

# STATISTICAL PROGRAM ANALYSIS AND SYNTHESIS

Martin Vechev

Department of Computer Science  
ETH Zurich

# Motivation

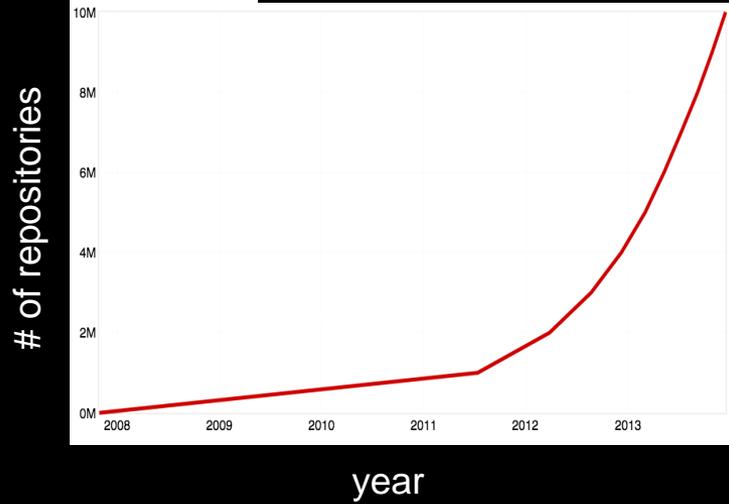
Unprecedented access to **massive codebases**  
(termed “Big Code” by a recent DARPA initiative)



# Motivation



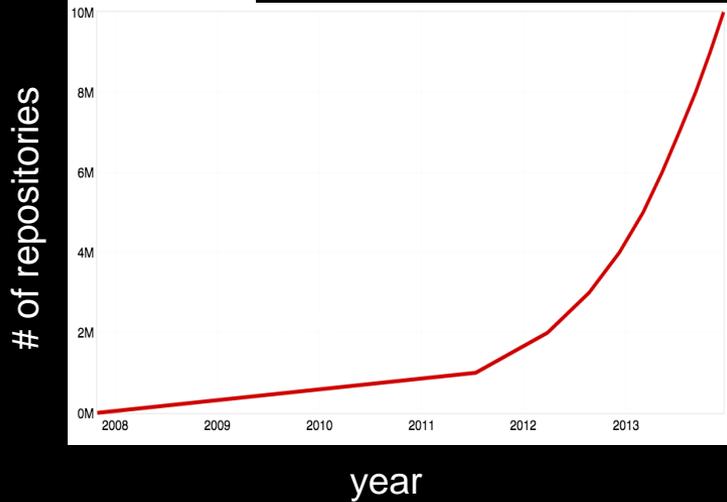
~16M repos  
~ 7M users



# Motivation



~16M repos  
~ 7M users



**Atlassian Bitbucket Passes 1 Million Users, Another Validation Of The Fast-Growing Developer Market**

Posted Jun 4, 2013 by Alex Williams (@alexwilliams)

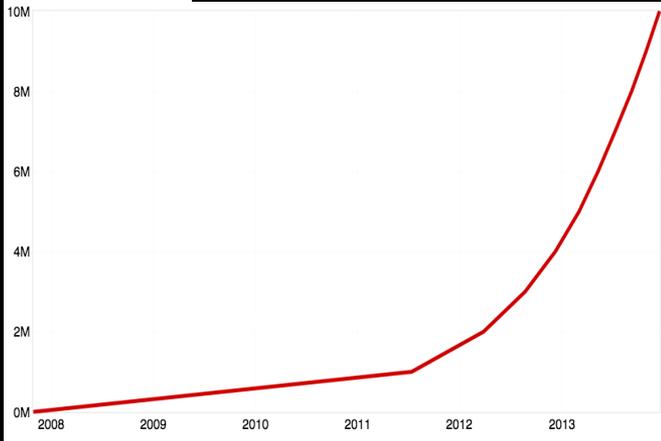
3 Share 128 Share 45 Tweet 126

# Motivation

# of repositories



~16M repos  
~ 7M users



year



**Atlassian Bitbucket Passes 1 Million Users, Another Validation Of The Fast-Growing Developer Market**

Posted Jun 4, 2013 by Alex Williams (@alexwilliams)  
3 Share 128 Share 45 Tweet 126



## Number of Android applications

Current number of Android apps in the market:

**1,375,385**

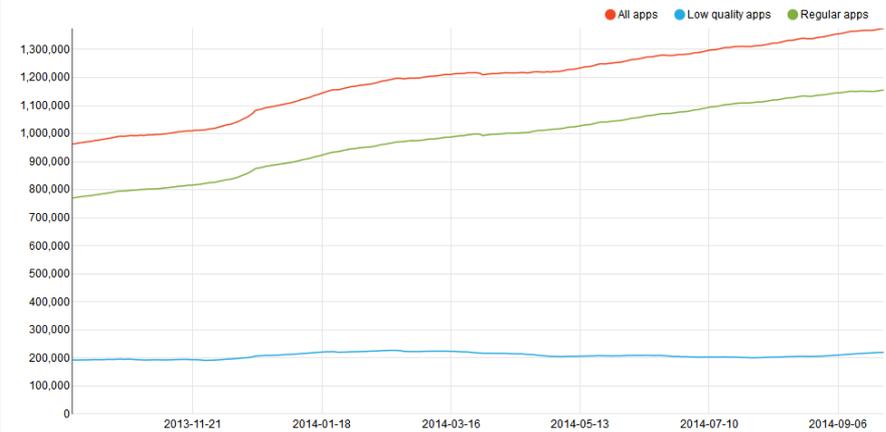
Percentage of low quality apps: 16%

## Development over time

This graph shows the number of available applications on Google Play. Last update: September 29, 2014.

The AppBrain low quality app detection filter detects automatically which apps are unlikely to be useful ("low quality apps"). Google seems to remove apps from the market roughly once a quarter, in which case the total number of available Android apps goes down. The removed apps are almost always classified by our system as low quality apps.

