

# Recognizing structural patterns in code

---

BACHELOR PROJECT

CEDRIC WALKER



# Motivation

```
1 TAP version 13 (ABACUS remix)
2 ok (recorded) 1 - Tests_Actions_Closures::test_action_closure
3 ok (recorded) 2 - Tests_Actions::test_simple_action
4 ok (recorded) 3 - Tests_Actions::test_remove_action
5 ok (recorded) 4 - Tests_Actions::test_has_action
6 ok (recorded) 5 - Tests_Actions::test_multiple_actions
7 ok (recorded) 6 - Tests_Actions::test_action_args_1
8 ok (recorded) 7 - Tests_Actions::test_action_args_2
9 ok (recorded) 8 - Tests_Actions::test_action_priority
10 ok (recorded) 9 - Tests_Actions::test_did_action
11 ok (recorded) 10 - Tests_Actions::test_all_action
12 ok (recorded) 11 - Tests_Actions::test_remove_all_action
13 ok (recorded) 12 - Tests_Actions::test_action_ref_array
14 ok (recorded) 13 - Tests_Actions::test_action_keyed_array
15 ok (recorded) 14 - Tests_Actions::test_action_self_removal
16 ok (recorded) 15 - Tests_Admin_includesMisc::test_shorten_url
17 ok (recorded) 16 - Tests_Admin_includesPlugin::test_get_plugin_data
18 ok (recorded) 17 - Tests_Admin_includesPlugin::test_menu_page_url
19 ok (recorded) 18 - Tests_Admin_includesScreen::test_set_current_screen_with
20 ok (recorded) 19 - Tests_Admin_includesScreen::test_post_type_as_hookname
21 ok (recorded) 20 - Tests_Admin_includesScreen::test_post_type_with_special
22 ok (recorded) 21 - Tests_Admin_includesScreen::test_taxonomy_with_special
23 ok (recorded) 22 - Tests_Admin_includesScreen::test_post_type_with_edit_pre
24 ok (recorded) 23 - Tests_Admin_includesScreen::test_post_type_edit_collisio
25 not ok (recorded) 24 - Failure: Tests_Admin_includesScreen::test_help_tabs
26 ---
27 message: 'Failed asserting that Array ('
28 severity: fail
29 ...
30 ok (recorded) 25 - Tests_Admin_includesScreen::test_in_admin
31 ok (recorded) 26 - Tests_Admin_includesTemplate::test_equal
32 ok (recorded) 27 - Tests_Admin_includesTemplate::test_notequal
33 ok (recorded) 28 - Tests_Admin_includesTheme::test_page_templates
34 ok (recorded) 29 - Tests_Admin_includesTheme::test_theme_files
```

```
30 INFO: Pre-instantiating singletons in org.springframework.beans.factory.support.DefaultListableBeanFactory@83e1e: defining beans [dataSource,transac
31 Sep 22, 2011 11:43:18 AM org.springframework.web.context.ContextLoader initWebApplicationContext
32 INFO: Root WebApplicationContext: initialization completed in 2419 ms
33 Sep 22, 2011 11:43:18 AM org.apache.catalina.util.SessionIdGenerator createSecureRandom
34 INFO: Creation of SecureRandom instance for session ID generation using [SHA1PRNG] took [266] milliseconds.
35 Sep 22, 2011 11:43:18 AM org.springframework.web.servlet.FrameworkServlet initServletBean
36 INFO: FrameworkServlet 'petstore': initialization started
37 Sep 22, 2011 11:43:18 AM org.springframework.context.support.AbstractApplicationContext prepareRefresh
38 INFO: Refreshing org.springframework.web.context.support.XmlWebApplicationContext@1ca209e: display name [WebApplicationContext for namespace 'petstore']
39 Sep 22, 2011 11:43:18 AM org.springframework.beans.factory.xml.XmlBeanDefinitionReader loadBeanDefinitions
40 INFO: Loading XML bean definitions from ServletContext resource [/WEB-INF/petstore-servlet.xml]
41 Sep 22, 2011 11:43:18 AM org.springframework.context.support.AbstractApplicationContext obtainFreshBeanFactory
42 INFO: Bean factory for application context [org.springframework.web.context.support.XmlWebApplicationContext@1ca209e]: org.springframework.beans.factory
43 Sep 22, 2011 11:43:18 AM org.springframework.beans.factory.support.DefaultListableBeanFactory preInstantiateSingletons
44 INFO: Pre-instantiating singletons in org.springframework.beans.factory.support.DefaultListableBeanFactory@13829d5: defining beans [viewResolver,def
45 Sep 22, 2011 11:43:18 AM org.springframework.web.servlet.FrameworkServlet initServletBean
46 INFO: FrameworkServlet 'petstore': initialization completed in 398 ms
47 Sep 22, 2011 11:43:19 AM org.springframework.web.servlet.FrameworkServlet initServletBean
48 INFO: FrameworkServlet 'remoting': initialization started
49 Sep 22, 2011 11:43:19 AM org.springframework.context.support.AbstractApplicationContext prepareRefresh
50 INFO: Refreshing org.springframework.web.context.support.XmlWebApplicationContext@817e66: display name [WebApplicationContext for namespace 'remoting']
51 Sep 22, 2011 11:43:19 AM org.springframework.beans.factory.xml.XmlBeanDefinitionReader loadBeanDefinitions
52 INFO: Loading XML bean definitions from ServletContext resource [/WEB-INF/remoting-servlet.xml]
53 Sep 22, 2011 11:43:19 AM org.springframework.context.support.AbstractApplicationContext obtainFreshBeanFactory
54 INFO: Bean factory for application context [org.springframework.web.context.support.XmlWebApplicationContext@17e66]: org.springframework.beans.factory
55 Sep 22, 2011 11:43:19 AM org.springframework.beans.factory.support.DefaultListableBeanFactory preInstantiateSingletons
56 INFO: Pre-instantiating singletons in org.springframework.beans.factory.support.DefaultListableBeanFactory@8990e4: defining beans [/OrderService-ha
57 Sep 22, 2011 11:43:19 AM org.springframework.web.servlet.FrameworkServlet initServletBean
58 INFO: FrameworkServlet 'remoting': initialization completed in 188 ms
59 Sep 22, 2011 11:43:20 AM org.apache.axis.utils.JavaUtils isAttachmentSupported
60 WARNING: Unable to find required classes (javax.activation.DataHandler and javax.mail.internet.MimeMultipart). Attachment support is disabled.
61 Sep 22, 2011 11:43:20 AM org.apache.coyote.AbstractProtocol start
62 INFO: Starting ProtocolHandler ["http-bio-8080"]
63 Sep 22, 2011 11:43:20 AM org.apache.coyote.AbstractProtocol start
64 INFO: Starting ProtocolHandler ["ajp-bio-8010"]
65 Sep 22, 2011 11:43:20 AM org.apache.catalina.startup.Catalina start
66 INFO: Server startup in 7795 ms
67 Sep 22, 2011 11:43:49 AM org.apache.jasper.compiler.TldLocationsCache tldScanJar
```

