Exploratory Test Automation

Harry Robinson
CAST 2010
What is Software Testing?

“Software testing is an investigation conducted to provide stakeholders with information about the quality of the product or service under test.

Software testing also provides an objective, independent view of the software to allow the business to appreciate and understand the risks at implementation of the software.

Test techniques include, but are not limited to, the process of executing a program or application with the intent of finding software bugs.”

- Wikipedia
What is Test Automation?

“Test automation is the use of software to control the execution of tests, the comparison of actual outcomes to predicted outcomes, the setting up of test preconditions, and other test control and test reporting functions. Commonly, test automation involves automating a manual process already in place that uses a formalized testing process.”

- Wikipedia
What is a Test Case?

“A set of test inputs, execution conditions, and expected results developed for a particular objective, such as to exercise a particular program path or to verify compliance with a specific requirement.”

What is Exploratory Testing?

“Exploratory testing is simultaneous learning, test design, and test execution.

In other words, exploratory testing is any testing to the extent that the tester actively controls the design of the tests as those tests are performed and uses information gained while testing to design new and better tests.”

- James Bach,
Exploratory Testing Explained
v.1.3 4/16/03
What is Engineering?

According to Billy Vaughn Koen in *Discussion of the Method*:

“The *engineering method* [is] the strategy for causing the best change in a poorly understood situation with the available resources.” (p. 7)

The fundamental Rule of Judgment in engineering is to evaluate an engineer or his engineering design against the [state of the art] that defines best practice at the time the design was made” (p. 51)
Conceptual blocks

“... mental walls that block the problem-solver from correctly perceiving a problem or conceiving its solution.”

- James L. Adams

Conceptual Blockbusting

**Perceptual**
- Stereotyping
- Difficulty isolating the problem
- Delimiting the problem too closely
- Saturation

**Emotional**
- Fear of risk
- No appetite for chaos
- Judging rather than generating ideas
- Inability to incubate
- Complacency

**Cultural**
- Taboos
- Humor (lack of)
- Reason vs intuition
- Tradition vs change

**Environmental**
- Supportive team
- Supportive management
- Incorporating feedback
- Hidden agendas
“Automated tests represent the automation of a manual process.”

– Harold F. Tipton, Micki Krause

“The most important benefit of automated testing over conventional manual testing is the minimization of costs over repeated tests.”

– Markus Helfen, Michael Lauer
Blocks about Automation & Bugs

“I have never been convinced that finding ‘new’ bugs is a realistic expectation for test automation.”

– I.M. Testy

“After you have run your automated tests for the first time, you are done finding new bugs. Never again will you find a new issue.”

– Steve Rowe

“Running automated test scripts can not be used to find new bugs in the software ...”

– Cordell Vail
Blocks about Test Oracles

“Oracle: A mechanism to produce the predicted outcomes to compare with the actual outcomes of the software under test.”
– BS 7925-1

“A test oracle is a mechanism for predicting the expected results.”
- Andreas Spillner; Tilo Linz; Hans Schaefer

“In real testing the outcome is predicted and documented before the test is run.” – Boris Beizer
Horseless Carriage Thinking

The earliest automobiles were actually horse-drawn carriages where the horse had been replace by an engine.

In fact, most early automobiles came with a buggy whip holder as standard equipment.

*Horseless-carriage thinking* is the tendency of people to use new technologies in exactly the same ways they used earlier technologies.

- William Horton
  April 8, 2000
Challenge assumptions

• Is it necessary to predict the outcome of a test?
• Must an automated test behave like a human?
• Can random inputs be used well?
• Does a test program really need to run quickly?
• How do humans judge if a result is correct?
• How efficient must a test program be?
• Why can’t a test program be used for exploratory testing?
Blockbusting

- Challenge assumptions
- Investigate the problem
- Question all constraints
- Look at object attributes
- Brainstorm ideas
- Who else has done something like this?
- Imagine you have unlimited resources ...
What’s wrong with manual testing?
What’s wrong with test automation scripts?
Humans & Machines

• Moravec’s Paradox
• Humans provide
  – Insight
  – Experience
• Computers provide
  – Calculation
  – Relentlessness
Our Current Challenge

How can we combine the best aspects of human testing with the best aspects of machines to truly deliver state of the art software?
Testing the Square Root Function

\[ \sqrt{a} \]
Testing the Sine Function

Inspired by Doug Hoffman’s article “Heuristic Test Oracles”
Testing Driving Directions
Investigating the Human-Machine Boundary

1.4 miles

11 miles

bug?
Nope, the Lat/Long Heuristic Misled Us
How About Reverse Directions?

