

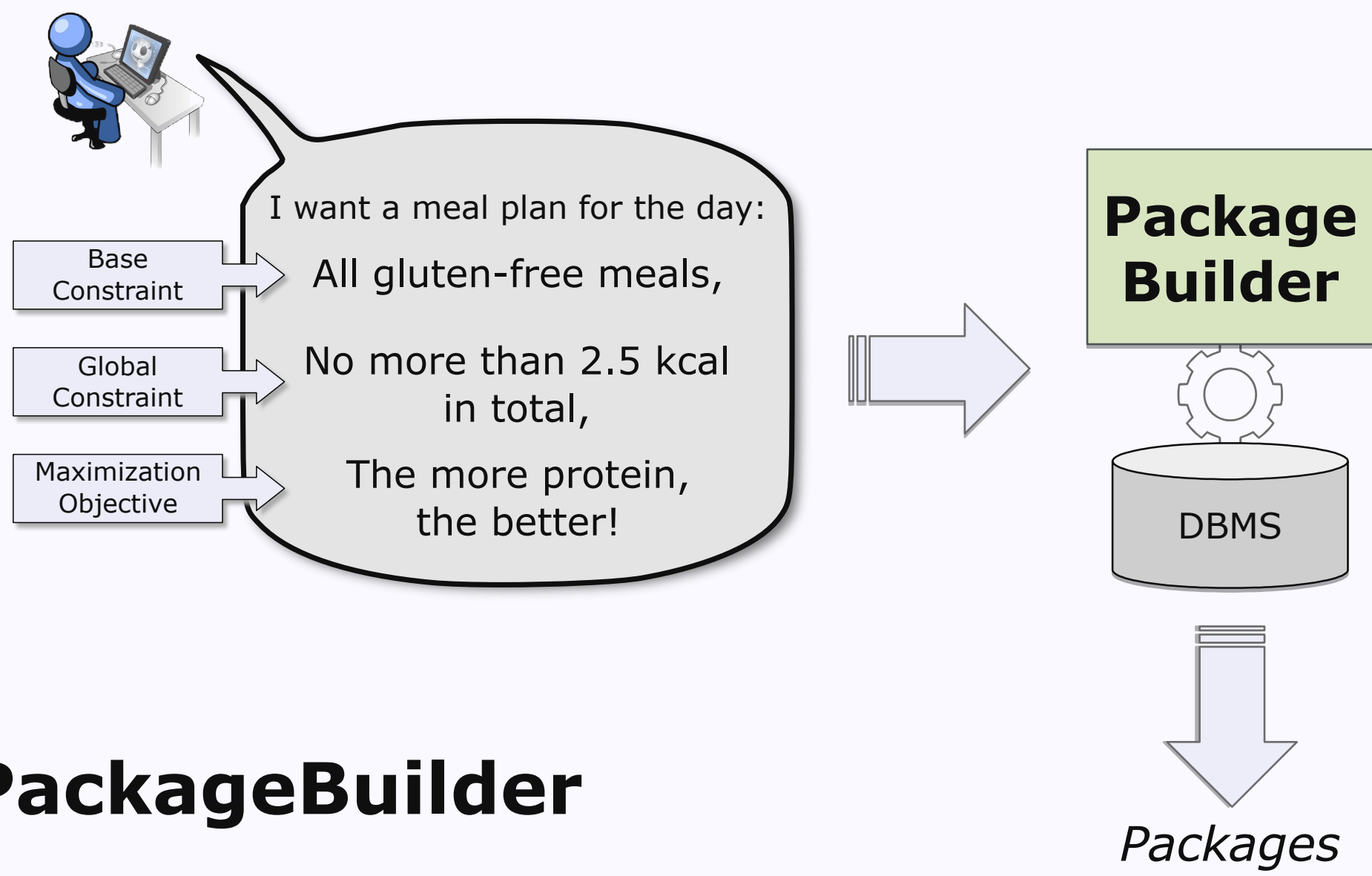
PackageBuilder: From Tuples to Packages

Matteo Brucato*, Rahul Ramakrishna*, Alexandra Meliou* and Azza Abouzied[§]
 *UMass Amherst §New York University, Abu Dhabi

<http://packagebuilder.cs.umass.edu>



Motivation



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.