# General Idea

---

Chop Text into interesting pieces (statements)

Create abstract representation of statements

Find dependencies between the representation (k-means)

Ideal Solution: One Parser rule per found dependency



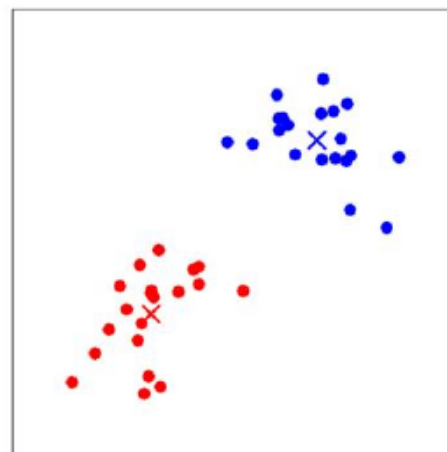
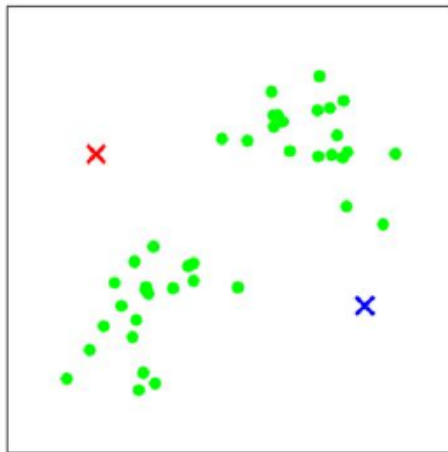
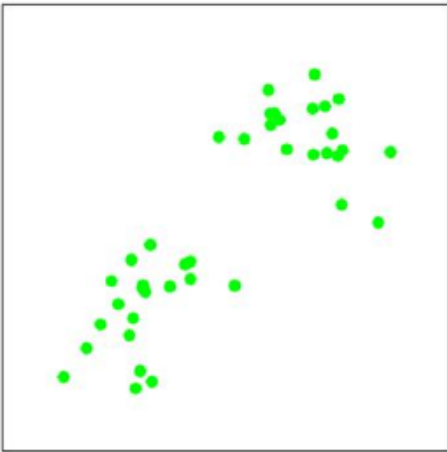
# K-Means

---

Unsupervised learning Algorithm

Clusters Data in k clusters with k centroids

Points of the Cluster have minimal distance to the centroid



# Abstract Representation

---

## Example

```
if ( x < 0 ) return 0 ;
```

```
0 2 0 2 1 2 0 1 2 -> (0 2 0 2 1 2 0 1 2)
```

```
if ( x = 0 ) return 1 ;
```

```
0 2 0 2 1 2 0 1 2 -> (0 2 0 2 1 2 0 1 2)
```

# Problems

---

## Example

```
if ( x < 0 ) return 0 ;
```

```
0 2 0 2 1 2 0 1 2 -> ( 0 2 0 2 1 2 0 1 2 0 )
```

```
if ( x => 6 ) return 1 ;
```

```
0 2 0 2 2 1 2 0 1 2 -> ( 0 0 2 0 2 1 2 0 1 2 )
```

0 for padding



comparing different types of data



Euclidean distance of  
 $\sqrt{8} \approx 2,8$

# Results

---

Change to Notepad++

# Future Work

---

Reversing centroids to text

Find “smarter” way to chop text into pieces

Can this approach distinguish between different languages