**Maps**

<table>
<thead>
<tr>
<th>Start address</th>
<th>End address</th>
</tr>
</thead>
<tbody>
<tr>
<td>CA 92564</td>
<td>CA 90747</td>
</tr>
</tbody>
</table>

Distance: 79.8 mi (about 1 hour 21 mins)
Yes, That Works Very Nicely
Sometimes Very Nicely Indeed

Start address: CA 92564
End address: Long Beach, CA 90813
Distance: 180 mi (about 3 hours 3 mins)

Start address: Long Beach, CA 90813
End address: CA 92564
Distance: 75.1 mi (about 1 hour 21 mins)
Inuyasha Episode 10 Video on AOL Video - Inuyasha episode ten in english.

Plus, watch full episodes of classic TV shows on In2TV, the first broadband video service on the web.

Inuyasha Episode 10 English - AOL Video

Inuyasha Episode 10 English on AOL Video - Inuyasha Episode 10 English


Inuyasha Episode 3 | Watch Inuyasha

Watch Inuyasha Every Single Inuyasha Episode ... Jordan Johnson was just getting on AOL to ... Inuyasha Episode 10, Inuyasha Episode 9, Inuyasha Episode 8, Inuyasha Episode 3

InuYasha Episode 165 | Watch InuYasha

Watch InuYasha Every Single InuYasha Episode ... Jordan Johnson was just getting on AOL to ... Inuyasha Episode 10, Inuyasha Episode 9, Inuyasha Episode 8, InuYasha Episode 165

Inuyasha episode 141 Video

Watch Inuyasha episode 141 and hundreds of other videos about Inuyasha episode 141, kagura, inuyasha... 10

vodpod.com/watch/1004765-inuyasha-episode-141

Safe Search Strict

Hey that's all u got to say INUYASHA+KAGOME=FOREVER NO DUH!!! ur a bitch i wish i could slap the fuck out of u quit hatin on my friend KAGOMEFAN ur probaby some fat kid with no friends sitin at home eating cheetos so FUCK OFF!!!

Go to the page

Internal preview Help improve Bing
what does dilligaf mean?

ALL RESULTS
1-10 of 6,220,000 results

DI- or De-? A Spelling Challenge
... with words beginning with the prefixes do- and di-. There is a difference if you know what the prefixes mean, you can keep track of the difference. So What Does Defenestrate Mean?
englishplus.com/news/news0999.htm · Cached page · Mark as spam

What does MTFBWY mean? - MTFBWY Definition - Meaning of MTFBWY...
Internet Slang. A list of common slang words, acronyms and abbreviations as used in websites, ICQ chat rooms, blogs, SMS, and internet forums.
www.internetslang.com/MTFBWY.asp · Cached page · Mark as spam

Kevheads - Kevin Bloody Wilson Membership area - FAQs
q) what does dilligaf mean? It one of the most common expressions that kev uses is the word dilligaf- meaning to look like I give a dam for the kids does it look like i got ...
kevheads.com.au/pages/faqs · Cached page · Mark as spam
Dead Space - USA XBOX360 (download torrent) - TPB
Region free? Im searching this doing game ... Be warned, the Dead Space developers are watching this torrent and reporting the ... through that exact bunting process with 2 ...

Dead Space wants to update my xbox360 or else it won't play the game...

Wow... one of the best games i played for ages. Amazing job of the creators...

Torrent Reactor - The most active torrents on the web

www.torrentreactor.net - Cached page - Mark as spam

MORE ON THIS PAGE

Yeah .... really what happened to the PC one ... its been a few days since it came out and still no torrent ... wtf

Is this one is next game but this is going to fuck up my 360?
That's Phucked Blog - Information, Comments, Reviews // BlogCatalog
Crazy phucked up blog and articles about anything and everything. Articles include Vagina a how to guide. Why do we love...
www.blogcatalog.com/blogs/thatspucked.html
Cached page - Mark as spam

MySpace - That's Phucked - 31 - Male - BEVERLY HILLS, California...
That's Phucked's Latest Blog Entry [Subscribe to this Blog]. Blog about anything and... hope everything goes well with the site, can't wait for it to be back up!
www.myspace.com/thatspucked
Cached page - Mark as spam