## Android apps on Google Play



# Vision

New techniques which **solve programming challenges** by leveraging availability of **massive codebases**

# General Approach

Phrase the problem in a way which allows creative combination of advanced program analysis and powerful machine learning techniques

# Statistical Program Analysis and Synthesis

Statistical Prediction of Program Properties, [POPL'15](#)  
(V. Raychev, M. Vechev, A. Krause)



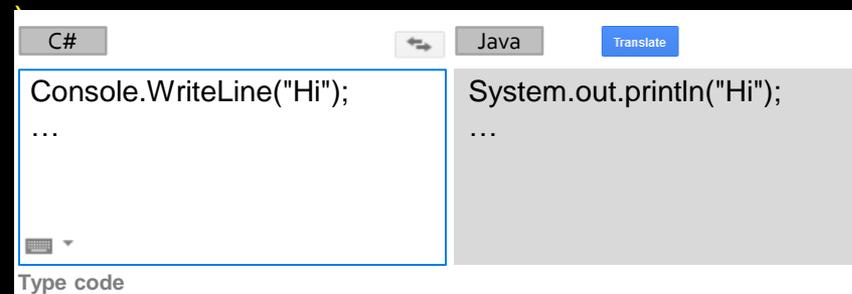
The screenshot shows the JS RICE web interface. The left pane, titled 'ENTER JAVASCRIPT', contains the following code:

```
// Put your JavaScript here that you want to rename, deobfuscate, or infer types
// Example:
function chunkData(e, t) {
  var n = t.length;
  var s = 0;
  for (var i = 0; i < n; i++) {
    m.push(e.substring(i, i + t));
  }
  return m;
}
```

The right pane, titled 'RESULT', shows the transformed code:

```
function chunkData(str, step) {
  var colnames = [];
  var len = str.length;
  var group = [number] +
  for (i = 0; i < len; i += step) {
    colnames.push(str.substring(i, i + step));
  }
  return colnames;
}
```

Statistical PL translation, [Onward'14](#)  
(S. Karaivanov, V. Raychev, M. Vechev)



The screenshot shows a web interface for translating code. The left pane is labeled 'C#' and contains the code:

```
Console.WriteLine("Hi");
...
```

The right pane is labeled 'Java' and contains the translated code:

```
System.out.println("Hi");
...
```

A 'Translate' button is visible between the panes.

Statistical Code Synthesis, [PLDI'14](#)  
(V. Raychev, M. Vechev, E. Yahav)

```
camera.setDisplayOrientation(90);
```

```
camera.unlock();
```

```
SurfaceHolder holder = getHolder();
```

# Statistical Program Analysis and Synthesis

Statistical Code Synthesis, [PLDI'14](#)

(V. Raychev, M. Vechev, E. Yahav)

```
camera.setDisplayOrientation(90);
```

```
camera.unlock();
```

```
SurfaceHolder holder = getHolder();
```

# Slang: Statistical Code Synthesis

```
void exampleMediaRecorder() throws IOException {  
    Camera camera = Camera.open();  
    camera.setDisplayOrientation(90);
```

```
    SurfaceHolder holder = getHolder();  
    holder.addCallback(this);  
    holder.setType(SurfaceHolder.SURFACE_TYPE_PUSH_BUFFERS);  
    MediaRecorder rec = new MediaRecorder();
```

```
    rec.setAudioSource(MediaRecorder.AudioSource.MIC);  
    rec.setVideoSource(MediaRecorder.VideoSource.DEFAULT);  
    rec.setOutputFormat(MediaRecorder.OutputFormat.MPEG_4);
```

```
}
```

# Slang: Statistical Code Synthesis

```
void exampleMediaRecorder() throws IOException {
```

```
    Camera camera = Camera.open();  
    camera.setDisplayOrientation(90);
```

```
    camera.unlock();
```

```
    SurfaceHolder holder = getHolder();  
    holder.addCallback(this);  
    holder.setType(SurfaceHolder.SURFACE_TYPE_PUSH_BUFFERS);  
    MediaRecorder rec = new MediaRecorder();
```

```
    rec.setCamera(camera);
```

```
    rec.setAudioSource(MediaRecorder.AudioSource.MIC);  
    rec.setVideoSource(MediaRecorder.VideoSource.DEFAULT);  
    rec.setOutputFormat(MediaRecorder.OutputFormat.MPEG_4);
```

```
    rec.setAudioEncoder(1);  
    rec.setVideoEncoder(3);
```

```
}
```

# Slang: Statistical Code Synthesis

```
void exampleMediaRecorder() throws IOException {  
    Camera camera = Camera.open();  
    camera.setDisplayOrientation(90);
```

```
    camera.unlock();
```

```
    SurfaceHolder holder = getHolder();  
    holder.addCallback(this);  
    holder.setType(SurfaceHolder.SURFACE_TYPE_PUSH_BUFFERS);  
    MediaRecorder rec = new MediaRecorder();
```

```
    rec.setCamera(camera);
```

```
    rec.setAudioSource(MediaRecorder.AudioSource.MIC);  
    rec.setVideoSource(MediaRecorder.VideoSource.DEFAULT);  
    rec.setOutputFormat(MediaRecorder.OutputFormat.MPEG_4);
```

```
    rec.setAudioEncoder(1);  
    rec.setVideoEncoder(3);
```

```
}
```

sequence not in  
training data

# Slang: Statistical Code Synthesis

```
void exampleMediaRecorder() throws IOException {  
    Camera camera = Camera.open();  
    camera.setDisplayOrientation(90);
```

```
    camera.unlock();
```

```
    SurfaceHolder holder = getHolder();  
    holder.addCallback(this);  
    holder.setType(SurfaceHolder.SURFACE_TYPE_PUSH_BUFFERS);  
    MediaRecorder rec = new MediaRecorder();
```

```
    rec.setCamera(camera);
```

```
    rec.setAudioSource(MediaRecorder.AudioSource.MIC);  
    rec.setVideoSource(MediaRecorder.VideoSource.DEFAULT);  
    rec.setOutputFormat(MediaRecorder.OutputFormat.MPEG_4);
```

```
    rec.setAudioEncoder(1);  
    rec.setVideoEncoder(3);
```

```
}
```

