Inductive Meta-Synthesizers

Sumit Gulwani
Researcher @ Microsoft Research

March 2014
ExCAPE Meeting, Berkeley
A Generic Methodology for Inductive Synthesis

1. **Problem Definition**: Identify a vertical domain of tasks that users struggle with.

2. **Domain-Specific Language (DSL)**: Design a DSL that can succinctly describe tasks in that domain.

3. **Synthesis Algorithm**: Develop an algorithm that can efficiently translate examples into likely programs in DSL.

4. **Machine Learning**: Rank the various programs.

5. **User Interface**: Provide an appropriate interaction mechanism to resolve ambiguities.

*CACM 2012: “Spreadsheet Data Manipulation using Examples”, Gulwani, Harris, Singh*
A Generic Methodology for Inductive Synthesis

Some Instantiations

• Syntactic String Transformations (POPL 2011)
  - Flash Fill feature in Excel 2013
• Semantic String Transformations (VLDB 2012)
• Number Transformations (CAV 2012)
• Table Transformations (PLDI 2011)

Advantages (compared to a generic synthesizer)

• Can leverage DSL properties to design efficient synthesizer.
• Easier control over ranking to enable learning from few examples.
• Tighter integration with user interaction model
  - E.g., soft constraints, search guidance

Disadvantages

• Non-constructive: synthesizer requires thinking & implementation.
• Non-modular: DSL is not extensible.
Inductive Meta-Synthesizers

- Alleviate disadvantages of the generic methodology.

- Meta-Synthesizer: A synthesizer for a related family of DSLs that supports a common user interaction model.

- A Generic Methodology for Inductive Meta-Synthesizers
  1. Identify a family of vertical task domains.
     - that allow a common user interaction model.
  2. Design an algebra for DSLs.
     - A DSL is an ordered set of grammar rules (to model ranking).
  3. Design a search algorithm for each algebra operator such that it is compositional and inductive.
Ana Trujillo
357 21th Place SE
Redmond, WA
(757) 555-1634

Antonio Moreno
515 93th Lane
Renton, WA
(411) 555-2786

Thomas Hardy
742 17th Street NE
Seattle, WA
(412) 555-5719

Christina Berglund
475 22th Lane
Redmond, WA
(443) 555-6774

Hanna Moos
785 45th Street NE
Puyallup, WA
(376) 555-2462
FlashExtract: Text File Instantiation

Ana Trujillo
357 21th Place SE
Redmond, WA
(757) 555-1634

Antonio Moreno
515 93th Lane
Renton, WA
(411) 555-2786

Thomas Hardy
742 17th Street NE
Seattle, WA
(412) 555-5719

Christina Berglund
475 22th Lane
Redmond, WA
(443) 555-6774

Hanna Moos
785 45th Street NE
Puyallup, WA
(376) 555-2462
FlashExtract: Text File Instantiation

Ana Trujillo
357 21th Place SE
Redmond, WA
(757) 555-1634

Antonio Moreno
515 93th Lane
Renton, WA
(411) 555-2786

Thomas Hardy
742 17th Street NE
Seattle, WA
(412) 555-5719

Christina Berglund
475 22th Lane
Redmond, WA
(443) 555-6774

Hanna Moos
785 45th Street NE
Puyallup, WA
(376) 555-2462
FlashExtract: Text File Instantiation

Ana Trujillo
357 21th Place SE
Redmond, WA
(757) 555-1634

Antonio Moreno
515 93th Lane
Renton, WA
(411) 555-2786

Thomas Hardy
742 17th Street NE
Seattle, WA
(412) 555-5719

Christina Berglund
475 22th Lane
Redmond, WA
(443) 555-6774

Hanna Moos
785 45th Street NE
Puyallup, WA
(376) 555-2462
FlashExtract: Text File Instantiation

Ana Trujillo
357 21th Place SE
Redmond, WA
(757) 555-1634

Antonio Moreno
515 93th Lane
Renton, WA
(411) 555-2786

Thomas Hardy
742 17th Street NE
Seattle, WA
(412) 555-5719

Christina Berglund
475 22th Lane
Redmond, WA
(443) 555-6774

Hanna Moos
785 45th Street NE
Puyallup, WA
(376) 555-2462
FlashExtract: Text File Instantiation

Ana Trujillo
357 21th Place SE
Redmond, WA
(757) 555-1634

Antonio Moreno
515 93th Lane
Renton, WA
(411) 555-2786

Thomas Hardy
742 17th Street NE
Seattle, WA
(412) 555-5719

Christina Berglund
475 22th Lane
Redmond, WA
(443) 555-6774

Hanna Moos
785 45th Street NE
Puyallup, WA
(376) 555-2462
<table>
<thead>
<tr>
<th>Name</th>
<th>Address</th>
<th>City, State</th>
<th>Phone</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ana Trujillo</td>
<td>357 21th Place SE</td>
<td>Redmond, WA</td>
<td>(757) 555-1634</td>
</tr>
<tr>
<td>Antonio Moreno</td>
<td>515 93th Lane</td>
<td>Renton, WA</td>
<td>(411) 555-2786</td>
</tr>
<tr>
<td>Thomas Hardy</td>
<td>742 17th Street NE</td>
<td>Seattle, WA</td>
<td>(412) 555-5719</td>
</tr>
<tr>
<td>Christina Berglund</td>
<td>475 22th Lane</td>
<td>Redmond, WA</td>
<td>(443) 555-6774</td>
</tr>
<tr>
<td>Hanna Moos</td>
<td>785 45th Street NE</td>
<td>Puyallup, WA</td>
<td>(376) 555-2462</td>
</tr>
</tbody>
</table>
FlashExtract: Text File Instantiation