NFL Player Has His COUSIN BEAT UP His 4 Month PREGNANT Girlfriend
THAT'S PHUCKED UP!!! ----- RIP 2 Da KING of ProP, 1988-2009 ---- That's some straight up Ray Carruth shyt! Speakin of dat fool... wonder how's he's doing since...
www.medialinkout.com/2008/26235-nfl_player_has_his_cousin_beat_up_his_4_month_pregnant...
Cached page - Mark as spam

PHUCKED UP - Episode 1
RED That's nothing. One time I stole that mother fuckers pudding cup. You know what that... Until the next random ass episode of "Phucked Up"... ANDREW(VO) Peace.
End
www.simplescripts.com/scripts/PhuckedUp-Episode1.pdf
Cached page - PDF file - Mark as spam

EX-DESTINYS CHURN MEMBERS ON E! // 'CONCRETELOOP.COM' - QUALITY...
If even HALF of that shyt is true, I don't care what NOBODY says, that's PHUCKED UP. Just too much SHADY-NESS going around for me. I guess I'll just say that I'm
Stopping Adult Text

“porn video”
Stopping Adult Text

• Generate and check
  – Input from
    • Log data
    • Mutated combinations
  – Check with
    • Homemade utility
    • PoliCheck
Stopping Adult Text

• Use Best-of-Breed as a test oracle

web

Commercial filter

bing

Results

bing

Results

Do these results differ?
Automating Rules of Thumb

bingrulz
Checking Birthdates

Abraham Lincoln
Birth date: 2/12/1809
Occupation: Lawyer
Spouse: Mary Todd Lincoln

bingrulz
Checking Birthdates

Abraham Lincoln
Birth date: 2/12/1809
Occupation: Lawyer
Spouse: Mary Todd Lincoln

February 12, 1809 (200 years ago)
Checking Birthdates

Abraham Lincoln
Birth date: 2/12/1809
Occupation: Lawyer
Spouse: Mary Todd Lincoln

Abe Lincoln
Political party: National Union
Birth date: 12/2/1809
Occupation: Lawyer

February 12, 1809 (200 years ago)
Checking Entity Cards

Joe Biden

Dick Cheney - Official site
www.whitehouse.gov/vicepresident
Richard Bruce "Dick" Cheney (born January 30, 1941) is the forty-sixth and current Vice President of the...
Checking Entity Cards
Checking Entity Cards

George Bush
Birth name: George Walker Bush
Political party: Republican
Birth date: 6/7/1946
Occupation: Businessman

George H. W. Bush - Wikipedia, the free encyclopedia
George Herbert Walker Bush (born June 12, 1924) was the 41st President of the United States (1989-1993). He was also Ronald Reagan’s Vice President (1981–1989), a congressman...
Checking Entity Cards

George H. W. Bush - Wikipedia, the free encyclopedia
George Herbert Walker Bush (born June 12, 1924) was the 41st President of the United States (1989-1993). He was also Ronald Reagan’s Vice President (1981-1989), a congressman ... en.wikipedia.org/wiki/George_H_W_Bush  Enhanced view · Mark as spam
Checking Entity Cards

George H. W. Bush - Wikipedia, the free encyclopedia
George Herbert Walker Bush (born June 12, 1924) was the 41st President of the United States (1989-1993). He was also Ronald Reagan’s Vice President (1981-1989), a congressman ...
Checking Entity Cards

George Bush Picture  520 x 650  163 kB
Go to page: http://www.ebaumsworld.com/pictures/view/65158/

George H. W. Bush - Wikipedia, the free encyclopedia
George Herbert Walker Bush (born June 12, 1924) was the 41st President of the United States (1989–1993). He was also Ronald Reagan’s Vice President (1981–1989), a congressman ...
Richard M. Nixon
Reconciliation was the first goal set by President Richard M. Nixon. The Nation was painfully divided, with...

Abraham Lincoln
Lincoln warned the South in his Inaugural Address: "In your hands, my dissatisfied fellow countrymen, and not...

Harry S. Truman
During his few weeks as Vice President, Harry S Truman scarcely saw President Roosevelt, and received no...
Testing AutoSuggestion with Heuristics

Thanks to DJ Ramakrishnappa for this example.
Testing AutoSuggestion with Heuristics

Problem: How can we verify an enormous input space?

<table>
<thead>
<tr>
<th>Input</th>
<th>Suggestion</th>
<th>Reason</th>
</tr>
</thead>
<tbody>
<tr>
<td>def</td>
<td>def jam</td>
<td></td>
</tr>
</tbody>
</table>
Testing AutoSuggestion with Heuristics

Problem: How can we verify an enormous input space?

