

Word2Vec:

Noggrannhet: Högre än Random Indexing, särskilt vid högre dimensionalitet (t.ex 100 och 1000).

Beräkningstid och resurser:

Långsammare och mer resurskrävande eftersom det tränar hela vokabulären med neural nätverksbaserade tekniker.

Random Indexing:

Noggrannhet: Lägre än Word2Vec, men tillräcklig för enklare uppgifter.

Beräkningstid och resurser: Snabbare och kräver mindre resurser eftersom det endast skapar representationer för relevanta ord.

Förklaring av skillnader

Word2Vec bygger relationer mellan ord genom hela vokabulären, vilket leder till mer precisa representationer men är mer tidskrävande.

Random Indexing skapar representationer med enkla operationer och fokuserar endast på ord i uppgiften, vilket är snabbare men mindre exakt.

just nu koden fungerar att alltid använda två ord på vardera sidan (fönsterstorlek = 2). för att lösa begränsningen att fönsterstorleken är fast kan man istället lägga till en parameter som gör fönsterstorleken justerbar. genom att ändra hur embadding fungerar kan modellen göras mer flexibel och stödja olika fönsterstorlek istället för att vara begränsad till 2.

Mohraj-0@student.ltu.se

Jawad Rajabi