Meal	Receipe	Calories	Proteins	Gluten
Breakfast	Pancakes Rice Muffins	675	32	free
Lunch	Hamburger	850	51	free
Dinner	Tuna Rice Wrap	630	60	free

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

$$\text{COUNT}(\ast) = 3 \longrightarrow \sum_{i=1}^n t_i = 3$$

$$\text{SUM}(\text{cal}) \leq 2500 \longrightarrow \sum_{i=1}^n t_i \otimes u_{i,\text{cal}} \leq 2500$$

2. Pruning using cardinality constraints

Search space is 2^n (with REPEAT 0)

$l \leq \text{COUNT}(\ast) \leq u$ reduces the search space to:

$$\binom{n}{l} + \binom{n}{l+1} + \dots + \binom{n}{u-1} + \binom{n}{u}$$

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

$$\left\lfloor \frac{20}{\text{MAX}(\text{cal})} \right\rfloor \leq \text{COUNT}(\ast) \leq \left\lfloor \frac{2500}{\text{MIN}(\text{cal})} \right\rfloor$$

3. Searching using heuristics

Start with a candidate package P_0

Perform a **local search**,

by *adding, removing or replacing* tuples, e.g.:

Single-tuple replacement search:

```

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

↑
out of constraint's range

Interactive Exploration

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

Adaptive Exploration

Selecting a constraint shows the rows or columns affected

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

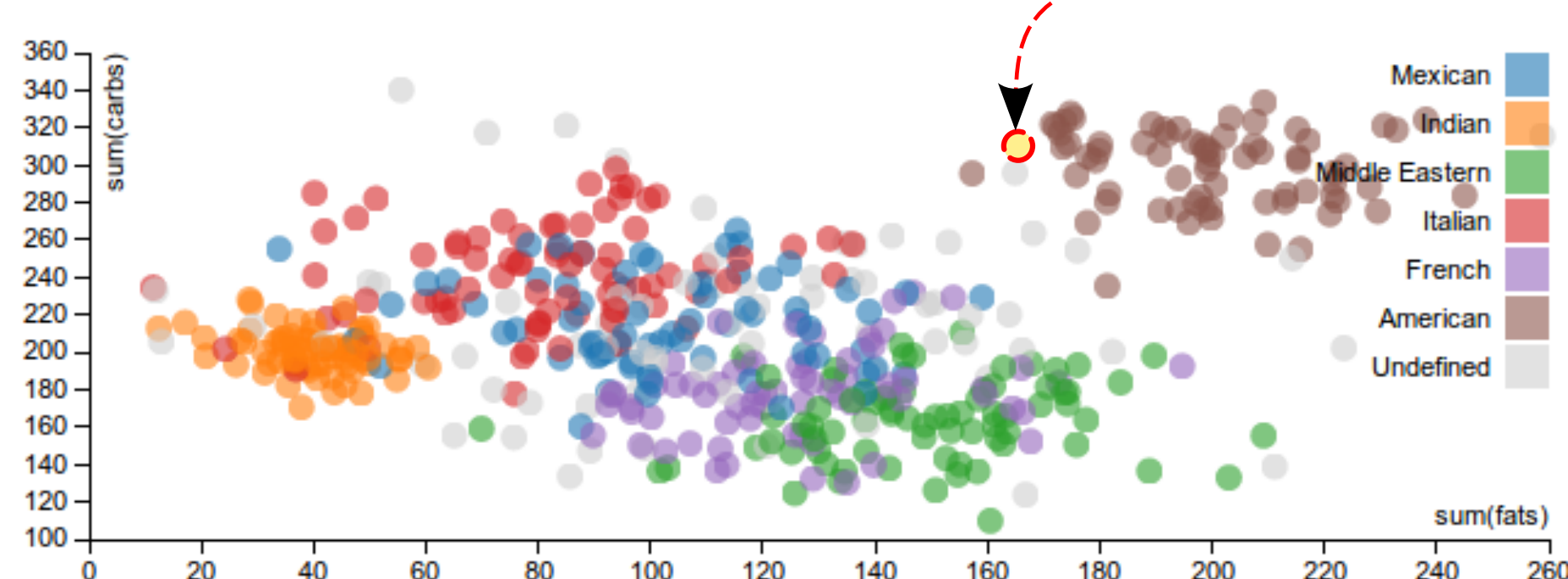
Natural language descriptions

SUGGESTIONS

- + Include if fats >= 12
- + exclude if fats <= 12
- + minimize(fats)
- + exactly 1 for fats == 12
- + avg(fats) >= 12

VISUAL SUMMARY

Packages Represented: 450; Search: Running



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

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