<table>
<thead>
<tr>
<th>Input</th>
<th>Suggestion</th>
<th>Reason</th>
</tr>
</thead>
<tbody>
<tr>
<td>def</td>
<td>def jam</td>
<td>prefix</td>
</tr>
</tbody>
</table>
Testing AutoSuggestion with Heuristics

Problem: How can we verify an enormous input space?

<table>
<thead>
<tr>
<th>Input</th>
<th>Suggestion</th>
<th>Reason</th>
</tr>
</thead>
<tbody>
<tr>
<td>def</td>
<td>def jam</td>
<td>prefix</td>
</tr>
<tr>
<td>nestb</td>
<td>bestbuy.com</td>
<td></td>
</tr>
</tbody>
</table>
Problem: How can we verify an enormous input space?

<table>
<thead>
<tr>
<th>Input</th>
<th>Suggestion</th>
<th>Reason</th>
</tr>
</thead>
<tbody>
<tr>
<td>def</td>
<td>def jam</td>
<td>prefix</td>
</tr>
<tr>
<td>nestb</td>
<td>bestbuy.com</td>
<td>spell correction</td>
</tr>
</tbody>
</table>

Testing AutoSuggestion with Heuristics
Problem: How can we verify an enormous input space?

<table>
<thead>
<tr>
<th>Input</th>
<th>Suggestion</th>
<th>Reason</th>
</tr>
</thead>
<tbody>
<tr>
<td>def</td>
<td>def jam</td>
<td>prefix</td>
</tr>
<tr>
<td>nestb</td>
<td>bestbuy.com</td>
<td>spell correction</td>
</tr>
<tr>
<td>dond</td>
<td>nbc.com</td>
<td></td>
</tr>
</tbody>
</table>
Testing AutoSuggestion with Heuristics

Problem: How can we verify an enormous input space?

<table>
<thead>
<tr>
<th>Input</th>
<th>Suggestion</th>
<th>Reason</th>
</tr>
</thead>
<tbody>
<tr>
<td>def</td>
<td>def jam</td>
<td>prefix</td>
</tr>
<tr>
<td>nestb</td>
<td>bestbuy.com</td>
<td>spell correction</td>
</tr>
<tr>
<td>dond</td>
<td>nbc.com</td>
<td>?</td>
</tr>
</tbody>
</table>
Problem: How can we verify an enormous input space?

Heuristic: Autosuggestions should be “explainable”

<table>
<thead>
<tr>
<th>Input</th>
<th>Suggestion</th>
<th>Reason</th>
</tr>
</thead>
<tbody>
<tr>
<td>def</td>
<td>def jam</td>
<td>prefix</td>
</tr>
<tr>
<td>nestb</td>
<td>bestbuy.com</td>
<td>spell correction</td>
</tr>
<tr>
<td>dond</td>
<td>nbc.com</td>
<td>?</td>
</tr>
</tbody>
</table>
### Testing AutoSuggestion with Heuristics

**Problem:** How can we verify an enormous input space?

<table>
<thead>
<tr>
<th>Input</th>
<th>Suggestion</th>
<th>Reason</th>
</tr>
</thead>
<tbody>
<tr>
<td>def</td>
<td>def jam</td>
<td>prefix</td>
</tr>
<tr>
<td>nestb</td>
<td>bestbuy.com</td>
<td>spell correction</td>
</tr>
<tr>
<td>dond</td>
<td>nbc.com</td>
<td>?</td>
</tr>
</tbody>
</table>

**Heuristic:** Autosuggestions should be “explainable”
Problem: How can we verify an enormous input space?

<table>
<thead>
<tr>
<th>Input</th>
<th>Suggestion</th>
<th>Reason</th>
</tr>
</thead>
<tbody>
<tr>
<td>def</td>
<td>def jam</td>
<td>prefix</td>
</tr>
<tr>
<td>nestb</td>
<td>bestbuy.com</td>
<td>spell correction</td>
</tr>
<tr>
<td>dond</td>
<td>nbc.com</td>
<td>?</td>
</tr>
</tbody>
</table>

Heuristic: Autosuggestions should be “explainable”
Testing AutoSuggestion with Heuristics

Problem: How can we verify an enormous input space?

<table>
<thead>
<tr>
<th>Input</th>
<th>Suggestion</th>
<th>Reason</th>
</tr>
</thead>
<tbody>
<tr>
<td>def</td>
<td>def jam</td>
<td>prefix</td>
</tr>
<tr>
<td>nestb</td>
<td>bestbuy.com</td>
<td>spell correction</td>
</tr>
<tr>
<td>dond</td>
<td>nbc.com</td>
<td>?</td>
</tr>
</tbody>
</table>

Heuristic: Autosuggestions should be “explainable”
Testing AutoSuggestion with Heuristics

Problem: How can we verify an enormous input space?

