DeTEXT: Programming By Demonstration

Konstantin Weitz
Jonathan Bragg
Md Tanvir Islam Aumi
Example

Input → Output

<html> → $<$html$>$$
<head> → $<$head$>$$
<title> → $<$title$>$$
<body> → $<$body$>$$

1000 more inputs........
Motivation

- Data Cleaning\textsuperscript{[2]}
- Simple Repetitive Tasks
- Complex Macro Recorders
- PBD is Intuitive
Version Space\textsuperscript{[1]}

\{h"h", h"he", h"hel",
  h"hell", h"hello", h"hello!"\}

= vs("h","hello!")
Version Space Algebra

- Union
- Join
- Transform
Our Version Space Tree

Programs

Program

Action

Search

Insert

Suffix

Prefix

abc

aabcc

aabcc

aabcc

aabcc
Our Version Space Tree

Programs

Program

Action

Search

Insert

Suffix

Prefix
Our Version Space Tree
Hi Grandpac!

Not Grandpac?

$<$html$>$

<head>

<title>DeTEXT</title>

</head>

<body>

<h1>

This is a heading

</h1>

<p>

This is a <strong>paragraph</strong>,

and it has a <a href="test.com"><img src="img.jpg"/></a>

</p>

</body>

$<$html$>$
Challenges

● No Program Trace
● General Solution
  ○ Insertion and Deletion together
  ○ Wide Range of Actions
Extensions

- Tokenization
- Line Numbers
- Deletion
- Probability
Evaluation

- Version Space Size
  - Hypothesis
  - Tree
- User Study
References

1. [Lau et al., MLJ 2003]
   Programming by demonstration using version space algebra

2. [Kandel et al., CHI 2011]
   Wrangler: Interactive visual specification of data transformation scripts.
Questions?
# Finding Insert Positions

<table>
<thead>
<tr>
<th>Source</th>
<th>Destination</th>
</tr>
</thead>
<tbody>
<tr>
<td>abc</td>
<td>aabcc</td>
</tr>
</tbody>
</table>

"_" is Insertion