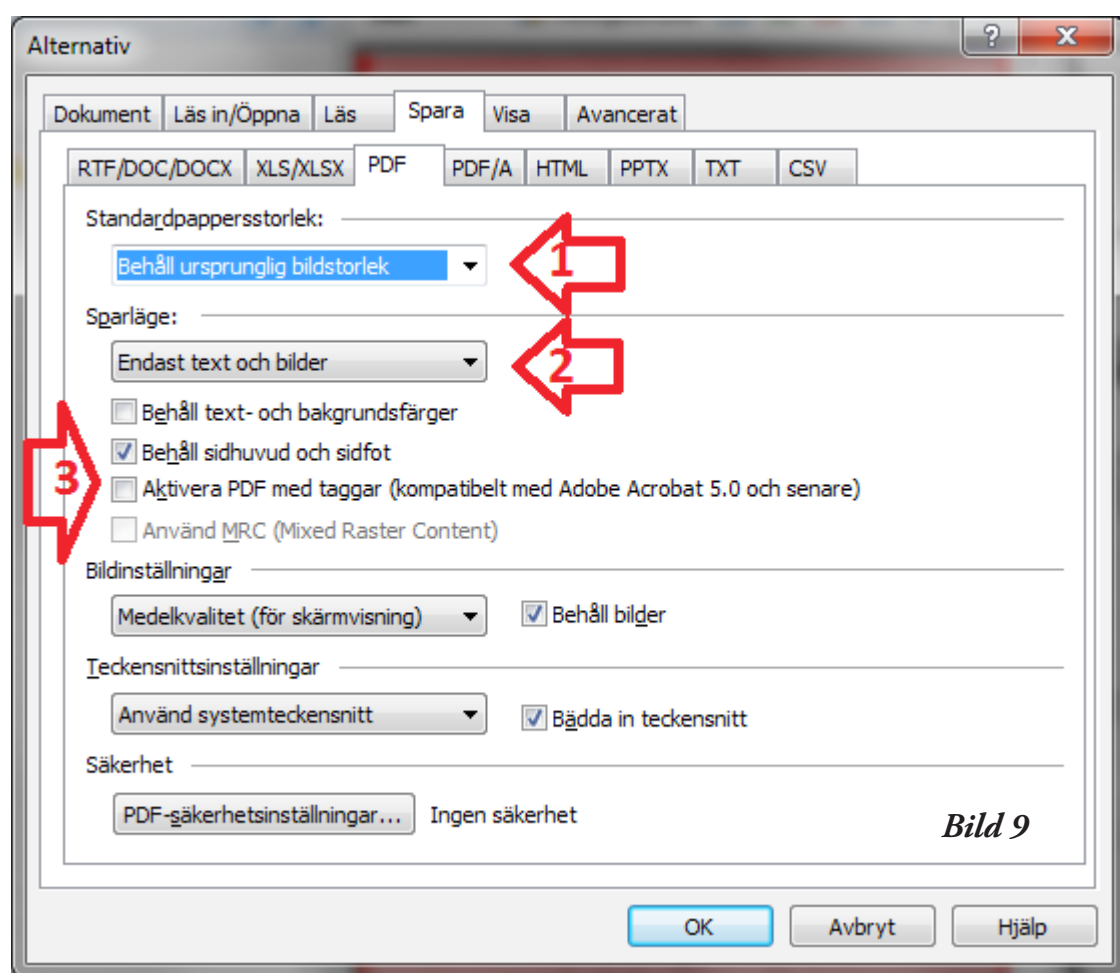


Bild 9

Sedan är det bara stänga menyn och klicka på spara!