sequence not in  
training data

multiple objects

# Slang: Statistical Code Synthesis

```
void exampleMediaRecorder() throws IOException {  
    Camera camera = Camera.open();  
    camera.setDisplayOrientation(90);
```

```
    camera.unlock();
```

```
    SurfaceHolder holder = getHolder();  
    holder.addCallback(this);  
    holder.setType(SurfaceHolder.SURFACE_TYPE_PUSH_BUFFERS);  
    MediaRecorder rec = new MediaRecorder();
```

```
    rec.setCamera(camera);
```

```
    rec.setAudioSource(MediaRecorder.AudioSource.MIC);  
    rec.setVideoSource(MediaRecorder.VideoSource.DEFAULT);  
    rec.setOutputFormat(MediaRecorder.OutputFormat.MPEG_4);
```

```
    rec.setAudioEncoder(1);  
    rec.setVideoEncoder(3);
```

```
}
```

sequence not in training data

multiple objects

scalar parameters

# Slang: Statistical Code Synthesis

```
void exampleMediaRecorder() throws IOException {  
    Camera camera = Camera.open();  
    camera.setDisplayOrientation(90);
```

```
    camera.unlock();
```

```
    SurfaceHolder holder = getHolder();  
    holder.addCallback(this);  
    holder.setType(SurfaceHolder.SURFACE_TYPE_PUSH_BUFFERS);  
    MediaRecorder rec = new MediaRecorder();
```

```
    rec.setCamera(camera);
```

```
    rec.setAudioSource(MediaRecorder.AudioSource.MIC);  
    rec.setVideoSource(MediaRecorder.VideoSource.DEFAULT);  
    rec.setOutputFormat(MediaRecorder.OutputFormat.MPEG_4);
```

```
    rec.setAudioEncoder(1);  
    rec.setVideoEncoder(3);
```

```
}
```

sequence not in training data

prediction is instantaneous

multiple objects

scalar parameters

# Statistical Program Analysis and Synthesis

Statistical Code Synthesis, PLDI'14

(V. Raychev, M. Vechev, E. Yahav)

```
camera.setDisplayOrientation(90);
```

```
camera.unlock();
```

```
SurfaceHolder holder = getHolder();
```

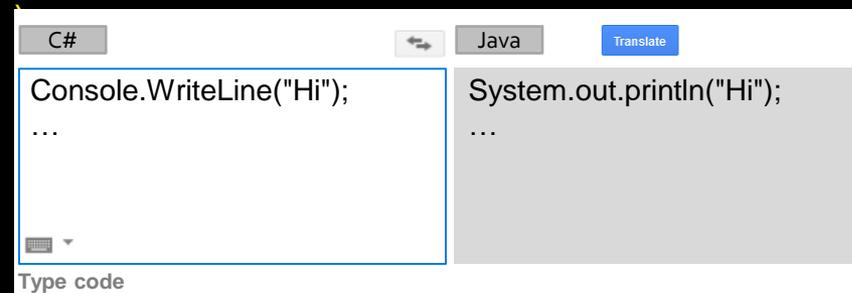
score programs with **generative models** such as n-gram and recurrent neural networks (RNN)

+

sequence extraction with **alias and typestate static analysis**

# Statistical Program Analysis and Synthesis

Statistical PL translation, [Onward'14](#)  
(S. Karaivanov, V. Raychev, M. Vechev)



```
C#                                     Java                                     Translate
Console.WriteLine("Hi");                System.out.println("Hi");
...                                     ...
Type code
```

Statistical Code Synthesis, [PLDI'14](#)  
(V. Raychev, M. Vechev, E. Yahav)

```
camera.setDisplayOrientation(90);
camera.unlock();
SurfaceHolder holder = getHolder();
```

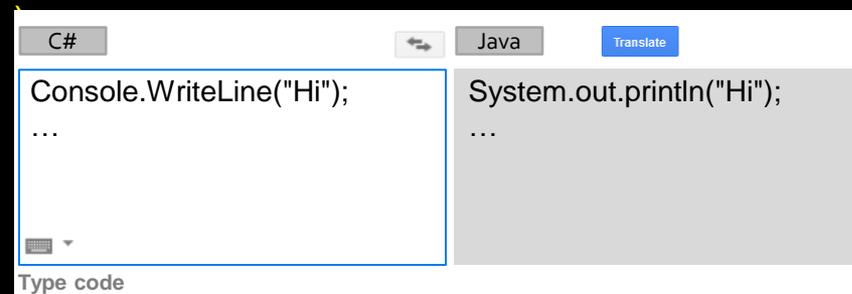
score programs with **generative models** such as  
n-gram and recurrent neural networks (RNN)

+

sequence extraction with **alias and typestate**  
**static analysis**

# Statistical Program Analysis and Synthesis

Statistical PL translation, Onward'14  
(S. Karaivanov, V. Raychev, M. Vechev)



```
C#                                     Java                                     Translate
Console.WriteLine("Hi");               System.out.println("Hi");
...                                     ...
```

Type code

phrase-based statistical machine translation

+

parse prefix grammars during translation

Statistical Code Synthesis, PLDI'14  
(V. Raychev, M. Vechev, E. Yahav)

```
camera.setDisplayOrientation(90);
```

```
camera.unlock();
```

```
SurfaceHolder holder = getHolder();
```

score programs with generative models such as n-gram and recurrent neural networks (RNN)

+

sequence extraction with alias and typestate static analysis

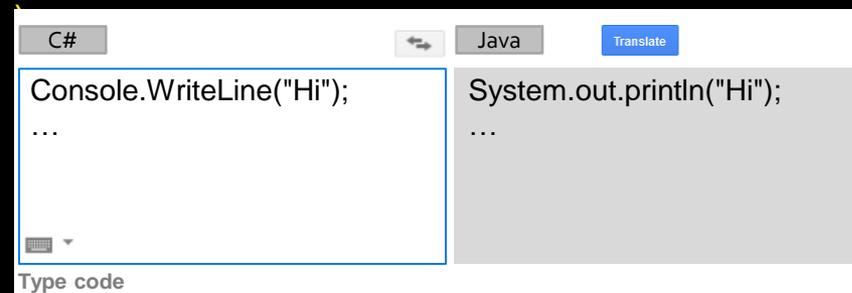
# Statistical Program Analysis and Synthesis

Statistical Prediction of Program Properties, [POPL'15](#)  
(V. Raychev, M. Vechev, A. Krause)



The screenshot shows the JS RICE web interface. On the left, under 'ENTER JAVASCRIPT', there is a code block with a function `chunkData` that takes a string and a number and returns an array of substrings. On the right, under 'RESULT', the transformed code is shown, where the function is renamed to `chunkData` and the variable `colNames` is used to store the results.