<table>
<thead>
<tr>
<th>Input</th>
<th>Suggestion</th>
<th>Reason</th>
</tr>
</thead>
<tbody>
<tr>
<td>def</td>
<td>def jam</td>
<td>prefix</td>
</tr>
<tr>
<td>nestb</td>
<td>bestbuy.com</td>
<td>spell correction</td>
</tr>
<tr>
<td>dond</td>
<td>nbc.com</td>
<td>?</td>
</tr>
</tbody>
</table>

Heuristic: Autosuggestions should be “explainable”

- We filter for prefixes
- We filter for spell corrections
- We look at what’s left
Testing AutoSuggestion with Heuristics

Problem: How can we verify an enormous input space?

<table>
<thead>
<tr>
<th>Input</th>
<th>Suggestion</th>
<th>Reason</th>
</tr>
</thead>
<tbody>
<tr>
<td>def</td>
<td>def jam</td>
<td>prefix</td>
</tr>
<tr>
<td>nestb</td>
<td>bestbuy.com</td>
<td>spell correction</td>
</tr>
<tr>
<td>dond</td>
<td>nbc.com</td>
<td>?</td>
</tr>
</tbody>
</table>

Heuristic: Autosuggestions should be “explainable”

Results:
- Covered all 22 million triggers in 7 days
- Detected 2000 bad suggestions
- Eliminated 1.9 million misspelled suggestions
**WEATHER IN LINCOLN, MONTANA**

<table>
<thead>
<tr>
<th></th>
<th>Today</th>
<th>Thu</th>
<th>Fri</th>
<th>Sat</th>
<th>Sun</th>
</tr>
</thead>
<tbody>
<tr>
<td>Foreca</td>
<td>38° / 24°</td>
<td>38° / 19°</td>
<td>44° / 25°</td>
<td>48° / 28°</td>
<td>50° / 31°</td>
</tr>
<tr>
<td>Intellicast</td>
<td>41° / 27°</td>
<td>41° / 21°</td>
<td>48° / 28°</td>
<td>54° / 32°</td>
<td>57° / 34°</td>
</tr>
<tr>
<td>iMap Weather</td>
<td>37° / 24°</td>
<td>38° / 19°</td>
<td>44° / 25°</td>
<td>48° / 28°</td>
<td>49° / 31°</td>
</tr>
</tbody>
</table>

**30°F • (°C)**

Wind: 20 mph W
Humidity: 86%

Turn off History on · Manage history
Table of Contents

Map
Weather
Events
Hotels
Tourism

Seattle
Minneapolis
San Diego
...

SEATTLE
Seattle Weather
Seattle Airport
Seattle Hotels
Seattle Attractions
Map of Seattle
Reference
Table of Contents

Map
Weather
Events
Hotels
Tourism

Seattle
Minneapolis
San Diego
...

Coldplay
Beatles
Pink Floyd
...

Music
Lyrics
Songs
Albums
Tour
Table of Contents

Seattle
Minneapolis
San Diego
...

Map
Weather
Events
Hotels
Tourism

Coldplay
Beatles
Pink Floyd
...

Music
Lyrics
Songs
Albums
Tour
Problem: How can we verify E-card links are relevant?

Testing E-cards with Heuristics
Testing E-cards with Heuristics

Problem: How can we verify E-card links are relevant?
Testing E-cards with Heuristics

Problem: How can we verify E-card links are relevant?

Heuristic: Linked pages should mention the original query term
Testing E-cards with Heuristics

Problem: How can we verify E-card links are relevant?

Heuristic: Linked pages should mention the original query term

Results:

- Tiny test program
- Ran overnight
- Found dozens of errors

Sample results: Country e-cards
How a Production Grammar Works

A → BB
B → CAC
C → DD

A
  └── B
    │  └── B
    │       └── B
    │           └── C
    │               └── A
    │                   └── C
    │                       └── B
    │                               └── B
    │                                           └── D
    │                                               └── D
    │                                                   └── D
    │                                                       └── D
    │                                                            └── D
    │                                                                └── D
    │                                                                    └── D
    │                                                                         └── D
    │                                                                              └── D
    │                                                                                     └── D
Evaluate the following expressions

