

ADVANCED FUZZY MATCHING IN THE TRANSLATION OF EU TEXTS

Margita Šoštarić

Sammanfattning

Idag är översättningsprocessen otänkbar utan översättningsverktyg. Eftersom översättningsminnen utgör en central del i hur dessa verktyg fungerar, uppstår naturligtvis frågan om hur man bäst kan utnyttja dem för att göra översättningsprocessen lättare och snabbare. Målet med den här undersökningen är att ta reda på om det finns effektivare metoder att identifiera användbara översättningsförslag än dem som nuförtiden används av översättningsverktyg. Det är av skärskilt intresse att skärskåda om en tillämpning av mer komplexa likhetsalgoritmer och lingvistiska egenskaper i jämförelseprocessen leder till någon betydelsefull förbättring av förslagets kvalitet. I experimentet används syntaktiska träd erhållna genom parsning av EK:s översättningsminne (DGT-TM) och ett antal olika konfigurationer tillämpas för att jämföra de data som träden innehåller. Algoritmerna testas både individuellt och genom att man kombinerar dem med hjälp av en regressionsmodell. De utvalda segmenten utvärderas genom automatisk utvärdering, baserad på korrelationer och automatiskt beräknade medelvärden. Dessutom används mänsklig utvärdering, baserad på korrelationer framkomna ur enkätsvar. Syftet med arbetet är att avgöra om någon av de undersökta metoderna skulle kunna implementeras i kommersiella översättningsverktyg för att förbättra översättningsprocessen.