

Keith Schwarz

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^[2]
- Simple Repetitive Tasks
- Complex Macro Recorders
- PBD is Intuitive

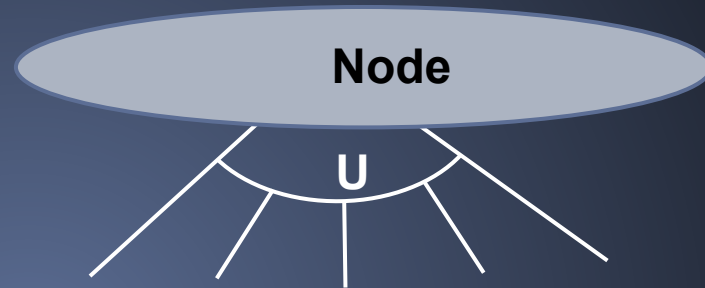
Version Space^[1]

$\{h_{\text{"h"}}, h_{\text{"he"}}, h_{\text{"hel"}},$
 $h_{\text{"hell"}}, h_{\text{"hello"}}, h_{\text{"hello!"}}\}$

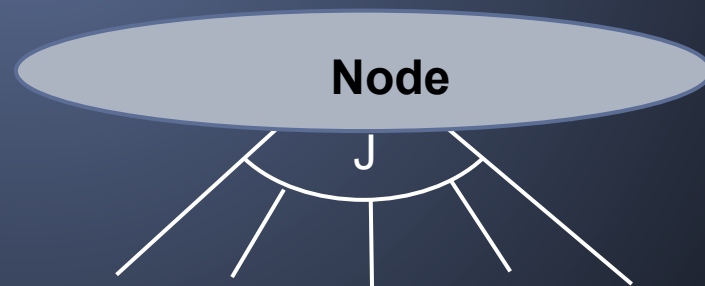
