

Arbetsprov

Bakgrund: OpenAlex är öppen datakälla med metadata om vetenskaplig publicering. Den innehåller information från flera internationella databaser och består av flera hundra miljoner entiteter (tidskriftsartiklar, bokkapitel, författare, organisationer, ämnesord, identifikatorer m.m.)

I dagsläget använder KIB licensierat metadata för att analysera Karolinska Institutets (KI) vetenskapliga publicering utifrån olika parametrar. En öppen datakälla så som OpenAlex skulle potentiellt kunna svara på samma frågor. Det finns ett strategiskt värde i att öppet kunna dela resultat av den här typen av analyser. Vi vill därför utforska datats kvalitet och eventuella brister.

En grundförutsättning är att kunna identifiera publikationer som producerats av forskare verksamma vid KI, dvs kunna knyta publikationer till KI som organisation.

OpenAlex:

Website: <https://openalex.org/>

Documentation: <https://docs.openalex.org/>

API: <https://docs.openalex.org/api>

Uppgift:

Hämta alla publikationer med unika DOI-nummer (identifikator för publikation) med författare från Karolinska Institutet och Karolinska Universitetssjukhuset. Begränsa resultatet till att omfatta publikationer av typen 'journal-article' och publikationsår '2020'. Följande information om varje artikel ska hämtas:

- artikeltitel
- publiceringsdatum
- författare
- institution
- förlag

Organisera datat på ett överskådligt sätt

Du får använda valfritt programmeringsspråk för att lösa denna uppgift.

Vid presentationstillfället går vi tillsammans igenom koden och resonerar kring hur du har löst uppgifterna.

Lycka till!