

FÖRSLAG PÅ EXAMENSARBETE

Text Mining av statistiska data om migration

Bakgrund

Migrationsverket förfogar över stora mängder ostrukturerade textdata. Inom ramen för EU-projektet Prisma försöker Migrationsverket se till att den information som återfinns i dessa data på ett bättre sätt kan tas om hand och tillgängliggöras för myndighetens analysarbete rörande till exempel migrantprofiler, rutter och drivkrafter bakom migration.

Ett av spåren i projektet behandlar de omvärldsanalyser, i form av bland annat rapporter, som Migrationsverket dagligen får skickade till sig från andra myndigheter runt om i Europa. Dessa analyser innehåller en hel del statistiska data, exempelvis hur många personer som anlänt till ett specifikt land under en viss tidsperiod, men dessa data presenteras ofta i den löpande texten, och inte i tabeller. Att sammanställa dessa data är av stort intresse för Migrationsverket, men att göra detta manuellt är ett tidsödande arbete.

Projektbeskrivning

Arbetets syfte är att föreslå, utveckla och utvärdera en eller flera möjliga lösningar för automatisk extraktion av statistiska data rörande migration ur löpande text. Lösningen ska kunna identifiera data som är av intresse för verksamheten och extrahera dessa data med referens till vad de givna siffrorna beskriver. De föreslagna lösningarna utvärderas på lämpligt sätt, exempelvis med avseende på precision och täckning relativt till en guldstandard.

Kund

Migrationsverket, Norrköping

Kontakt

Marco Kuhlmann, marco.kuhlmann@liu.se

Studentprofil

Kunskaper inom Natural Language Processing och Text Mining, till exempel genom kurser såsom TDDE09 och TDDE16