$= vs(\text{"h"}, \text{"hello!"})$

Version Space Algebra

- Union

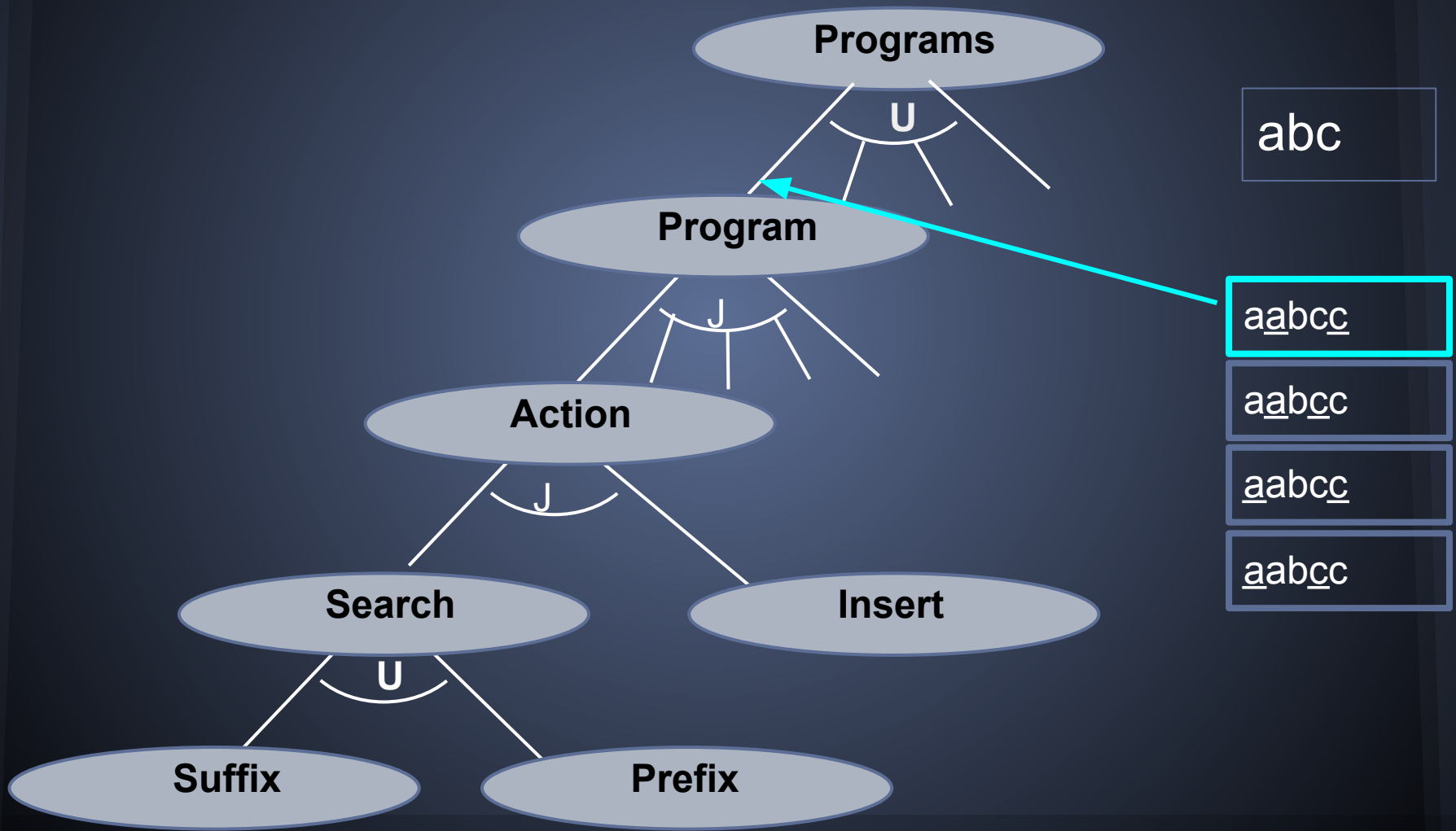


- Join

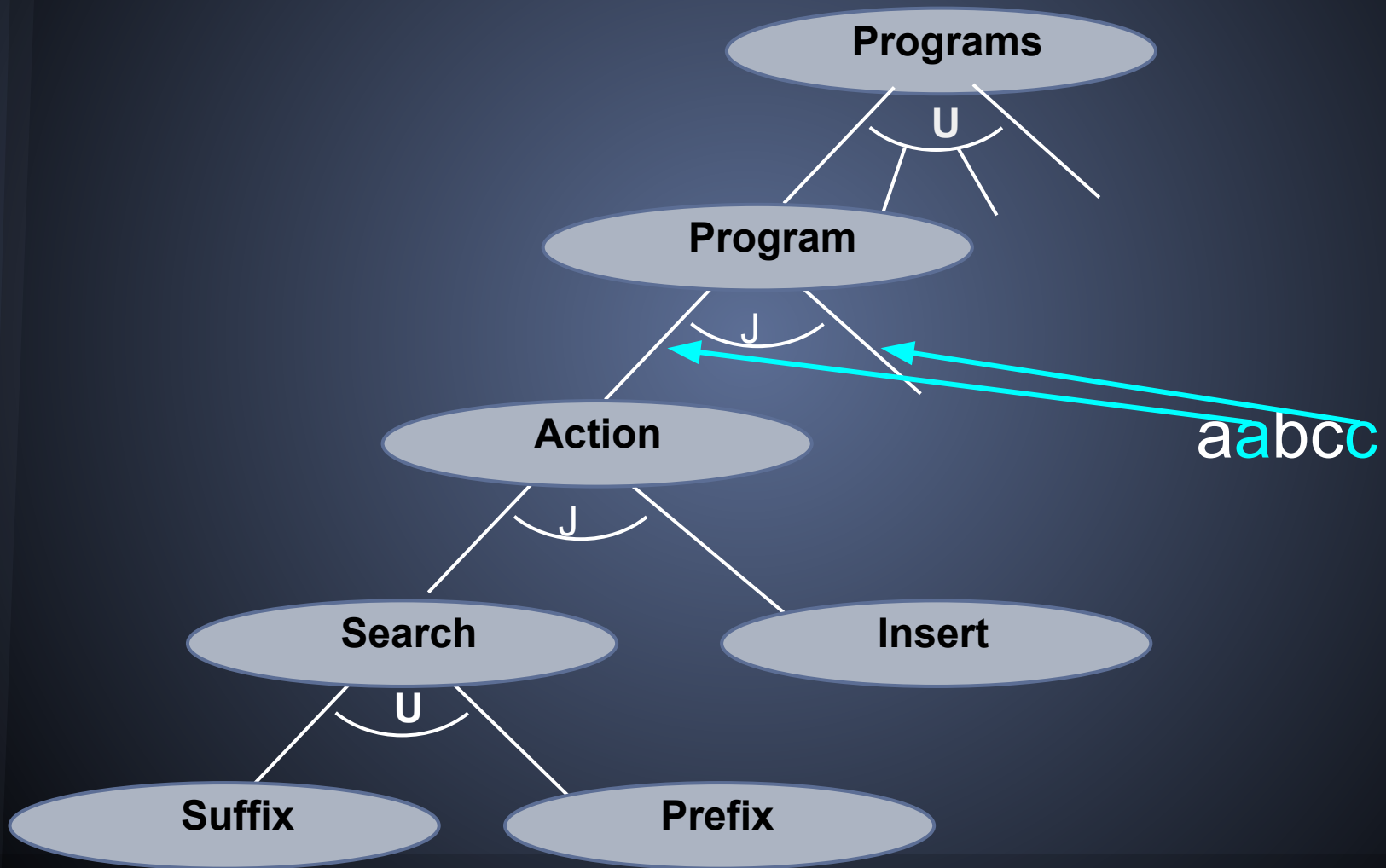


- Transform

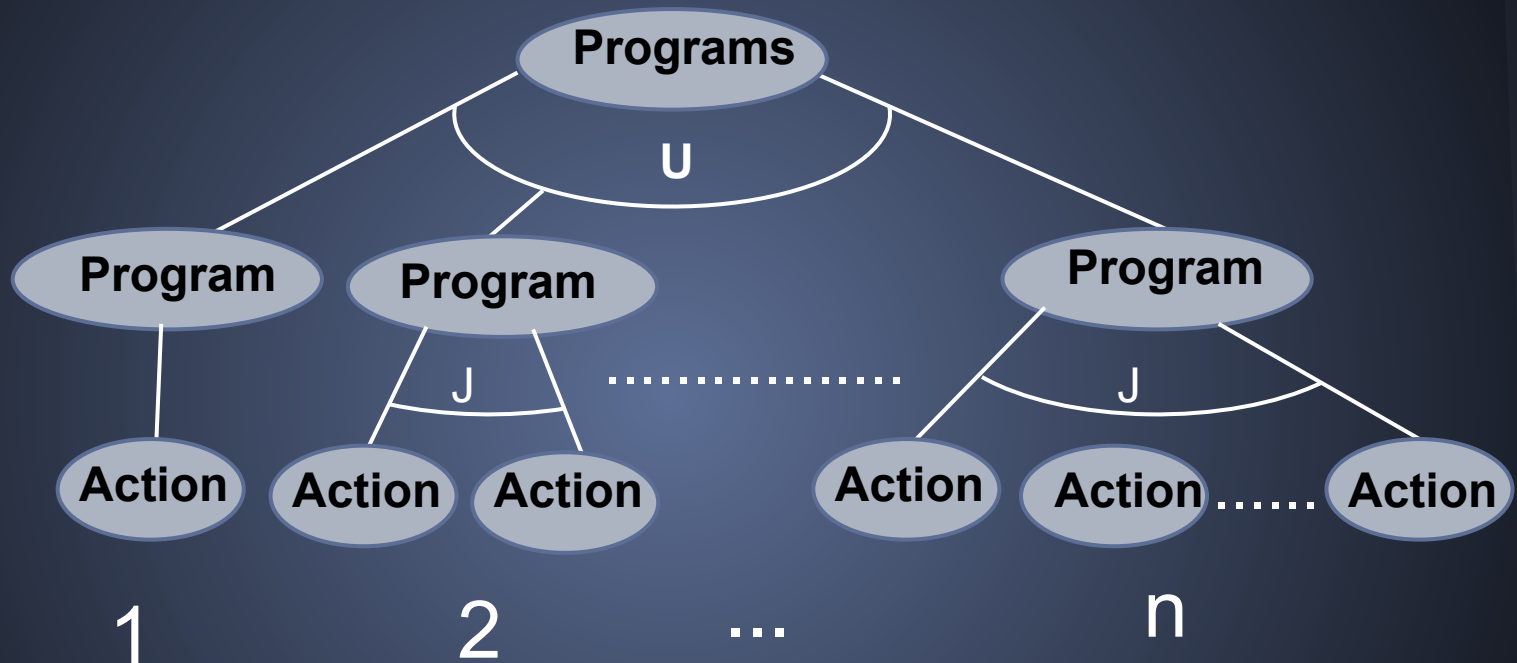
Our Version Space Tree



Our Version Space Tree



Our Version Space Tree



Hi Grandpac!

[Not Grandpac?](#)

\$_html\$_

`<h1>`

Submit

`<title>DeTEXT</title>`

edit

`</head>`

edit

`<body>`

edit

`<h1>`

edit

`This is a heading`

edit

`</h1>`

edit

`<p>`

edit

`This is a paragraph`

edit

`and it has a `

edit

`</p>`

edit

`</body>`

edit

Share

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



aabcc

aabcc

aabcc

aabcc

"_" is Insertion