1. $10 \times 1.0 \times 0.1 \times 6.0 \times 1.16666666667 \times 1.0 \times 1.14285714286 \times 5.25$
2. $2 \times 2.5 \times 0.4 \times 4.0 \times 1.25 \times 1.0 \times 0.8 \times 5.25$
3. $4 \times 0.5 \times 3.5 \times 1.0 \times 0.285714285714 \times 3.0 \times 0.166666666667 \times 42.0$
4. $2 \times 5.0 \times 0.2 \times 1.5 \times 1.0 \times 2.66666666667 \times 1.125 \times 4.66666666667$
5. $5 \times 1.6 \times 1.0 \times 0.5 \times 1.75 \times 0.428571428571 \times 0.333333333333 \times 42.0$
6. $4 \times 2.0 \times 0.75 \times 1.5 \times 0.444444444444 \times 0.75 \times 2.66666666667 \times 5.25$
7. $6 \times 1.5 \times 0.777777777778 \times 0.428571428571 \times 1.333333333333 \times 0.25 \times 7.0 \times 6.0$
8. $2 \times 1.5 \times 0.333333333333 \times 10.0 \times 0.5 \times 0.8 \times 1.5 \times 7.0$
9. $5 \times 0.6 \times 1.66666666667 \times 0.2 \times 1.0 \times 1.0 \times 10.0 \times 4.2$
5 * 8.4
5 * 1.0 * 8.4
1 * 5.0 * 1.0 * 8.4
1 * 1.0 * 5.0 * 1.0 * 8.4
10 * 0.1 * 1.0 * 5.0 * 1.0 * 8.4
8 * 1.25 * 0.1 * 1.0 * 5.0 * 1.0 * 8.4
7 * 1.14285714286 * 1.25 * 0.1 * 1.0 * 5.0 * 1.0 * 8.4
Evaluate the following expression

$$10 \times 0.6 \times 1.5 \times 0.666666666667 \times 1.0 \times 1.0 \times 0.833333333333 \times 1.2 \times 1.5 \times 0.666666666667 \times 1.33333333333 \times 0.375 \times$$

$$1.66666666667 \times 1.0 \times 2.0 \times 1.0 \times 0.9 \times 0.777777777778 \times 1.0 \times 0.428571428571 \times 3.0 \times 0.666666666667 \times 1.5 \times$$

$$0.222222222222 \times 0.5 \times 5.0 \times 2.0 \times 0.4 \times 2.0 \times 0.5 \times 0.75 \times 1.333333333333 \times 0.75 \times 2.33333333333 \times 1.428571428571 \times$$

$$0.7 \times 1.0 \times 1.28571428571 \times 1.0 \times 0.333333333333 \times 2.0 \times 0.833333333333 \times 0.2 \times 6.0 \times 1.33333333333 \times 0.375 \times$$

$$0.666666666667 \times 1.0 \times 2.0 \times 2.25 \times 0.222222222222 \times 4.5 \times 0.888888888889 \times 0.125 \times 4.0 \times 2.0 \times 0.75 \times 1.0 \times 1.5 \times$$

$$0.333333333333 \times 2.33333333333 \times 0.428571428571 \times 3.0 \times 0.888888888889 \times 0.625 \times 0.2 \times 8.0 \times 0.375 \times$$

$$0.333333333333 \times 7.0 \times 1.428571428571 \times 0.1 \times 3.0 \times 3.333333333333 \times 1.0 \times 0.6 \times 0.333333333333 \times 4.0 \times 0.5 \times 2.5 \times$$

$$0.5 \times 1.4 \times 1.14285714286 \times 0.25 \times 0.5 \times 7.0 \times 0.428571428571 \times 2.0 \times 1.0 \times 0.5 \times 0.666666666667 \times 4.5 \times$$

$$0.333333333333 \times 2.33333333333 \times 0.571428571429 \times 1.75 \times 0.428571428571 \times 2.0 \times 0.5 \times 14.0$$
Here’s a grammar for “person”:

\[
\begin{align*}
\text{query} & := \text{<job>} \ \text{<name>} \ | \ \text{<name>} \ \text{<job>} \ | \ \text{<name>} \\
\text{job} & := \text{<job1>} \ | \ \text{<job2>}
\end{align*}
\]

which produces the following strings:

- name
- name job1
- name job2
- job1 name
- job2 name

In the case of name = "brad pitt", job1 = "actor" and job2 = "actress":

- brad pitt
- brad pitt actor
- brad pitt actress
- actor brad pitt
- actress brad pitt
Brad Pitt

