Large Scale Lookups From and End-User and a Programmer Perspective

Peter Stegehuis and Naxin Zheng
Westat, USA

IBUC - October 2023

Overview

- Lookups in Blaise
  - Trigrams
- Location-specific lookups
- Adding to the look and feel

Blaise lookups with trigrams

- Creating trigram lookup file

  - Datamodel with trigram secondary key

  - Three characters snippets, lookup file indexed

- Using in a lookup

  - Interviewer types in search string

  - That gets chopped up the same way

  - Immediate results

  - More 'resistant' to typos than e.g. SQL 'LIKE' search

Location-dependent lookup files

For many purposes a lookup will be the same for all interviews

- Example: Prescribed medicines

In our case, we needed to show only relevant medical providers

- For an interview in New York, a Family Doctor in California is not relevant

- Subdivision or filtering needed

- Seven million provider records!

Implementation

Use zip code (postal code), and include all records within a 100 mile radius

- Pro: very localized

- Con: There are 42 K zip codes in the US

Too many lookup files, and too big (20 TB) to create all and install on interviewer's laptop

- Solution: create on laptop as needed

Issues

Even such a local lookup file can get big

- Possibly hundreds of thousands of records

- Takes time to create

Manipula: superfast in processing text files

- Funny finding: two-step process faster than one single setup

- Size of a source file: too big for in-memory processing

Presentation

Presentation

- Checkbox for filtering

- Details of selected record at the top

- Interviewer instructions to guide successful search

Conclusions

- The use of trigram lookups in Blaise is very powerful

- Lookups with thousands of records can be complicated for field interviewers
  - make lookups as user-friendly as possible
  - with options and pointers for efficient use

- It pays off to find the most efficient way to create lookups, at least if you have to create many
  - Sometimes better results with what may seem a counter-intuitive method