<table>
<thead>
<tr>
<th>Label 1</th>
<th>Label 2</th>
<th>Label 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ana Trujillo</td>
<td>Redmond</td>
<td>57</td>
</tr>
<tr>
<td>Antonio Moreno</td>
<td>Renton</td>
<td>11</td>
</tr>
<tr>
<td>Thomas Hardy</td>
<td>Seattle</td>
<td>12</td>
</tr>
<tr>
<td>Christina Berglund</td>
<td>Redmond</td>
<td>43</td>
</tr>
<tr>
<td>Hanna Moos</td>
<td>Puyallup</td>
<td>76</td>
</tr>
<tr>
<td>Frederique Citeaux</td>
<td>Redmond</td>
<td>89</td>
</tr>
</tbody>
</table>
FlashExtract: Text File Instantiation

Ana Trujillo
357 21th Place SE
Redmond, WA
(757) 555-1634

Antonio Moreno
515 93th Lane
Renton, WA
(411) 555-2786

Thomas Hardy
742 17th Street NE
Seattle, WA
(412) 555-5719

Christina Berglund
475 22th Lane
Redmond, WA
(443) 555-6774

Hanna Moos
785 45th Street NE
Puyallup, WA
(376) 555-2462
<table>
<thead>
<tr>
<th>Name</th>
<th>Address</th>
<th>City, State</th>
<th>Phone Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trujillo, Ana</td>
<td>Redmond, WA</td>
<td>(757) 555-1634</td>
<td></td>
</tr>
<tr>
<td>Antonio Moreno</td>
<td>515 93th Lane</td>
<td>Renton, WA</td>
<td>(411) 555-2786</td>
</tr>
<tr>
<td>Thomas Hardy</td>
<td>742 17th Street NE</td>
<td>Seattle, WA</td>
<td>(412) 555-5719</td>
</tr>
<tr>
<td>Christina Berglund</td>
<td>475 22th Lane</td>
<td>Redmond, WA</td>
<td>(443) 555-6774</td>
</tr>
<tr>
<td>Hanna Moos</td>
<td>785 45th Street NE</td>
<td>Puyallup, WA</td>
<td>(376) 555-2462</td>
</tr>
</tbody>
</table>
Ana Trujillo
357 21st Place SE
Redmond, WA
(757) 555-1634

Antonio Moreno
515 93rd Lane
Renton, WA
(411) 555-2786

Thomas Hardy
742 17th Street NE
Seattle, WA
(412) 555-5719

Christina Berglund
475 22nd Lane
Redmond, WA
(443) 555-6774

Hanna Moos
785 45th Street NE
Puyallup, WA
(376) 555-2462
<table>
<thead>
<tr>
<th>Name</th>
<th>Address</th>
<th>City</th>
<th>State</th>
<th>Phone</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trujillo, ana</td>
<td>357 21th Place SE</td>
<td>Redmond</td>
<td>WA</td>
<td>(757) 555-1634</td>
</tr>
<tr>
<td>Moreno, antonio</td>
<td>515 93th Lane</td>
<td>Renton</td>
<td>WA</td>
<td>(411) 555-2786</td>
</tr>
<tr>
<td>Hardy, thomas</td>
<td>742 17th Street NE</td>
<td>Seattle</td>
<td>WA</td>
<td>(412) 555-5719</td>
</tr>
<tr>
<td>Berglund, christina</td>
<td>475 22th Lane</td>
<td>Redmond</td>
<td>WA</td>
<td>(443) 555-6774</td>
</tr>
<tr>
<td>Moos, hanna</td>
<td>785 45th Street NE</td>
<td>Puyallup</td>
<td>WA</td>
<td>(376) 555-2462</td>
</tr>
</tbody>
</table>
FlashExtract: An inductive meta-synthesizer

[PLDI 2014; with Vu Le]

- **Family of vertical domains**: Data Extraction from documents of various types e.g., text files, web pages, spreadsheets.

- **User Interaction model**: Give positive/negative examples of rectangular regions on a document.
  - Different colors & nested regions enables data extraction into a data structure with struct/sequence constructs.

- **Algebra**: Decomposable Map, FilterBool, FilterInt, Merge, Pair
TDS (Test-Driven Synthesis): An inductive meta-synthesizer

[PLDI 2014; with Daniel Perelman, Grossman & Provost]

• Family of vertical domains: Transformation of various kinds of data e.g., strings, tables, XML

• User Interaction model: Provide an increasingly sophisticated sequence of input-output examples.

• Algebra: Typed base operators, Conditional, Specialized Foreach and For Loop patterns that can only occur at the top of the program.

• Key Synthesis Strategy: Minimally repair a program to make it pass the next test case.