

1. Introduktion

I denna lärresurs får du se hur det går till att ladda ner stora mängder data från Twitter.

Lärresursen går igenom:

- hur du skaffar behörighet till Twitterdata,
- hur ett program för att hämta Twitterdata är uppbyggt,
- hur du modifierar programmets kod för att göra olika sökningar och
- hur du kör programmet

Lärandemål

Lärandemålen för denna lärresurs är att få

- ökad förståelse för hur data från sociala medier är strukturerad
- grundläggande kännedom om hur programmeringskod ser ut och fungerar
- ökad förståelse för hur metoder för datainsamling från webbsidor praktiskt går till
- kunskaper för att kritiskt kunna granska artiklar och andra arbeten där data från sociala medier, i synnerhet Twitter, används som forskningsdata
- en ingång till att själv kunna hämta stora mängder data från Twitter.

Vad är webbskrapning?

” Webbskrapning (web scraping) – insamling av information direkt från webbsidor med hjälp av speciella program. Exempel på webbskrapning är de tjänster som gör automatiska prisjämförelser mellan postorderföretag på internet. Det kallas för skrapning därför att programmen hämtar informationen från webbsidor, inte från de bakomliggande databaserna. Webbskrapning påminner om den indexering som sökmotorer gör när de samlar in innehållet på webbsidor. Skillnaden är att sökmotorer samlar in allt de hittar på webbsidorna, medan program för webbskrapning letar efter särskilda typer av information.”

Computer Sweden, IDG. 2018. *webbskrapning*. I: IT-ord, <https://it-ord.idg.se/ord/webbskrapning/> [Hämtad 2021-12-08]

Syftet med webbskrapning är alltså att med hjälp av ett datorprogram ladda ner data från olika webbplatser. Anledningen till att forskare använder denna automatiserade metod istället för att använda webbtjänstens sökfunktion och spara ned data från resultaten manuellt är att ett program kan hämta ner och hantera större mängder data på betydligt kortare tid än om ett manuellt arbetssätt hade använts.

Någonstans mellan små datamängder och *big data*

Digitaliseringen av samhället har möjliggjort att samla in och analysera stora mängder data på ett sätt som förut inte varit möjligt. *Big data* är ett begrepp som brukar användas om datamängder som är för stora för att bearbeta med traditionella databasmetoder.

Den data du samlar in med hjälp av webbskrapning befinner sig någonstans mittemellan den manuella datainsamlingen och insamling av *big data*; den är enklare att samla in med hjälp av ett program, men du kan förmodligen hantera datamängden med hjälp av digitala verktyg som inte kräver någon större beräkningskraft.

Digitala verktyg och metoder ger oss sätt att angripa forskningsfrågor på nya sätt och göra analyser som annars hade varit alltför tidskrävande (t.ex. genom att använda *text mining*, en metod för att upptäcka mönster och samband i ostrukturerad textdata).

Lärresursens begränsningar

Eftersom både Twitters Developer Account, Python och Pythonbibliotek uppdateras med jämna mellanrum och nya versioner kräver ändringar i kod, sökformuleringar och uppdaterad programvara så är denna lärresurs *inte* en regelrätt guide till hur du själv genomför webbskrapning från Twitter. Detta eftersom lärresursen inte kommer att uppdateras i takt med tjänsterna som används.

Däremot är kunskaperna kring data- och kodstruktur och det generella tillvägagångssättet överförbara även till nyare programversioner, samt till stor hjälp för att förstå hur (hämtning av) data från webben fungerar generellt.