

Vejledning til korrektur af Trykkefrihedens Skrifter

Webapplikationen AltoFaxEdit benyttes til korrekturarbejdet: altofaxedit-test.kb.dk

Se brugervejledningen.

Efter login kan man klikke på "Læs om fraktur" og printe den vejledning så den altid er tilgængelig.

I princippet ønskes alle OCR-fejllæsninger rettet i skrifterne, men det må accepteres at perfektion ikke kan opnås.

I det følgende beskrives hvorledes forskellige problemstillinger skal håndteres.

Generelt

De originale stavemåder skal beholdes, også hvis der er stavfejl.

Fejl lavet af sætteren skal ikke rettes. Fx beholdes "Riser" hvor sætteren har sat "R" i stedet for "K" i filen ..._1_002_004_010.xml linje 15.

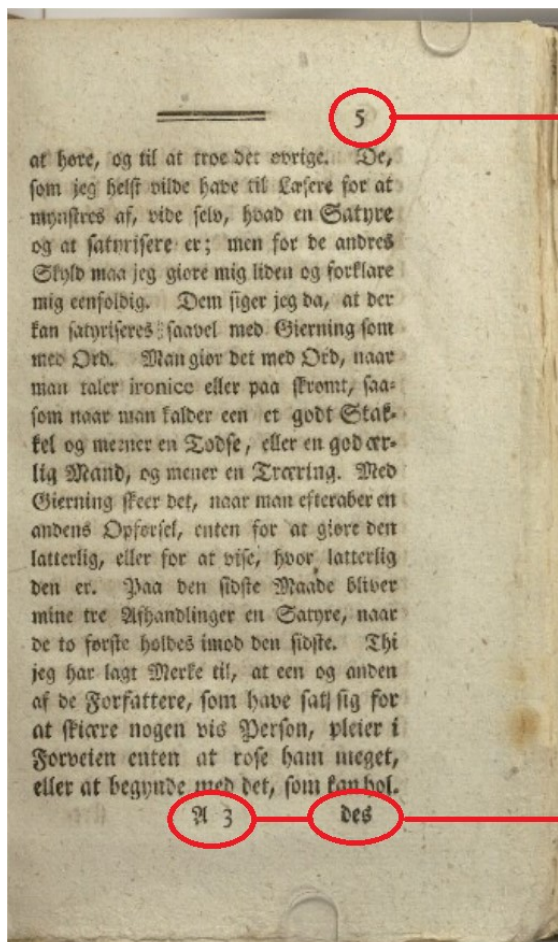
Fejl pga. manglende sværte skal rettes hvis det er muligt at se hvilke bogstaver der mangler.

Fx "S kke" rettes til "Sukke" i filen ..._1_002_006_011 linje 12.

Sidetal og kustoder

Vi må ikke slette sidetallet der står øverst på mange sider. Hvis sidetallet mangler, skal vi indsætte det. Hvis der ikke findes en tom linje øverst til sidetallet, indsættes en ny linje.

Arknummer og kustode slettes, dvs. hele linjen slettes (ved at klikke på skraldespand-ikonen).



Sidetall skal bevares

Originale stavemåder
må ikke ændres

Arknummer og kustode
skal slettes

Brøker

Brøker kan optræde i teksterne. Det drejer sig fx om $1/5$, $1/4$, $1/2$, $3/4$ etc.

Nogle af disse brøker eksisterer som specialtegn på computere: $\frac{1}{4}$, $\frac{1}{2}$ og $\frac{3}{4}$

Vi må dog ikke bruge disse specialtegn, men skal skrive med en skråstreg.

Fx $3/4$ skrives altså med tre tegn (3 skråstreg 4) uden mellemrum mellem tegnene.

Mellemrum mellem sammenskrevne ord

Hvis OCR-softwaren fejlagtigt har indsat et mellemrum i et ord som if. faksimilen tydeligvis er sammenskrevet, skal vi fjerne mellemrummet. Det gøres vha. en af de to ordsammenlægningsknapper som ses når man klikker på et ord.

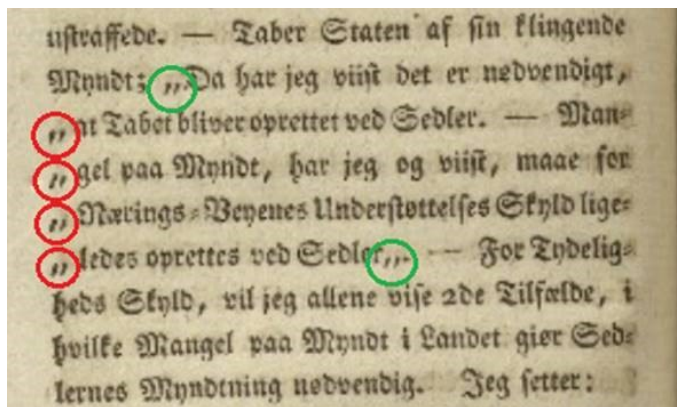
Manglende mellemrum

Hvis OCR-softwaren fejlagtigt har sammenskrevet to ord som if. faksimilen tydeligvis er adskilt, skal vi indsætte et mellemrum på rette sted.

