Recording and Replaying Navigations on AJAX Web Sites

Alberto Bartoli
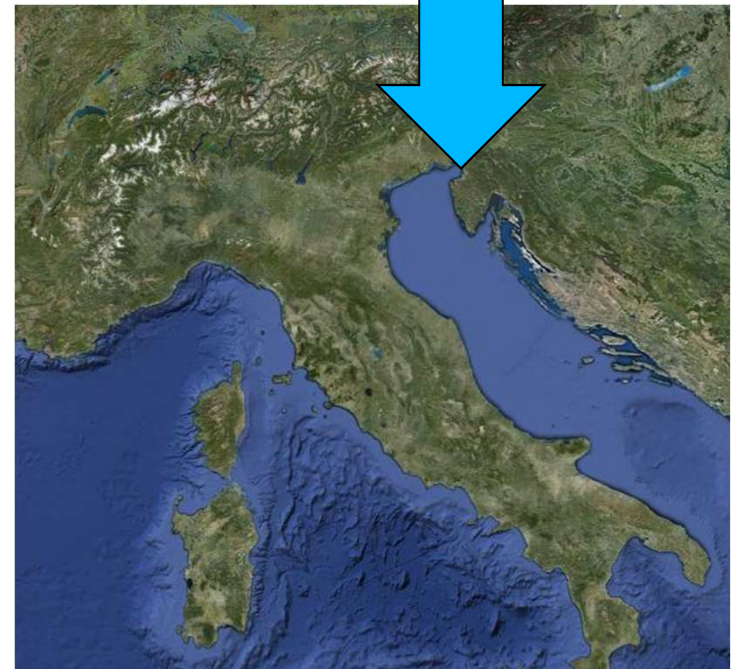
Eric Medvet

Marco Mauri

DI[3] — Università di Trieste, Italy
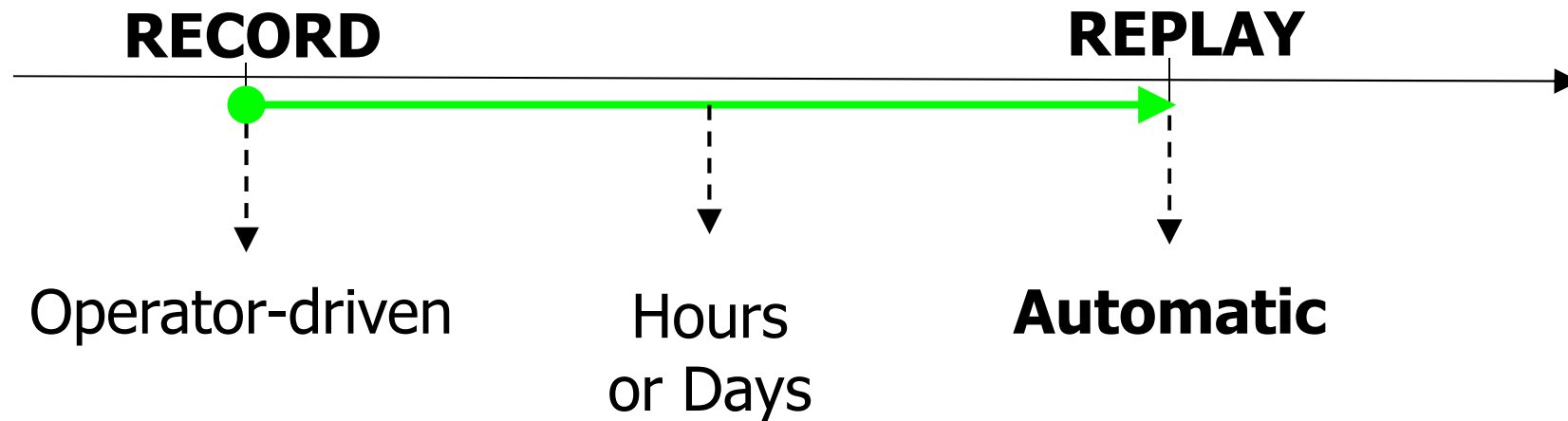

`http://bartoli.inginf.units.it`

July 2012

The Problem

- ❑ Rich Internet Applications (AJAX)

- ❑ Sources not available (Third party, Live site)