Statistical PL translation, [Onward'14](#)  
(S. Karaivanov, V. Raychev, M. Vechev)



The screenshot shows the Onward'14 web interface. On the left, under 'C#', there is a code block with `Console.WriteLine("Hi");`. On the right, under 'Java', the translated code is shown as `System.out.println("Hi");`. A 'Translate' button is visible between the two panels.

phrase-based statistical machine translation

+

parse prefix grammars during translation

Statistical Code Synthesis, [PLDI'14](#)  
(V. Raychev, M. Vechev, E. Yahav)

camera.setDisplayOrientation(90);

camera.unlock();

SurfaceHolder holder = getHolder();

score programs with generative models such as n-gram and recurrent neural networks (RNN)

+

sequence extraction with alias and typestate static analysis

# Statistical Program Analysis and Synthesis

Statistical Prediction of Program Properties, [POPL'15](#)  
(V. Raychev, M. Vechev, A. Krause)

DEMO



The screenshot shows the JS-RICE web interface. The left pane, titled 'ENTER JAVASCRIPT', contains the following code:

```
1 // Put your JavaScript here that you want to rename, deobfuscate, or infer types
2 foo!
3
4 // Example:
5 function chunkData(e, t) {
6   var n = t.length;
7   var s = 0;
8   for (i = 0; i < n; i++) {
9     if (i % t == 0) {
10      m.push(e.substring(i, i + t));
11    }
12  }
13  return m;
14 }
```

The right pane, titled 'RESULT', shows the transformed code:

```
1 /**
2  * @param {string} str
3  * @param {number} step
4  * @return {Array}
5  */
6 function chunkData(str, step) {
7   var chunk = [];
8   var colnames = [];
9   var len = str.length;
10  var group = [number] + [];
11  for (i = 0; i < len; i += step) {
12    chunk.push(str.substring(i, i + step));
13  }
14  colnames.push(str.substring(0, len));
15  return chunk;
16 }
```

# JSNice



800+ Tweets (sample below):



**Kamil Tomšík** @cztomsik · Jun 6

tell me [how this](#) works!

de-minify #jquery #javascript incl. args, vars & #jsdoc  
impressive! [jsnice.org](http://jsnice.org)

# JSNice



800+ Tweets (sample below):



**Alex Vanston** @mrvdot · Jun 7

I've been looking for this for years: JS NICE [buff.ly/1pQ5qfr](https://buff.ly/1pQ5qfr) #javascript #unminify #deobfuscate #makeItReadable

Expand

↩ Reply ↻ Retweet ★ Favorite ⋮ More



**Kamil Tomšik** @cztomsik · Jun 6

tell me [how this](#) works!  
de-minify #jquery #javascript incl. args, vars & #jsdoc  
impressive! [jsnice.org](http://jsnice.org)

# JSNice



800+ Tweets (sample below):



**Alvaro Sanchez** @alvasavi · Jun 10

This is gold.

Statistical renaming, Type inference and Deobfuscation.

[jsnice.org](http://jsnice.org)

Expand

Reply Retweet Favorite More



**Alex Vanston** @mrvdot · Jun 7

I've been looking for this for years: JS NICE [buff.ly/1pQ5qfr](http://buff.ly/1pQ5qfr) #javascript

#unminify #deobfuscate #makeItReadable

Expand

Reply Retweet Favorite More



**Kamil Tomšik** @cztomsik · Jun 6

tell me how this works!

de-minify #jquery #javascript incl. args, vars & #jsdoc

impressive! [jsnice.org](http://jsnice.org)

# JSNice



800+ Tweets (sample below):



**Brevity** @seekbrevity · Jul 28

JSNice is an a [amazing](#) tool for de-minifying [#javascript](#) files. [JSNice.org](#), Its great for [#learning](#) and reverse engineering.

Expand

↩ Reply ↻ Retweet ★ Favorite ⋮ More



**Alvaro Sanchez** @alvasavi · Jun 10

[This is gold.](#)

Statistical renaming, Type inference and Deobfuscation.

[jsnice.org](#)

Expand

↩ Reply ↻ Retweet ★ Favorite ⋮ More



**Alex Vanston** @mrvdot · Jun 7

I've been looking for this for years: JS NICE [buff.ly/1pQ5qfr](#) [#javascript](#) [#unminify](#) [#deobfuscate](#) [#makeItReadable](#)

Expand

↩ Reply ↻ Retweet ★ Favorite ⋮ More



**Kamil Tomšik** @cztomsik · Jun 6

tell me [how this](#) works!

de-minify [#jquery](#) [#javascript](#) incl. args, vars & [#jsdoc](#) impressive! [jsnice.org](#)

# JSNice



800+ Tweets (sample below):



**Ingvar Stepanyan** @RRReverser · Aug 6

JSNice.org became my **must-have** tool for code deobfuscation.

Expand

↩ Reply ↻ Retweet ★ Favorite ⋮ More



**Brevity** @seekbrevity · Jul 28

JSNice is an a **amazing** tool for de-minifying #javascript files. JSNice.org, Its great for #learning and reverse engineering.

Expand

↩ Reply ↻ Retweet ★ Favorite ⋮ More



**Alvaro Sanchez** @alvasavi · Jun 10

**This is gold.**

Statistical renaming, Type inference and Deobfuscation.

