

Rapport fra deltakelse på Nicar-konferanse 2016

Her kommer en rapport fra [konferansen](#), som i år gikk av stabelen i Denver.

Klok av skade fra i fjor, da jeg mistet en konferansedag grunnet uventet dårlig vær, valgte jeg i år å reise en dag i forveien. Det ga en «bonus-dag» ekstra før konferansstart, med tid til å bli bedre kjent med Denver. Anbefaler andre som skal reise til USA eller andre steder langt unna, å legge inn en slik buffer. Det gir forutsigbarhet.

Fra årets konferanse var det spesielt et verktøy som var spesielt nyttig for mitt vedkommende, nemlig verktøyet [CSV-kit](#). Utviklet av og for journalister, er CSV-kit et perfekt verktøy for å håndtere store datamengder lagret i CSV-formatet. Det kan oppleves som en terskel å bruke det da man må inn i kommandolinja for både å installere det og bruke det, men det var langt enklere enn jeg hadde trodd. CSV-kit er egentlig et Python-bibliotek, noe som innebærer at Python må være installert for å få det til å virke. Python kan enkelt installeres fra [Pythons hjemmeside](#).

-
Når Python er installert, kan man bruke pip for å installere CSV-kit. Enkle instruksjoner for hvordan man går frem for å gjøre dette finnes i dokumentasjonen jeg har lenket til. Etter CSV-kit er installert, kan man plutselig begynne å «slize og dize» i dataene. Fjerne kolonner, gjøre enkle sql-lignende spørringer for å «intervjue» dataene, og ikke minst eksportere dataene til en rekke kjente formater, eks. JSON, excel, mysql etc. Rett og slett et «must have» for journalister som jobber med store datamengder.

Jeg deltok dessuten på flere sesjoner om såkalt «Crowd sourcing», som går ut på at journalisters research baserer seg på datainput fra lesere og/eller myndigheter. Journalistenes oppgave blir da å fasilitere for at dataene kommer inn, deretter kvalitetssikre og behandle dataene og presentere dem. Et prosjekt baserte seg på narkotika-dødsfall i en amerikansk by, hvor journalisten fikk lesernes hjelp til å fortelle historien om hvem den avdøde var (+ bilder, etc.). Sesjonene inspirerer meg til å gjennomføre lignende prosjekter på kommune-nivå i Norge. Jeg har flere konkrete ideer til

prosjekter å gjennomføre som følge av dette.

[Tipsheets & Links fra årets konferanse](#) ligger som i fjor på konferansens nettsted. Dette er i praksis en stor kunnskapsbank til alle som ønsker å få gratis opplæring og/eller henvisning til verktøyer innenfor datastøttet journalistikk.

Jeg var på en svært nyttig sesjon om bruk av Google-verktøyet [Open Refine](#). Dette er et verktøy jeg lærte om på fjorårets konferanse, men i år deltok jeg på en mer avansert klasse. Her lærte vi å bruke såkalte regulere uttrykk (Regular expressions, aka Regex) i kombinasjon med verktøyet. Veldig nyttig for «datavasking» av ustrukturerte datasett, som ofte er en problemstilling i datastøttet journalistikk.

Et av verktøyene som ble undervist i flere ganger, og som også er utviklet av og for journalister, heter [Tabula](#). Det er et verktøy som lar deg hente ut tabeller fra pdf-filer hvor teksten er lesbar. Forsøkt å kopiere tabeller fra PDF-før og lime inn i Excel? Det går som regel ikke, men med Tabula går det som bare det.

Jeg takker igjen for mottatt stipend, og anbefaler flere fagpressemedarbeidere å søke!