Data-Driven Evidence-Based Syntactic Sugar Design

David OBrien, Robert Dyer, Tien Nguyen, and Hridesh Rajan

davidob@iastate.edu, rdyer@unl.edu, tien.n.nguyen@utdallas.edu, hridesh@iastate.edu

46th International Conference on Software Engineering
PEP 622 – Structural Pattern Matching

Author: Brandt Bucher <brandt at python.org>, Daniel F. Moisset <dfmoisset at gmail.com>, Tobias Kohn <kohn at tobiaskohn.ch>, Ivan Levkivskyi <ilevkivskyi at gmail.com>, Guido van Rossum <guido at python.org>, Talin <viridia at gmail.com>

BDFL Delegate: Discussions-To: Python-Dev Int
Status: Superseded
Type: Standards.Track
Created: 23-Jun-2020
Python-Version: 3.10
Post-History: 23-Jun-2020, 08-Jul-2020
Superseded-By: 834

Based off anecdotal experiences

Navigated by designer’s intuition

Low impact / irrelevant changes
Contributions

1. Technique for recommending data-driven evidence-based programming language evolution.

2. Evaluation of our approach against a baseline.

3. A catalog of 7 data-driven Java syntactic sugars.

4. Survey evaluating user appreciation of catalog.
**Data-Driven Evidence-Based Syntactic Sugars**

*Syntactic Sugar* is a local transformation in programming languages, providing alternative syntax to express or condense code.

```java
result = result ?: 0;

if (result == null) {
    result = 0;
}
```

Elvis Operator ( `?:` )

Desugared Equivalent
Frequent idioms can serve as data-driven evidence for programming language evolution!

Elvis Operator ( ?? )  Desugared Equivalent
Overview of Approach

Source Code Dataset → 166 Million+ Modified CFGs → Frequent Subgraphs → Data-Driven Syntactic Sugars
Overview of Approach

Boa Java Dataset

<table>
<thead>
<tr>
<th>Granularity</th>
<th>Amount</th>
</tr>
</thead>
<tbody>
<tr>
<td>Projects</td>
<td>380,125</td>
</tr>
<tr>
<td>Recent Snapshot Functions</td>
<td>166,827,154</td>
</tr>
</tbody>
</table>
Overview of Approach

Source Code Dataset → 166 Million+ Modified CFGs → Frequent Subgraphs → Data-Driven Syntactic Sugars
Control Flow Graph Overview

Example Java Code

```java
if (result == null) {
  result = 0;
}
```

Control Flow Graph of Example Java Code
Control Flow Graph Overview

Control Flow Graph of Example Java Code

result == null

TRUE

result = 0;

Generalized CFG of Example Java Code

NULL EQUALITY

USE-DEF TRUE

int ASSIGN LIT
Overview of Approach

- Source Code Dataset
- 166 Million+ Modified CFGs
- Frequent Subgraphs
- Data-Driven Syntactic Sugars
Frequent Subgraph Mining

Graph Database

Found Frequent Subgraphs
(> 66%)
Overview of Approach

Source Code Dataset → 166 Million+ Modified CFGs → Frequent Subgraphs → Data-Driven Syntactic Sugars
Data-Driven Evidence-Based Syntactic Sugars

166 Million+ Generalized CFG Database

1,865
Frequent Subgraphs
## Data-Driven Evidence-Based Syntactic Sugars

<table>
<thead>
<tr>
<th>Proposed Syntactic Sugar</th>
<th>Amount of Subgraphs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Multiple Assignment</td>
<td>5,543,853</td>
</tr>
<tr>
<td>Multiple ++</td>
<td>292,449</td>
</tr>
<tr>
<td>Unless</td>
<td>9,574,658</td>
</tr>
<tr>
<td>Any &amp; All</td>
<td>10,267,105</td>
</tr>
<tr>
<td>Null if Null</td>
<td>236,567</td>
</tr>
<tr>
<td>requireType</td>
<td>116,742</td>
</tr>
<tr>
<td>Rethrow</td>
<td>1,994,698</td>
</tr>
</tbody>
</table>
Data-Driven Evidence-Based Syntactic Sugars

