Flettemotor

Har sett på bibliotekene docx4j og ODFDOM, begge for java. Docx4j brukes for å arbeide med Microsoft Office (Office OpenXML) Word dokumenter (.docx), mens ODFDOM brukes for å arbeide med OpenDocument Format dokumenter (OpenOffice/LibreOffice og lignende).

Docx4j:

* Opprette, endre og lagre MS word dokumenter.
* Kan også lagre dokumenter som PDF.

ODFDOM:

* Opprette, endre og lagre OpenDocument Format dokumenter.

# Kodeeksempel:

Har skrevet litt kode for å arbeide med .docx dokumenter, ved å bruke docx4j.

## Funksjonalitet:

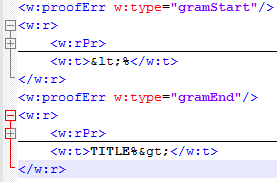
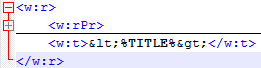
* Bytte ut et placeholder-felt (som f.eks. <%CONTENT%>) med en gitt tekst.
* Bytte ut et placeholder-felt med innholdet i et annet dokument (innholdet i body, ikke header eller footer).
* Sette inn blokker med elementer (tekst, tabeller, bilder o.l.) fra et word dokument inn i et annet dokument.

# Erfaringer og lignende

MS Office dokumenter (Office OpenXML) er I bunn og grunn en zipfil med diverse XML filer (en for innhold, en for styling osv.), så det er mulig å arbeide mot dem direkte ved å pakke ut, parse xml og bruke xpath eller lignende, hvis man føler behovet for det.

Hvis man oppretter et dokument som skal fungere som mal med diverse placeholder-felt (som f. eks. <%TITTEL%>), så må man passe på at ingen av disse placeholder-feltene blir oppfattet som, eller inneholder elementer som oppfattes som stave- eller grammatiske feil. Da vil Word putte inn noen xml-elementer som representerer disse feilene, noe som vil føre til at xml-elementet som inneholder placeholder-feltet vil bli delt opp i flere elementer. Hvis man ikke har tatt hensyn til dette i programmet, så vil ting slutte å fungere slik man har tenkt.

Eksempel: Venstre inneholder stavefeil, som fører til at elementet blir delt opp.



Har lagt noen eksempler hvor jeg tar hensyn til at placeholder elementer kan bli delt opp, så dette er egentlig ikke noe problem lenger, men noe man bare må være obs på.

Det hjelper mye å ha kjennskap til strukturen i xml-en til et Word dokument (Office OpenXML). Hvis noe ikke fungerer slik man hadde tenkt, så kan det være lurt å åpne dokumentet som zipfil og sjekke hva som faktisk ligger i xml-en.

Bilder er lagret som ressurser i Word, og trenger litt ekstra arbeid for å bli kopiert fra et dokument til et annet. Man må blant annet passe på at relasjoner og lignende er riktige. Det finnes et eksempel i kodeeksempelet på hvordan dette kan gjøres.