

EXAMENSARBETE Interactive Iterative Patent Search**STUDENT** Daniel Jogstad**HANDLEDARE** Pierre Nugues (LTH), Fredrik Edman (LTH), Henrik Benckert (Mindified)**EXAMINATOR** Jacek Malec (LTH)

Interaktiv Iterativ Patentsökning

POPULÄRVETENSKAPLIG SAMMANFATTNING **Daniel Jogstad**

Vid sökning efter matchande dokument i stora patentdatabaser används metoder för att effektivisera sökningen. Detta arbete har undersökt möjligheten att förbättra resultat från en sökning med hjälp av Sentence embeddings i en för användaren interaktiv, iterativ process.

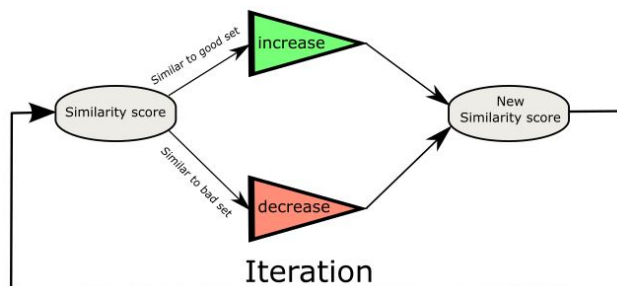
De senaste åren har användandet av neurala nätverk blivit allt vanligare för att hitta semantiska likheter mellan texter. Dessa nätverk blir tränade på en stor mängd data för att skapa numeriska vektorrepresentationer, så kallade embeddings, av ord och meningar. Vektorer kan enkelt jämföras matematiskt för att se hur lika varandra de är. Därmed reducerar man den, för en maskin, komplicerade processen att säga hur lika två texter är varandra, till ett enklare matematiskt problem.

Eftersom det inte alltid finns ett exakt svar på hur lika olika texter är varandra, är resultatet av en jämförelse väldigt subjektiv. Detta avspeglar sig i att det uppstår skillnader i vad ett program tycker är den mest lika träffen till en text och vad användaren tycker.

I detta examensarbete har jag använt mig av en databas med titlar och abstrakt från 310 024 patent. Varje patent har representerats av en sammanslagen titel och abstrakt översatt till embedding-format av ett SBERT-nätverk. Sökningar kan då ske, där man jämför en embedding från ett input-patent med alla andra embeddings i databasen. Resultatet från en sökning är en sorterad lista med de, för input-patentet, matematiskt mest lika träffarna.

Examensarbetet har haft som syfte att testa om det går att ranka om resultaten från en så-

dan sökning efter vad användaren tycker om de högst rankade patenten. Genom att interaktivt markera för programmet vilka träffar som var bra och dåliga, får man ett underlag för en ny sökning. Sökningen kan därmed göras om från början, men med skillnaden att man nu har en större mängd patent som man vill att de nya träffarna skall bli lika, och även en mängd patent som man vill att de inte skall bli lika. Genom att göra detta i flera steg blir processen iterativ.



Jag har testat 2 enkla matematiska metoder för att ranka om träffar. En testgrupp har sedan testat och utvärderat resultat från metoderna i 3 iterationer för 5 olika input-patent. Resultaten visar att testgruppens betyg för de bästa träffarna växer med stigande antal iterationer och att metoderna därmed fungerar.