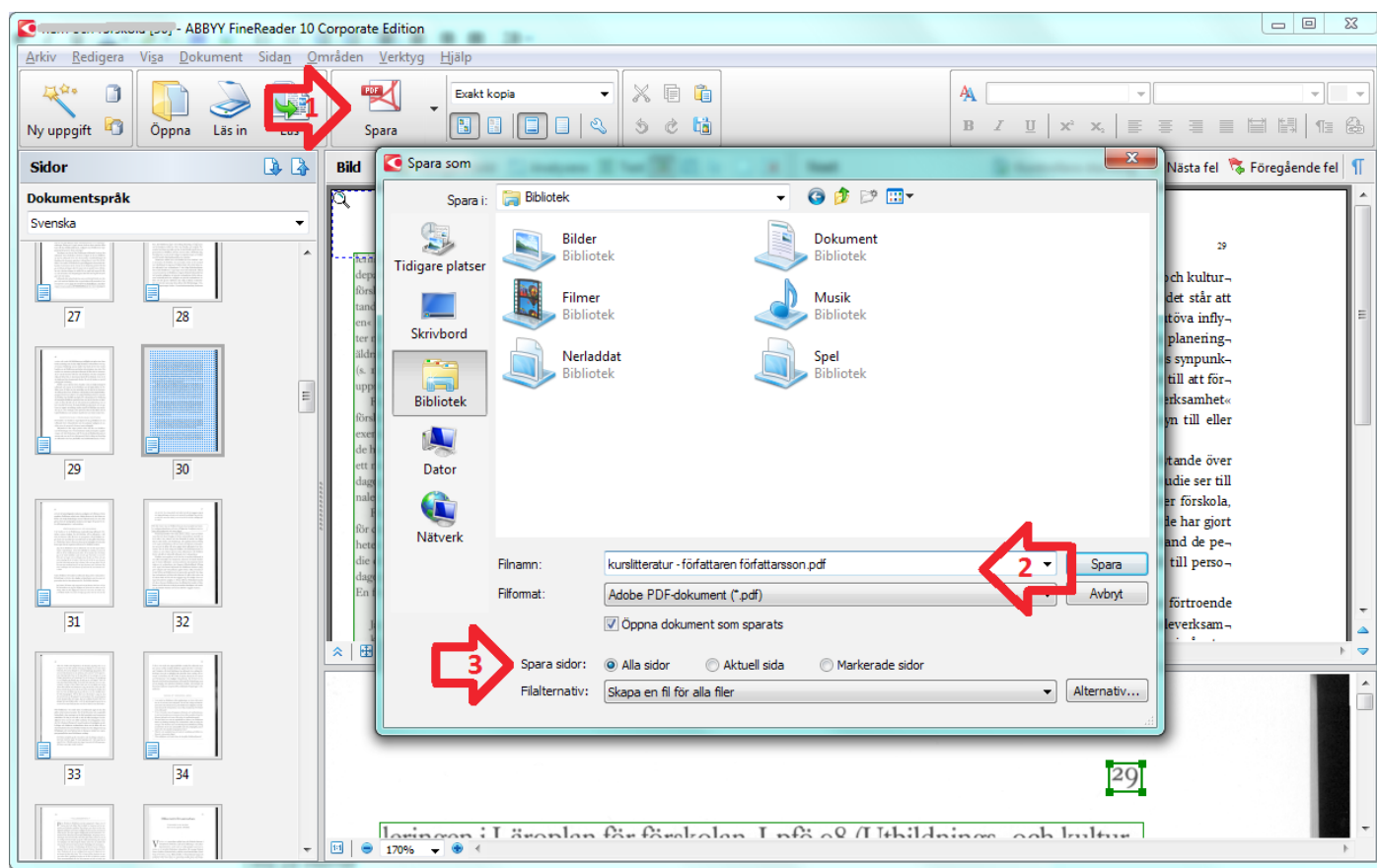
9.1. Klicka på Spara-knappen så öppnas en dialogruta.

9.2. Namnge filen. Förslagvis med titel och författarnamn.

9.3. Se till att alternativet Spara sidor: Alla sidor är ifyllt. Klicka sedan på spara!

Dokumentet kommer nu automatiskt öppnas i din PDF-läsare och så du kan kika igenom det och se till att allt är som det ska. Om du känner dig nöjd så kan du ta bort FineReader-dokumentet, för att frigöra mer plats på din hårddisk. Det går även att redigera PDF-filer i FineReader ifall du skulle hitta något fel, men tänk på då på att du inte har originalbilderna kvar i dokumentet att utgå ifrån.

Grattis, du har nu skannat och OCR:at boken!



Dela på internet

När boken är färdigscannad och OCR:ad så har du en fil på mellan 5-30 mb. För att dela med dig till andra så har du nu olika alternativ: antingen så kan du skicka boken via mail, lägga den på ett USB och ge det till dina kompisar, eller så lägger du upp den på nätet. Du kan t.ex. lägga upp boken på Piratebay eller Rapidshare.

Vi vill förstås att alla som skannar hjälps åt och sprider det skannade materialet så vi slipper göra dubbel-arbete. Det handlar också om att hjälpas åt, inte bara dra nytta av saker som andra har lagt ner en massa tid på. Vi som gjort den här guiden är delaktiga i sidan *kunskapsfabriken.nu*. Om du skickar en länk till *kunskapsfabriken@gmail.com* så lägger vi upp det på *kunskapsfabriken.nu* i vårän länklsta. Du kan också maila en PDF om du inte har möjlighet att själv lägga upp det på nätet.

Länkar till bra sidor

thepiratebay.com – här ligger en hel del av tidigare skannad litteratur

frikurslitteratur.blogspot.com – några andra som också skannar

books.google.com – inte ett projekt för att hjälpas åt, men i värsta fall kan man hitta svåra utländska böcker