Anførselstegn (gåseøjne)

Anførselstegn anvendes ofte til at markere citater. I Trykkefrihedens Skrifter er der ikke blot anførselstegn i begyndelsen og slutningen af citatet, men længere citater er typisk markeret med et citationstegn i linjestart (og evt. i linjeslut).

Sådanne anførselstegn skal håndteres som forklaret ved dette eksempel:



Ved citater

beholdes anførselteget ved start og slut (grøn markering)

mens de mellemliggende anførselstegn i margen/linjestart slettes (rød markering).

Tabeller

Vi ikke skal lave korrektur af tabeller. Så selve tabellerne springer vi simpelthen over. Men tekst over/under tabellerne skal rettes.

Også på sider hvor der er tabeller skal vi indsætte en ny linje øverst med sidetal hvis det mangler, og vi skal også slette den nederste linje med arknummer og kustode hvis de findes.

Skrifter på tysk, fransk og latin

Nogle skrifter (hele hæfter) er forfattet på tysk og fransk. Vi skal springe dem over. Der skal altså ikke laves nogen korrektur på dem.

Men hvis et skrift er forfattet på dansk med ord eller små segmenter på tysk, fransk eller latin skal vi lave korrektur af disse ord/segmenter.

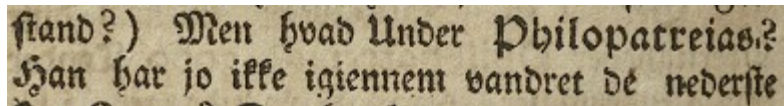
Fjernelsen/indsættelse af orddelingsstreg ved linjeskift

OCR-filerne (XML-filerne i ALTO-format) indeholder information om automatisk identificerede orddelingsstreger ved linjeslut. Det ses ved at en blå bindestreg i OCR-teksten når ordet i linjeslut ikke er i redigeringsmodus:

Aarsagen, hvorfor vore høylovlige Kon-
ger har gjort alt dette, veed hver Borger og

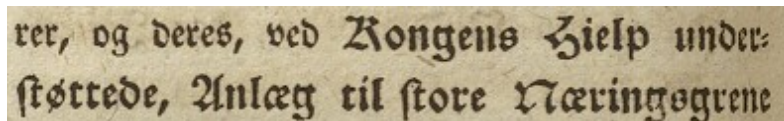
I nogle tilfælde har OCR-softwaren indsat en blå orddelingsstreg fejlagtigt. P.t. har applikationen AltoFaxEdit ikke en funktion til at slette disse blå orddelingsstreger, så i stedet markerer vi dem som skal slettes. Derved kan vi efter korrekturarbejdet er afsluttet anvende en søg-og-erstat-funktion (anden software) til at rette problemet.

Vi markerer at en blå orddelingsstreg skal slettes ved at indsætte @ (snabel-a) allersidst i linjen. Eksempel fra filen ..._1_002_005_028.xml linje 27:



stand?) Men hvad Under Philopatreias?
Han har jo ikke igiennem vandret de nederste

Det modsatte tilfælde ses også. I nogle tilfælde har OCR-softwaren ikke forstået at der var orddeling ved linjeskift. Tilføjelse af orddeling ved linjeskift håndteres ved at indsætte en afsluttende bindestreg. Fx i filen ..._1_003_003_078.xml linje 10 her med vist med rødt hvor der er rettet. Der er manuelt indsat en bindestreg efter "under". Altså "under-":



rer, og deres, ved Kongens Hielp under-
støttede, Anlæg til store Næringsgrene

Markering af noget som skal efterses senere

Fordi der kan være tilfælde som er svære at forklare, og fordi det er nemt at gennemse alle XML-filer for en unik tegnsekvens har vi vedtaget at der kan indskrives en særlig tre-tegnkode til at markere særtilfælde som Bo (boks@kb.dk) og evt. Ulrik Langen (frw104@hum.ku.dk) skal studere efterfølgende. Koden er:

#%=

Koden kan indsættes hvor som helst fordi den senere vil blive fjernet overalt i alle filer med en generel søg-og-erstat-operation

Bemærk, at det er meget vigtigt at der ikke er indsat mellemrum mellem de tre tegn.

OBS: Ved indsættelse af koden beder du hver gang Bo om at finde og åbne den aktuelle fil, og det er tidskrævende. Så vær rar at bruge koden med omhu.