With looks that have inspired countless People magazine covers, Internet shrines, and paparazzi blitzkriegs, Brad Pitt is an actor whose very name inspires drooling platitudes, more about male beauty than about acting.
{  brad pitt actor  }
{ brad pitt actress }
actor brad pitt
Angelina Jolie - Wikipedia, the free encyclopedia
Angelina Jolie (born Angelina Jolie Voight; June 4, 1975) is an American actress and ... Jonny Lee Miller and Billy Bob Thornton, Jolie currently lives with actor Brad Pitt, in a ... en.wikipedia.org/wiki/Angelina_Jolie
Enhanced view
· Mark as spam

Brad Pitt - Biography
Add Resume. Brad Pitt products. Shop at Amazon ... Engaged to actress Gwyneth Paltrow. [20 December 1996] Donated $100,000 to the Discovery ... www.imdb.com/name/nm0000093/bio · Cached page · Mark as spam

Videos of actress brad pitt
IE Installation Testing

Diagram:
- 0 to 7
- 7 to 0
- 0 to 8
- 8 to 0
- 0 to 9
- 9 to 0
- 7 to 8
- 8 to 9
- 7 to 9
IE Installation Testing

1. 0 to 7; 7 to 0
2. 0 to 8; 8 to 0
3. 0 to 9; 9 to 0
4. 0 to 7; 7 to 8; 8 to 0
5. 0 to 7; 7 to 9; 9 to 0
6. 0 to 8; 8 to 9; 9 to 0
7. 0 to 7; 7 to 8; 8 to 9; 9 to 0
Google Talk’s Multi-User Chat

#1

#2 #3 #4 #5 #6

me

Angela

Jacob

You have added Angela to this group chat.
You have added Jacob to this group chat.
Angela entered group chat.
Jacob entered group chat.
Jacob: Hey guys
me: hi!
Angela: where are we going hiking tomorrow?
MUC is Governed by a Small Set of Rules

- Anyone logged in can create a chat room
- Anyone in a chat room can invite others into it
- A person can join a chat room if invited
- A chat room disappears if everyone in it leaves
- ...

Now consider how you would test this system manually.
Scenario 1

Angela logs in

Jacob logs in

Angela creates a room

Angela invites Jacob

Jacob accepts
Scenario 2

Angela logs in

Jacob logs in

Angela creates a room

Angela invites Jacob

Jacob declines
Scenario 3

Angela logs in

Jacob logs in

Jacob creates a room

Jacob invites Angela

Angela accepts
Scenario 4

Both log in

Each creates a room

Each invites the other

Both accept

Angela logs out
Scenario 5

Both log in

Each creates a room

Each invites the other

Jacob accepts

Jacob logs out
Hold it!

• We are barely 5 test cases in, and already:
• I’m already losing track of what we’ve covered
• I’m starting to make reckless assumptions
• I haven’t even reached a 3-person scenario yet
Here’s What the Rules Look Like (version 1)
Here's What the Rules Look Like (version 2)
Here’s What the Rules Look Like (version 3)
The Machine Can Now Generate Test Sequences

1. Angela logs in
2. Jacob logs in
3. Angela creates a room
4. Angela invites Jacob
5. Jacob accepts
6. Jacob logs out
What Do We Have to Track?

• Who is logged in
• Who is invited into a room
• Who is in which room

**Bonus**
At any point we are able to oracle the state of the system.
Where’s the Human Touch?

You can start with a small model and build in the direction you think will do the most good for you.
You Can “Steer” the Test Sequences

1. Angela logs in
2. Angela logs out
3. Jacob logs in
4. Jacob logs out
5. Angela logs in
6. Jacob logs in
7. Angela logs out
Welcome to the Hotel California …

Enable arbitrary logouts to force scenarios
Or, Welcome to the Hotel Las Vegas

Penalize arbitrary logouts to ‘encourage’ scenarios
Shortening Repro Scenarios
The Motivation

“The fewer steps that it takes to reproduce a bug, the fewer places the programmer has to look (usually).

If you make it easier to find the cause and test the change, you reduce the effort required to fix the problem. Easy bugs gets fixed even if they are minor.”

- from Testing Computer Software
The Beeline Approach

A repro path is simply another traversal through the state model, so …

1. Choose any 2 nodes in the repro path
2. Find the shortest path between them
3. Execute the spliced ‘shortcut’ path
4. Evaluate the results and repeat
The repro path reduction problem
Random walk finds a bug

... but the repro path is inconveniently long
1. Choose any 2 nodes in the path
2. Find shortest path between them
3. Execute the spliced shortcut path

The bug repro’ed - this is the new shortest path
Continue trimming ...
... until you stop.
Why Reduce the Repro?