[jsnice.org](http://jsnice.org)

Expand

↩ Reply ↻ Retweet ★ Favorite ⋮ More



**Alex Vanston** @mrvdot · Jun 7

I've been looking for this for years: JS NICE [buff.ly/1pQ5qfr](https://buff.ly/1pQ5qfr) #javascript #unminify #deobfuscate #makeItReadable

Expand

↩ Reply ↻ Retweet ★ Favorite ⋮ More



**Kamil Tomšik** @cztomsik · Jun 6

tell me **how this** works!

de-minify #jquery #javascript incl. args, vars & #jsdoc impressive! [jsnice.org](http://jsnice.org)

# JSNice

- 30,000 users in 1<sup>st</sup> week
- used in > 170 countries



800+ Tweets (sample below):



**Ingvar Stepanyan** @RRReverser · Aug 6

JSNice.org beca

Expand

## 50 fantastic freebies for web designers, July 2014

By **Juan Pablo Sarmiento** | Resources | Jul 16, 2014



**Brevity** @seekbr

JSNice is an a  
great for #learnir

Expand



**Alvaro Sanchez**

This is gold.

Statistical renaming, Type inference and Deobfuscation.

[jsnice.org](http://jsnice.org)

Expand

← Reply ↻ Retweet ★ Favorite ⋮ More



**Alex Vanston** @mrvdot · Jun 7

I've been looking for this for years: JS NICE [buff.ly/1pQ5qfr](http://buff.ly/1pQ5qfr) #javascript  
#unminify #deobfuscate #makeItReadable

Expand

← Reply ↻ Retweet ★ Favorite ⋮ More



**Kamil Tomšik** @cztomsik · Jun 6

tell me [how this](#) works!

de-minify #jquery #javascript incl. args, vars & #jsdoc  
impressive! [jsnice.org](http://jsnice.org)

# JSNice

- 30,000 users in 1<sup>st</sup> week
- used in > 170 countries



800+ Tweets (sample below):



**Ingvar Stepanyan** @RRReverser · Aug 6

JSNice.org beca

Expand

## 50 fantastic freebies for web designers, July 2014

By **Juan Pablo Sarmiento** | Resources | Jul 16, 2014



**Brevity** @seekbr

JSNice is an a  
great for #learnin

Expand



**Alvaro Sanchez**

This is gold.

Statistical renaming, Type inference and Deobfuscation.

[jsnice.org](http://jsnice.org)

Expand

Reply Retweet Favorite More



**Alex Vanston** @mrvdot · Jun 7

I've been looking for this for years: JS NICE [buff.ly/1pQ5qfr](http://buff.ly/1pQ5qfr) #javascript #unminify #deobfuscate #makeItReadable

Expand

Reply Retweet Favorite More



**Kamil Tomšik** @cztomsik · Jun 6

tell me **how this** works!

de-minify #jquery #javascript incl. args, vars & #jsdoc impressive! [jsnice.org](http://jsnice.org)

The screenshot shows a webpage from CodeGeekz titled "20 Essential Tools for Coders" by Gavin, dated July 20, 2014. The article features a section on "JavaScript Diagrams" with a diagram showing the flow from code to a visual representation. The text discusses the benefits of development tools and lists "1. JS Nice" as one of the tools. The page includes a navigation menu, a comments section, and social media sharing options.

# JSNice

- 30,000 users in 1<sup>st</sup> week
- used in > 170 countries



800+ Tweets (sample below):



**Ingvar Stepanyan** @RRReverser · Aug 6

JSNice.org beca

Expand



**Brevity** @seekbr

JSNice is an a  
great for #learnir

Expand



**Alvaro Sanchez**

This is gold.  
Statistical renaming, Type infer  
[jsnice.org](http://jsnice.org)

Expand



**Alex Vanston** @mrvdot · Jun 7

I've been looking for this for years: JS NICE [buff.ly/1pQ5qfr](http://buff.ly/1pQ5qfr) #javascript  
#unminify #deobfuscate #makeItReadable

Expand

Reply Retweet Favorite More



**Kamil Tomšik** @cztomsik · Jun 6

tell me **how this** works!  
de-minify #jquery #javascript incl. args, vars & #jsdoc  
impressive! [jsnice.org](http://jsnice.org)

## 50 fantastic freebies for

## July 2014

AUGUST 28, 2014

### JavaScript迷宮救星 — JS nICE



我們網頁前端工程師常常要在網站上用瀏覽器的debug來替網站上的JavaScript檔案除錯，通常那些檔案都已經經過最小化、最佳化，還有變數混淆了，以增進檔案讀取的效率，並防止他人輕易盜用程式碼。不過一旦要除錯，就苦了前端工程師了。

就算現代的瀏覽器如Chrome, Safari與Firefox都支援美化程式碼的功能，讓最小化的程式還原成有排版的狀況，但是遇到變數混淆過的程式，面對一堆變數名稱是a, b, c, d的程式，還是會除錯除得很辛苦。

不過網路上有些神奇的工具可以幫助我們不僅是把糊成一團的JS檔排的美美的，甚至它會替我們逆推算已經混淆的變數名稱！

那就是這篇要介紹的JS nICE網站。

這個網站的介面超級乾淨，就像下面一樣，左邊是複製貼上的JS原始碼，只要按下NICIFY JAVASCRIPT按鈕，經過排版與替換變數的成果就會出現在右邊了！



COMMENTS

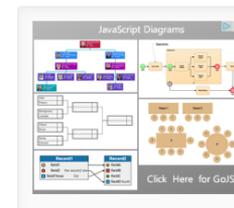
MOST RECENT

Frick Komz on  
20 Online Code Editors and Tools for Developers

HOME WORDPRESS WEB DESIGN TYPOGRAPHY RESOURCES ARCHIVES CONTACT ADVERTISE

### 20 Essential Tools for Coders

BY GAVIN IN DEVELOPMENT RESOURCES - 20 JUL, 2014



Over the last decade, third party tools and free development tools have become an increasingly popular solution for building web applications through a fast and cost-efficient development process. These tools are enriched with various features that turned out to be handy during the development process of an application. Probably this is the reason coders and developers keep looking for free and useful tools which make their work easier and quick.

