

EXAMENSARBETE Locally Generated Unique Identifiers for Geospatial Data**STUDENT** Tim Jangefeldt**HANDLEDARE** Patrick Cording (Apple Inc.) Jonas Skeppstedt (LTH)**EXAMINATOR** Flavius Gruian (LTH)

Hur mycket information behövs för att unikt identifiera alla världens vägar?

POPULÄRVETENSKAPLIG SAMMANFATTNING **Tim Jangefeldt**

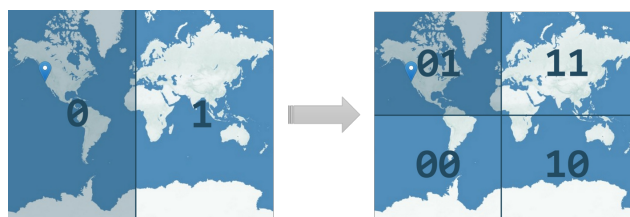
I vårt arbete presenterar vi en geografisk algoritm som genererar IDn på vägar i OpenStreetMap utifrån vägens utformning. Med en kombination av geografiskt läge, längd och vägvinkel lyckades jag identifiera 224 miljoner vägar utan en enda kollision med 71 bitar.

Ponera att du vill komma på ett sätt att ge varenda människa i världen ett unikt förnamn, utan vetskap om vad andra människor heter, samt hitta den kortaste namnlängden som uppnår detta. Din uppgift blir att komma på en optimal kombination av egenskaper hos personen som tillsammans är helt säregen. Eftersom alfabetet har 29 bokstäver och jordens befolkning ligger på ca åtta miljarder kan du uppnå detta med minimalt sju bokstäver teoretiskt ($29^7 > 8md$), men i verkligheten blir det mer komplicerat. Du kan exempelvis välja att representera personens ögonfärg med första positionen, längd med nästa, ålder med den tredje och fjärde kombinerad, och vikt med den femte och så vidare. Du kanske också upptäcker regler runt namngivningen såsom att fler än två av samma bokstav inte får förekomma i följd. Problemet blir snabbt komplext.

I detta examensarbete försöker vi uppnå detta på alla världens vägsegment i en karta, och vårt resulterande ID består av ettor och nollor (bitar) istället för bokstäver. På så sätt slipper man beroende av en central enhet som håller koll på vilka ID som använts hittills, där kommunikation går långsamt. Andra metoder som sommet UUID förlitar sig på sannolikhet, och ID unika till följd av storleken. Vårt ID är kortare, 71 mot

128 bitar, och har inga slumpmässiga element. Vi fann även att det kan användas för att snabbt verifiera om en nyskapad väg redan existerar, då en kollision uppkommer.

Egenskapen där mest antal bitar behövdes var geografiskt läge. Här användes en befintlig algoritm Geohash, där man kan beskriva rutor i världen vars storlek beror på hur många bitar som används. Eftersom världen är ofantligt stor behövs mycket information för att beskriva positioner till centimeternivå.



Vi lyckades namnge varenda väg med 71 bitar, utan att inkludera ett enda slumpmässigt element. Det visar sig även att algoritmen är mycket effektiv och snabbare än en central enhet, då kommunikation till denne måste ske över hela världen. Vi föreslår därmed en lösning som uppnår en markant mindre storlek än vanligt använda UUIDs, och som är funktionella och effektiva i jämförelse med en central enhet.