samediff

Skriv Din Egen Sang

Brug SameDiff til at analysere den lyriske stil fra to musikere og opfind en duet til dem

Hvad er SameDiff?

SameDiff sammenligner to eller flere tekstfiler, og fortæller dig, hvor ens eller forskellige de er. Det hjælper dig med at se forskelle og ligheder i de ord, der bruges i filerne, så du kan lære om kvantitativ tekstanalyse. Øvelsen hjælper deltagerne med at opbygge datakompetence ved at sammenligne teksterne fra to musikere og opfinde en ny sang, som sammenfletter de to.

Læringsmål

- Styrke evnen til at analysere tekstdata.
- Forståelse for, at sammenligning er en effektiv måde at finde datahistorier på.
- Viden om hvilke former for spørgsmål man kan/bør stille til tekstdata.
- Forståelse for, at algoritmisk analyse kan afsløre interessant information om dine data.

Tidsramme

30 til 45 minutter

Deltagere

3 – 100 personer. Alder: 12+.
Udviklet for 6.-9. klasse, ungdomsuddannelser, nyhedsorganisationer, non-profit og frivillige organisationer. Tidligere erfaring med data er ikke nødvendig.

Lokale

- Projektor og computer.
- Mulighed for at kunne arbejde i grupper af tre omkring en computer.
- Store borde, gulv- eller vægplads til placering af post-its og tegninger.

Udstyr

- En computer pr. gruppe
- Stort papir/flip-overs
- Skriveredskaber: Store tuscher, farveblyanter o.lign.

Øvelsen

Problemløsning

Det er svært at analysere en omfattende samling tekster manuelt. En måde kan være at sammenligne den med en anden samling, eller sammenligne dele af begge. Datamatikere har opfundet værktøjer, der kan gøre det lettere ved hjælp af opskrifter eller "algoritmer", der kan sammenligne de to samlinger for dig. SameDiff anvender nogle af disse algoritmer, så du lettere kan sammenligne to store samlinger med hinanden.

Hav de gode eksempler med

Store tekstdatasæt er overalt omkring os. Man kan downloade alle Hillary Clintons e-mails, som hun sendte mens hun var udenrigsminister, lækkede diplomatiske dokumenter fra wikileaks eller alle Sherlock Holmes bøgerne fra Gutenberg projektet. Analyse og visualisering af disse store tekstsamlinger er idag en helt almindelig ting at gøre, både for alvor og for sjov. Man kan f.eks. vise Jaz Parkinsons "Color Signatures", der sammenligner farver nævnt i forskellige bøger (http://jazparkinson.tumblr.com), og Tahir Hemphills "Rap Research Lab" (http://rapresearchlab.com).

Øvelsen (fortsat)

Introducer værktøjet

Åben SameDiff (https//databasic.io/samediff) og vælg Beyoncé og Aretha Franklin. Ved resultatsiden forklarer du, at den venstre kolonne viser de ord, som kun Beyoncé bruger, og den højre side viser de ord, som kun Aretha bruger. Det er deres forskelle. Den midterste kolonne viser de ord, de har tilfælles. Gør deltagerne opmærksom på, at der ved toppen af resultatsiden står, disse to dokumenter er "ret ens". SameDiff bruger en algoritme kaldet "cosine similarity (cosinus lighed)" for at give dig en sammenlignelighedsværdi. Cosine similarity skaber en liste af ord fra Beyoncé og fra Aretha Franklin. Den tæller, hvor ofte hvert ord forekommer i hvert dokument, hvorefter den sammenligner, hvor tæt de to lister er på hinanden. Dette er en effektiv algoritme til tekstanalyse.

Igangsæt øvelsen

1. Deltagerne har 15 minutter.
2. Deltagerne skal være i 3-mandsgrupper.
3. Hver gruppe bruger SameDiff til at sammenligne sangteksterne fra to kunstnere. Den slags kunstnersamarbejder blandt musikere er meget populære, og opgaven er derfor at vælge to kunstnere og forestille sig, hvordan en duet mellem dem ville se ud.
4. Hver gruppe skriver deres sangtekst på en flipover.

Grupperne får bonuspoint hvis: (a) deres sang rimer og/eller (b) de finder på en melodi at synge deres sang til og/eller (c) De synger sangen for resten af klassen.

Præsentationsrunde

Giv hver gruppe 1 minut til at fremlægge deres sang. Gode spørgsmål og temaer man kan tage op under diskussionen:

- Lagde I mærke til nogle fælles temaer?
- Er de nye tekster mere interessante, hvis de kommer fra kunstnere, hvis stile er meget forskellige?
- Sammenligning er en effektiv måde at finde datahistorier på.
- At arbejde med data kan være sjovt!

Husk

- Vi bruger algoritmer hver eneste dag. Hvis du fx taber dine nøgler, kører du en algoritme for at lede efter dem – først kigger du i dine lommer, så kigger du på bordet, ved døren osv.
- En Cosine similarity (cosinus lighed) på 1.0 betyder præcis den samme; 0 betyder fuldstændigt forskellige.

Følgende termer skal introduceres

Algoritme
En serie af skridt som du (eller en computer) gør for at løse et problem.

Cosine Similarity (Cosinus lighed)
Cosine similarity værdien bruges til at fortælle, hvor sammenlignelige to dokumenter er, baseret på antallet af gange specifikke ord bliver brugt i dem.

Eksempel