If you are a coder and looking for some useful development tools then you are at

right place, I have gathered 20 Essential Tools for Coders that would help you out with your development tasks. Following development tools let you write your ideas into code, debug, beautify, store your snippet and help you to simplify your development tasks quickly. We hope you will find the list handy and according to your requirements. Enjoy!

#### 1. JS Nice



# JSNice

- 30,000 users in 1<sup>st</sup> week
- used in > 170 countries



800+ Tweets (sample below):



**Ingvar Stepanyan** @RRReverser · Aug 6

JSNice.org beca

Expand



**Brevity** @seekbr

JSNice is an a  
great for #learnir

Expand



**Alvaro Sanchez**

This is gold.

Statistical renaming, Type infer

[jsnice.org](http://jsnice.org)

Expand



**Alex Vanston** @mrvdot · Jun 7

I've been looking for this for years: JS NICE buf  
#unminify #deobfuscate #makeltReadable

Expand



**Kamil Tomšik** @cztomsik · Jun 6

tell me how this works!

de-minify #jquery #javascript incl. args, vars  
impressive! [jsnice.org](http://jsnice.org)

## 50 fantastic freebies for

## July 2014

AUGUST 28, 2014

### JavaScript迷宮救星 — JS nICE



我們網頁前端工程師常常要在網站上用瀏覽器的debug來替網站上的JavaScript檔案除錯，通常那些檔案都已經經過最小化、最佳化，還有變數混淆了，以增進檔案讀取的效率，並防止他人輕易盜用程式碼。不過一旦要除錯，就苦了前端工程師了。

就算現代的瀏覽器如Chrome, Safari, Firefox, Opera, Edge, 透過的程式，面對一堆變數名稱是a, b, c, d, e, f, g, h, i, j, k, l, m, n, o, p, q, r, s, t, u, v, w, x, y, z, 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 那這就是這篇要介紹的JS nICE網站。

這個網站的介面超級乾淨，就像下面圖

標數的成果就會出現在右邊了！



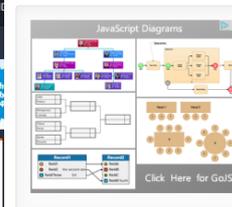
### Comment déminifier un code Javascript



COMMENTS

### 20 Essential Tools for Coders

BY GAVIN IN DEVELOPMENT RESOURCES - 20 JUL, 2014



Over the last decade, third party tools and free development tools have become an increasingly popular solution for building web applications through a fast and cost-efficient development process. These tools are enriched with various features that turned out to be handy during the development process of an application. Probably this is the reason coders and developers keep looking for free and useful tools which make their work easier and quick.

If you are a coder and looking for some useful development tools then you are at

right place, I have gathered 20 Essential Tools for Coders that would help you out with your development tasks. Following development tools let you write your ideas into code, debug, beautify, store your snippet and help you to simplify your development tasks quickly. We hope you will find the list handy and according to your requirements. Enjoy!

#### 1. JS Nice



# JSNice

```
function chunkData(e, t)
  var n = [];
  var r = e.length;
  var i = 0;
  for (; i < r; i += t)
    if (i + t < r)
      n.push(e.substring(i, i + t));
    else
      n.push(e.substring(i, r));
  return n;
```



```
function chunkData(str, step)
  var colNames = [];
  var len = str.length;
  var i = 0;
  for (; i < len; i += step)
    if (i + step < len)
      colNames.push(str.substring(i, i + step));
    else
      colNames.push(str.substring(i, len));
  return colNames;
```



**Kamil Tomšik** @cztomsik · Jun 6

tell me **how this** works!

de-minify #jquery #javascript incl. args, vars & #jsdoc

impressive! [jsnice.org](http://jsnice.org)

# JSNice

$$y = \underset{y \in \Omega}{\operatorname{argmax}} \Pr(y \mid x)$$

# JSNice

```
function chunkData(str, step)
  var colNames = [];
  var len = str.length;
  var i = 0;
  for (; i < len; i += step)
    if (i + step < len)
      colNames.push(str.substring(i, i+step));
    else
      colNames.push(str.substring(i, len));
  return colNames;
```

$$= \operatorname{argmax}_{y \in \Omega} \Pr(y | \quad)$$

```
function chunkData(e, t)
  var n = [];
  var r = e.length;
  var i = 0;
  for (; i < r; i += t)
    if (i + t < r)
      n.push(e.substring(i, i + t));
    else
      n.push(e.substring(i, r));
  return n;
```

# JSNice

```
function chunkData(str, step)
  var colNames = [];
  var len = str.length;
  var i = 0;
  for (; i < len; i += step)
    if (i + step < len)
      colNames.push(str.substring(i, i+step));
    else
      colNames.push(str.substring(i, len));
  return colNames;
```

$$= \underset{y \in \Omega}{\operatorname{argmax}} \Pr(y | \text{ )}$$

```
function chunkData(e, t)
  var n = [];
  var r = e.length;
  var i = 0;
  for (; i < r; i += t)
    if (i + t < r)
      n.push(e.substring(i, i + t));
    else
      n.push(e.substring(i, r));
  return n;
```

learned from  
"Big Code"



# JSNice

```
function chunkData(str, step)
  var colNames = [];
  var len = str.length;
  var i = 0;
  for (; i < len; i += step)
    if (i + step < len)
      colNames.push(str.substring(i, i+step));
    else
      colNames.push(str.substring(i, len));
  return colNames;
```

$$= \underset{y \in \Omega}{\operatorname{argmax}} \Pr(y | \text{...})$$

directly captured by a  
**discriminative** model

```
function chunkData(e, t)
  var n = [];
  var r = e.length;
  var i = 0;
  for (; i < r; i += t)
    if (i + t < r)
      n.push(e.substring(i, i + t));
    else
      n.push(e.substring(i, r));
  return n;
```

learned from  
“Big Code”