- The model can detect (and therefore reduce) both crashing AND non-crashing bugs.
- Finding a shortcut is simple in a model, so the reduction is more efficient.
- Finding bugs is good. Getting them fixed is better.
That Was The Year That Wasn’t

Start
Minimize
Stop
Start
Restore
Date

![Image of a clock showing 4:36 PM on 8/29/99. After clicking the Start button, the date updates to 8/29.]

bug
An 84-step repro sequence

invoke about ok_about no_title doubleclick seconds restore seconds doubleclick doubleclick date about ok_about restore gmt maximize doubleclick doubleclick date seconds date close_clock invoke close_clock invoke close_clock invoke seconds date restore about ok_about no_title doubleclick digital doubleclick doubleclick no_title doubleclick no_title doubleclick seconds restore restore doubleclick doubleclick gmt analog maximize date digital minimize restore minimize close_clock invoke restore digital date minimize close_clock invoke maximize gmt digital restore doubleclick doubleclick about ok_about maximize digital digital digital seconds analog about ok_about about ok_about minimize close_clock invoke restore date
Reducing the Sequence:

• Initial path length: 84 steps
• Shortcut attempt 2 : repro sequence: 83 steps
• Shortcut attempt 3 : repro sequence: 64 steps
• Shortcut attempt 4 : repro sequence: 37 steps
• Shortcut attempt 5 : repro sequence: 11 steps
• Shortcut attempt 7 : repro sequence: 9 steps
• Shortcut attempt 20 : repro sequence: 8 steps
• Shortcut attempt 29 : repro sequence: 6 steps
# Repro Steps Over Time
Useful Regression Testing
Q: What scenario does a developer use to test a fix?
A: The repro scenario you provided!

The Motivation

(type) → (format) → (spell check) → (print)

(type) → (spell check) → (format)
The Gawain* Approach

1. Assign the same weight to each arc in a graph
2. Choose a path through the graph
3. Assign a low weight to each arc in that path
4. Exercise paths in graph in weight-increasing order

* Graph Algorithm Without An Interesting Name
Assign the same weight to each arc
Assign a lower weight to each arc in that path

Weight of this path = 4
Choose a path through the graph
Execute all paths with total weight less than some amount “X”

E.g., weight of this path = 8
Weight of this path = 8
Weight of this path = 8
Weight of this path = 9
Weight of this path = 11
You end up “Cocooning” the regression path
What Kinds of Bugs We Find This Way?
The Incredible Shrinking Clock

Start
Maximize
Stop
Start
Minimize
Stop
Start
Restore
Stop
That Was the Year that Wasn’t

Start
Minimize
Stop
Start
Restore
Date

![Clock]

2:14 PM
1/23/2005

2:14 PM
1/23

⚠️

Bug
Why Does Computer-Assisted Testing Work?

“... I think that less than 10 percent of most programs’ code is specific to the application. Furthermore, that 10 percent is often the easiest 10 percent. Therefore, it is not unreasonable to build a model program to use as an oracle.”

-Boris Beizer, Black Box Testing, p.63
A Quick Look at the Economics of Test Generation

- A typical test engineer
  - Earns $50,000 per year
  - Works 40 to 50 hours per week

- A typical test machine
  - Costs $300 per year to buy and operate
  - Can work 100 hours per week
Should You Hire Additional Testers?

$402,400 per year for 20800 hours of testing
But one human with several machines …

$52,400 per year for 41600 hours of testing
Economics of Computer-Assisted Testing

- 1 tester, 1 cpu: 2080 hrs/yr
- 2 testers, 2 cpus: 4160 hrs/yr
- 1 tester, 10 cpus: 52000 hrs/yr

$
Metrics Issues

Should you count bugs you **prevented**?

Should you count how many test cases you’ve **generated**?
Benefits of Computer-Assisted Testing

- Easy test case maintenance
- Reduced costs
- More test cases
- Early bug detection
- Increased bug count
- Time savings
- Time to address bigger test issues
- Improved tester job satisfaction
Obstacles to Computer-Assisted Testing

• Comfort factor
  – This is not your parents’ test automation

• Skill sets
  – Need testers who can design

• Expectations
  – Some approaches can be a significant upfront investment
  – Will never catch all the bugs

• Metrics
  – Bad metrics: bug counts, number of test cases
  – Better metrics: spec coverage, code coverage
Thanks for coming!