Multiple Assignment Proposal

Java Code

```java
id = 0;
name = “Bob”;
age = 50;
```

Proposed Syntactic Sugar

```java
id, name, age = 0, “Bob”, 50;
```
Data-Driven Evidence-Based Syntactic Sugars

Any & All Proposal

Java Code

```java
if (cond1 && cond2 && cond3) {
    body
}
if (cond1 || cond2 || cond3) {
    body
}
```

Proposed Syntactic Sugar

```java
all(cond1, cond2, cond3) {
    body
}
any(cond1, cond2, cond3) {
    body
}
```
Data-Driven Evidence-Based Syntactic Sugars

**Unless Statement Proposal**

Java Code

```java
if (!cond) {
    body
}
```

Proposed Syntactic Sugar

```java
unless (cond) {
    body
}
```
Data-Driven Evidence-Based Syntactic Sugars

JEP 430: String Templates (Preview)

<table>
<thead>
<tr>
<th>Owner</th>
<th>Jim Laskey</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type</td>
<td>Feature</td>
</tr>
<tr>
<td>Scope</td>
<td>SE</td>
</tr>
<tr>
<td>Status</td>
<td>Closed/Delivered</td>
</tr>
<tr>
<td>Release</td>
<td>21</td>
</tr>
<tr>
<td>Component</td>
<td>specification / language</td>
</tr>
<tr>
<td>Discussion</td>
<td>amber dash dev at openjdk dot org</td>
</tr>
<tr>
<td>Effort</td>
<td>M</td>
</tr>
<tr>
<td>Duration</td>
<td>M</td>
</tr>
<tr>
<td>Relates to</td>
<td>JEP 459: String Templates (Second Preview)</td>
</tr>
<tr>
<td>Reviewed by</td>
<td>Alex Buckley, Brian Goetz, Maurizio Cimadamore</td>
</tr>
<tr>
<td>Endorsed by</td>
<td>Brian Goetz</td>
</tr>
<tr>
<td>Created</td>
<td>2021/09/17 13:41</td>
</tr>
<tr>
<td>Updated</td>
<td>2023/10/13 13:52</td>
</tr>
<tr>
<td>Issue</td>
<td>8273943</td>
</tr>
</tbody>
</table>

String name = "Joan";
String info = STR."My name is \{name}"
assert info.equals("My name is Joan"); // true
Survey

We survey a population consisting of:

Participants’ Occupation

Participants’ Java Experience

- Graduate Student
- Practitioner
- Other

- 0-2 Years
- 2-5 Years
- 5-10 Years
- 10-15 Years
- 15-20 Years
- 20+ Years
Survey Results
Survey

Survey Results:

(Multiple Assignment 58.06%): Seems to read ‘easier’ (matches how we would state the condition in natural language) and unambiguously, so I’m a fan of the new feature.

(Unless 41.94%): I like the idea of unless, but the “else” case seems awkward. Is there a better word for unless?
Conclusion

Overview of Approach

Data-Driven Evidence-Based Syntactic Sugars

<table>
<thead>
<tr>
<th>Proposed Syntactic Sugar</th>
<th>Amount of CFGs Present</th>
</tr>
</thead>
<tbody>
<tr>
<td>Multiple Assignment</td>
<td>3,343,853</td>
</tr>
<tr>
<td>Multiple ++</td>
<td>292,449</td>
</tr>
<tr>
<td>Unless</td>
<td>9,574,458</td>
</tr>
<tr>
<td>Any &amp; All</td>
<td>10,267,105</td>
</tr>
<tr>
<td>Null if Null</td>
<td>236,567</td>
</tr>
<tr>
<td>requireType</td>
<td>116,742</td>
</tr>
<tr>
<td>Rethrow</td>
<td>1,994,698</td>
</tr>
</tbody>
</table>

Survey Results:

Survey Results: Approved Replacements for Syntactic Sugar

Syntactic Sugar Replacements

- Approved Replacements
- Survey Results