# JSNice

captures **dependence**  
(**structure**) between properties

directly captured by a  
**discriminative** model

```
function chunkData(str, step)
  var colNames = [];
  var len = str.length;
  var i = 0;
  for (; i < len; i += step)
    if (i + step < len)
      colNames.push(str.substring(i, i+step));
    else
      colNames.push(str.substring(i, len));
  return colNames;
```

$$= \operatorname{argmax}_{y \in \Omega} \Pr(y | \dots)$$

```
function chunkData(e, t)
  var n = [];
  var r = e.length;
  var i = 0;
  for (; i < r; i += t)
    if (i + t < r)
      n.push(e.substring(i, i + t));
    else
      n.push(e.substring(i, r));
  return n;
```

learned from  
"Big Code"



# JSNice

captures **dependence (structure)** between properties

directly captured by a **discriminative** model

```
function chunkData(str, step)
  var colNames = [];
  var len = str.length;
  var i = 0;
  for (; i < len; i += step)
    if (i + step < len)
      colNames.push(str.substring(i, i+step));
    else
      colNames.push(str.substring(i, len));
  return colNames;
```

$$= \operatorname{argmax}_{y \in \Omega} \Pr(y | \dots)$$

```
function chunkData(e, t)
  var n = [];
  var r = e.length;
  var i = 0;
  for (; i < r; i += t)
    if (i + t < r)
      n.push(e.substring(i, i + t));
    else
      n.push(e.substring(i, r));
  return n;
```

learned from  
"Big Code"

constraints  
on properties



# JSNice

captures **dependence (structure)** between properties

directly captured by a **discriminative** model

```
function chunkData(str, step)
  var colNames = [];
  var len = str.length;
  var i = 0;
  for (; i < len; i += step)
    if (i + step < len)
      colNames.push(str.substring(i, i+step));
    else
      colNames.push(str.substring(i, len));
  return colNames;
```

$$= \operatorname{argmax}_{y \in \Omega} \Pr(y | \dots)$$

```
function chunkData(e, t)
  var n = [];
  var r = e.length;
  var i = 0;
  for (; i < r; i += t)
    if (i + t < r)
      n.push(e.substring(i, i + t));
    else
      n.push(e.substring(i, r));
  return n;
```

candidate properties

constraints on properties

learned from "Big Code"



# JSNice

make a **joint** prediction for most likely properties

captures **dependence (structure)** between properties

directly captured by a **discriminative** model

```
function chunkData(str, step)
var colNames = [];
var len = str.length;
var i = 0;
for (; i < len; i += step)
  if (i + step < len)
    colNames.push(str.substring(i, i+step));
  else
    colNames.push(str.substring(i, len));
return colNames;
```

$$= \operatorname{argmax}_{y \in \Omega} \Pr(y | \dots)$$

```
function chunkData(e, t)
var n = [];
var r = e.length;
var i = 0;
for (; i < r; i += t)
  if (i + t < r)
    n.push(e.substring(i, i + t));
  else
    n.push(e.substring(i, r));
return n;
```

candidate properties

constraints on properties

learned from "Big Code"



# JSNice: Inference

```
function chunkData(e, t)
  var n = [];
  var r = e.length;
  var i = 0;
  for (; i < r; i += t)
    if (i + t < r)
      n.push(e.substring(i, i + t));
    else
      n.push(e.substring(i, r));
  return n;
```

# JSNice: Inference

```
function chunkData(e, t)
  var n = [];
  var r = e.length;
  var i = 0;
  for (; i < r; i += t)
    if (i + t < r)
      n.push(e.substring(i, i + t));
    else
      n.push(e.substring(i, r));
  return n;
```



Unknown properties:

t      r      i      ...  
○      ○      ○      ...

Known properties:

o      []      length

# JSNice: Inference

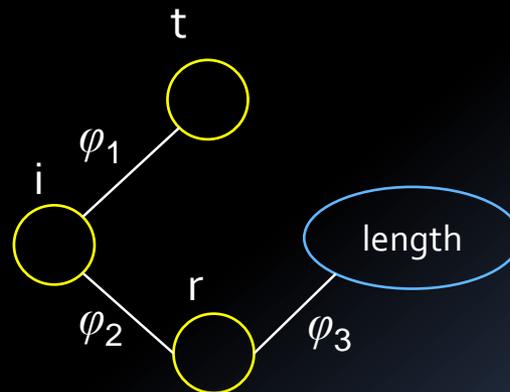
```
function chunkData(e, t)
  var n = [];
  var r = e.length;
  var i = 0;
  for (; i < r; i += t)
    if (i + t < r)
      n.push(e.substring(i, i + t));
    else
      n.push(e.substring(i, r));
  return n;
```



Unknown properties:



Known properties:



# JSNice: Inference

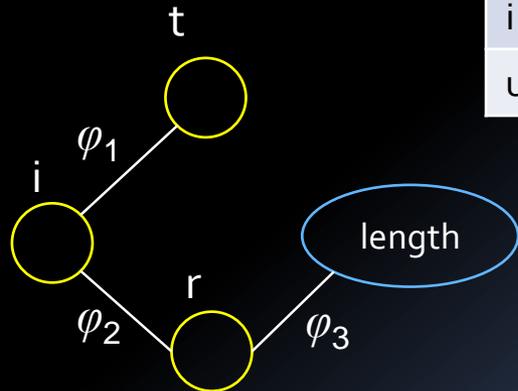
```
function chunkData(e, t)
  var n = [];
  var r = e.length;
  var i = 0;
  for (; i < r; i += t)
    if (i + t < r)
      n.push(e.substring(i, i + t));
    else
      n.push(e.substring(i, r));
  return n;
```



Unknown properties:



