

## PROJEKT «INGLÄMNLAGARE»

Riksantikvarieämbetets lämningss databaser innehåller mängder av fynduppgifter för intressanta arkeologiska analyser. Delar av databaserna uppvisar dock en struktur som försvårar sådana undersökningar. År 2018 påbörjade RAÄ migrationsprocessen från Fornminnesinformationssystemet till ett nytt sådant – Kulturmiljöregistret. Vid övergången, som var ämnad åt att effektivisera sökmekanismer och framtidssäkra plattformen, överfördes dock inte alla delar från FMIS på samma enhetliga sätt. Ett exempel på detta är spalten *ing\_lamn* i arkeologiska fornlämningss databaser som kan laddas ner från KMR. Denna spalt med ingående lämningar innehåller extrainformation rörande många fyndplatser, som dock utgörs i de nya databaserna av ett hopbuntat fält (rimligen en kvarleva från processen med vilken dessa värden sparades inför överföring från FMIS) av följande struktur:

```
[{"lamningstyp":"Stensättning","antal":3,"egenskap":"Form: Rund, Konstruktion: Stenfyllt"}, {"lamningstyp":"Stensättning","antal":1,"egenskap":"Form: Rund, Konstruktion: Övertorvad"}, {"lamningstyp":"Boplats","antal":1,"egenskap":null}]
```

För att göra en sådan uppsättning mer tillgänglig för överblick, sökning och statistisk analys, behöver dess data brytas ut på motsvarande sätt:

| Stensättning - Form: Rund;<br>Konstruktion: Stenfyllt | Boplats | Stensättning - Form: Rund;<br>Konstruktion: Övertorvad |
|---|---------|--|
| 3   | 1       | 1  |

Detta kan göras med hjälp av *Inglämnlagaren*. Programmet är skrivet i språket Python, och dess interna bearbetning av spalten *ing\_lamn* görs med så kallade reguljära uttryck. Reguttryck används flitigt inom datorlingvistik som ett mycket kraftfullt verktyg för arbete med text på maskinväg, och kan brukas för att fånga samt ändra godtyckliga typer av textstrukturer. Som ett enkelt exempel: om reguttrycket `\d+` appliceras på strängen `{"lamningstyp":"Boplats","antal":6,"egenskap":null}`, kommer den enda siffran i strängen att brytas ut och svaret blir – talet 6. Meningen med reguttryck är följaktligen inte att i förväg söka efter en viss text, utan ett visst *textmönster*. Programmeringsspråket Python använder anropet `re.` för att aktivera sitt interna maskineri för reguljära uttryck, och du kan titta i programmets kod för att försöka lista ut vad de olika uttrycken skulle kunna fånga för text. I övrigt innehåller koden en del kommentarer, så vid intresse för programmering överlag kan det vara intressant att gå igenom den och få en uppfattning om programmets ingående delar.

För att kunna bryta ut informationen ur spalten *ing\_lamn*, behöver infilen vara av typen CSV samt även inneha spalten *FID*, vilken används för att hålla reda på raderna i filen. Det finns flera sätt att få igång programmet beroende på vilken variant av *Inglämnlagaren* som finns tillgänglig.

- Enklast är det vid tillgång till en kompilerad version (tex. *Ing.exe* för Windows): dubbelklicka på programfilen för att starta.
- Om endast källfilen *Ing.py* finns att tillgå, men datorn har en uppsatt Python-miljö, kan den köras direkt i operativsystemets kommandotolk genom att exempelvis ställa sig i källfilens mapp och skriva `python Ing.py`. Vid behov går även *Ing.exe* bra att starta genom kommandotolken.

- Ifall det varken finns tillgång till en kompilerad version eller en Python-miljö, går det dessutom att få igång Inglämnlagaren genom Qvantums **geoinfosystem**. Öppna dess kommandotolk (menyerna *Plugin* → *Pythonkonsol*) och mata in raden...

```
ing = r"Sökväg\till\Ing.py"
```

...där sökvägen inom citattecken får ändras till där Inglämnlagares källfil ligger på din maskin (tex. `ing = r"C:\Ny mapp\Ing.py"`). Och sedan ytterligare en sträng för att starta:

```
import subprocess; subprocess.run("cmd /c python \"" + ing + "\""); print("\n\n\u004B\u006C\u0061\u0072.", end = "")
```

I programmets fönster anges sedan sökvägen till CSV-infilen som ska bearbetas (alternativt bara dess namn om den ligger i samma mapp som programmet). Utfilen som fås efter bearbetningen motsvarar utseendemässigt infilen – med den skillnaden att tabellen får nya spalter med lämningsantal på höger sida, benämnda efter kombinationen av *lämningstyp* och *egenskap* ur spalten `ing_lamn` på samma rad (samt efter de ensamstående spalterna med samma namn om dessa också fanns i infilen). Observera att CSV-utfilen, som grafiskt ser ut som en tabell, separerar spalterna internt med kommatecken: beroende på vilket tabellprogram utfilen öppnas i (därav i Qvantum självt), kan alltså *spaltavgränsaren* behöva ställas in till att *endast* vara *komma* (och inte tex. semikolon). Ytterligare ett tips är att om infilen saknar spalten FID, kan detta enkelt åtgärdas genom att spara om tabellen som en CSV-fil från Qvantum – då skapas en FID-spalt automatiskt.

Lycka till med analyserna!

[illegible]