PackageBuilder: From Tuples to Packages

Matteo Brucato*, Rahul Ramakrishna*, Alexandra Meliou* and Azza Abouzied§
*UMass Amherst §New York University, Abu Dhabi
http://packagebuilder.cs.umass.edu

**Motivation**

I want a meal plan for the day:
- All gluten-free meals,
- No more than 2.5 kcal in total,
- The more protein, the better!

**PackageBuilder**

PackageBuilder is a system that extends query engines to support package generation.

A package is a collection of tuples with certain global properties defined on the collection as a whole.

**Language Specification: PaQL**

```
SELECT PACKAGE(R) AS P
FROM Recipes R REPEAT 0
WHERE R.gluten = 'free'
SUCH THAT count(*) = 3 AND sum(cal) <= 2500
MAXIMIZE sum(proteins)
```

**Query Evaluation**

1. **Using MIP solvers**

\[
\begin{align*}
\text{COUNT}(\ast) &= 3 \\
\sum_{i=1}^{n} t_i &= 3 \\
\text{SUM}(\text{cal}) &\leq 2500 \\
\sum_{i=1}^{n} t_i \otimes u_{i,\text{cal}} &\leq 2500
\end{align*}
\]

2. **Pruning using cardinality constraints**

Search space is \(2^n\) (with REPEAT 0)\n
\[
I \leq \text{COUNT}(\ast) \leq u \quad \text{reduces the search space to:}
\]

\[
\begin{align*}
\binom{n}{1} + \binom{n}{2} + \cdots + \binom{n}{u-1} + \binom{n}{u}
\end{align*}
\]

\[
20 \leq \text{SUM}(\text{cal}) \leq 2500 \quad \text{can be converted into:}
\]

\[
\left[ \frac{20}{\text{MAX}(\text{cal})} \right] \leq \text{COUNT}(\ast) \leq \left[ \frac{2500}{\text{MIN}(\text{cal})} \right]
\]

3. **Searching using heuristics**

Start with a candidate package \(P_0\)

Perform a local search,
by adding, removing or replacing tuples, e.g.:

```
SELECT T1.recipe, R1.recipe
FROM P0 AS T1, CORE-table AS R1
WHERE 2535 = T1.cal + R1.cal <= 2500
```

**Interactive Exploration**

Users can directly add constraints. An auto-suggest feature helps with syntax

Selecting a constraint shows the rows or columns affected

Highlight values, cell, rows or columns to get suggestions for constraints

Natural language descriptions

The current package’s position in the result space is highlighted

Only packages found so far are visualized. Running indicates incomplete result space