Known properties:



i	t	$\varphi_1$
i	step	0.5
j	j	0.4
i	j	0.2
u	q	0.05

# JSNice: Inference

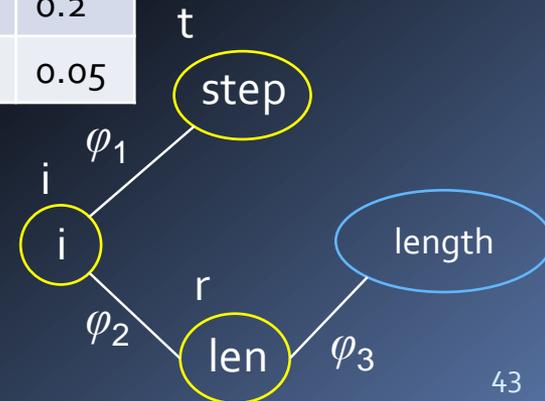
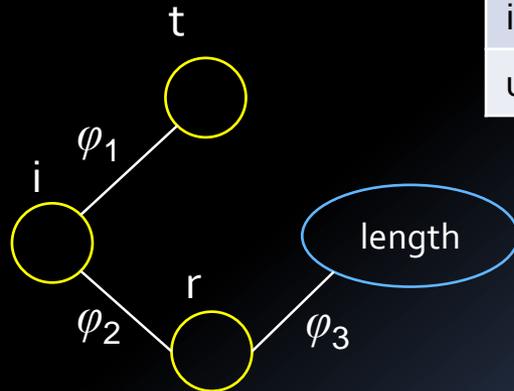
```
function chunkData(e, t)
  var n = [];
  var r = e.length;
  var i = 0;
  for (; i < r; i += t)
    if (i + t < r)
      n.push(e.substring(i, i + t));
    else
      n.push(e.substring(i, r));
  return n;
```



Unknown properties:



Known properties:



i	t	$\varphi_1$
i	step	0.5
j	j	0.4
i	j	0.2
u	q	0.05

# JSNice: Inference

```
function chunkData(e, t)
  var n = [];
  var r = e.length;
  var i = 0;
  for (; i < r; i += t)
    if (i + t < r)
      n.push(e.substring(i, i + t));
    else
      n.push(e.substring(i, r));
  return n;
```

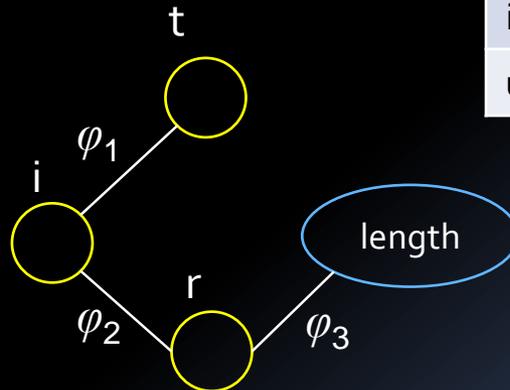
```
function chunkData(str, step)
  var colNames = [];
  var len = str.length;
  var i = 0;
  for (; i < len; i += step)
    colNames.push(str.substring(i, i + step));
  else
    colNames.push(str.substring(i, len));
  return colNames;
```



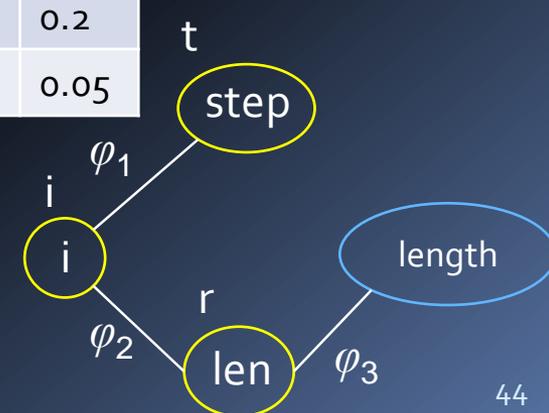
Unknown properties:



Known properties:



i	t	$\varphi_1$
i	step	0.5
j	j	0.4
i	j	0.2
u	q	0.05



# Statistical Program Analysis and Synthesis

Statistical Prediction of Program Properties, [POPL'15](#)  
(V. Raychev, M. Vechev, A. Krause)



The screenshot shows the JS-RICE web interface. The left pane, titled 'ENTER JAVASCRIPT', contains the following code:

```
1 // Put your JavaScript here that you want to rename, deobfuscate, or infer types
2 foo!
3
4 // Example:
5 function chunkData(e, t) {
6   var n = e.length;
7   var s = 0;
8   for (; t <= n; t += 5) {
9     if (t <= 1 + 2 * t)
10      m.push(e.substring(t, t + 3));
11     m.push(e.substring(t, t));
12   }
13   return m;
14 }
15
```

The right pane, titled 'RESULT', shows the transformed code:

```
1 /**
2  * @param {string} str
3  * @param {number} step
4  * @return {[]}
5  */
6 function chunkData(str, step) {
7   var chunk = [];
8   var colnames = [];
9   var len = str.length;
10  var group = [number] + [];
11  for (i = 0; i < len; i += step) {
12    if (i <= 1 + 2 * i)
13      colnames.push(str.substring(i, i + step));
14    colnames.push(str.substring(i, len));
15  }
16  return colnames;
17 }
18
```

structured prediction with discriminative graphical models

+

features based on alias analysis important

# Statistical Program Analysis and Synthesis

Statistical Prediction of Program Properties, [POPL'15](#)  
(V. Raychev, M. Vechev, A. Krause)

```
// Put your JavaScript here that you want to rename, deobfuscate, or infer types
// Example:
function chunkData(str, n) {
  var m = str.length;
  var i = 0;
  for (; i < m; i += n) {
    m.push(str.substr(i, n));
  }
  return m;
}

// RESULT
function chunkData(str, step) {
  var chunk (str, step);
  var colnames = [];
  var len = str.length;
  var group (number) =
  for (i = 0; i < len; i += step) {
    colnames.push(str.substr(i, step));
  }
  return colnames;
}
```

structured prediction with **discriminative graphical models**

+

features based on **alias analysis** important

Statistical PL translation, [Onward'14](#)  
(S. Karaivanov, V. Raychev, M. Vechev)

```
C#
Console.WriteLine("Hi");
...

Java
System.out.println("Hi");
...
```

**phrase-based** statistical machine translation

+

parse **prefix grammars** during translation

Statistical Code Synthesis, [PLDI'14](#)  
(V. Raychev, M. Vechev, E. Yahav)

camera.setDisplayOrientation(90);

camera.unlock();

SurfaceHolder holder = getHolder();

score programs with **generative models** such as n-gram and recurrent neural networks (RNN)

+

sequence extraction with **alias and typestate static analysis**