RECORD   **REPLAY**

Operator-driven   Hours or Days   **Automatic**
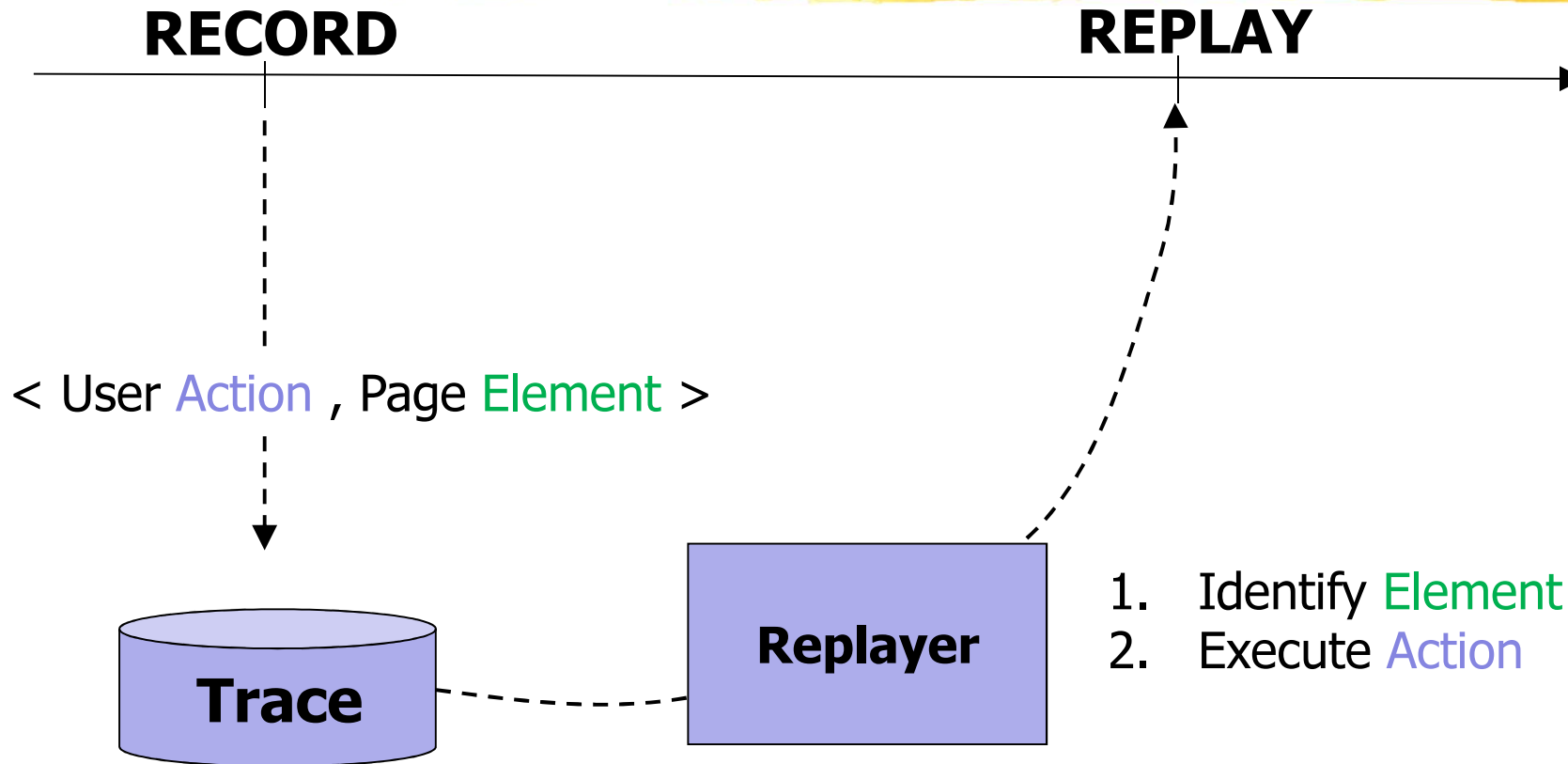
...why bother ?

- Very useful for GUI testing
    - Security
    - Functional
    - Regression
    - ...
    - Web macros
    - Reverse engineering
    - ...

Key Requirement

RECORD   **REPLAY**

< User Action , Page Element >

Trace   **Replayer**

1. Identify Element
2. Execute Action

Identify Element...hhmmm...

❑ DOM@Replay may be very **different** from DOM@Record

    ❑ **Dynamic** content

        ❑ Element in different position...or no longer exists

    ❑ **Session**-based element id

    ❑ Elements **without** id / with **duplicate** id
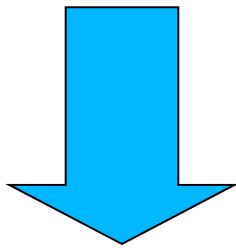
❑ *How to identify "the same" element when replaying a recorded trace ?*

❑ *How to make sure "that" element still exists ?*

That's why...

❑ **Tools for Record & Replay have severe limitations**

    ❑ Non AJAX

    ❑ Assume (implicitly) that Replay occurs immediately after the Record

    ❑ Assume (implicitly) that Replay is always possible

Our Contribution

- **Transparent** Record
  - With Firefox
- **Automated** Replay
  - On the live site**s**

- Works with real sites of practical complexity

- Methodology
- Tool

# Implementation (in a nutshell)

- *Transparent Record*
  - Proxy **injects** Javascript for intercepting user events and recording them in a trace

- *Automated Replay*
  - Several **heuristics** for locating elements

It works...

❑ Realistic complexity
  ❑ Google Search, Shopping, Monclick, +1
  ❑ Twitter login, user search, respond to a tweet
  ❑ Amazon search, share, login and send by email
  ❑ ...

❑ Correct replay
❑ Even after several days
❑ Ability to detect when replay is no longer possible
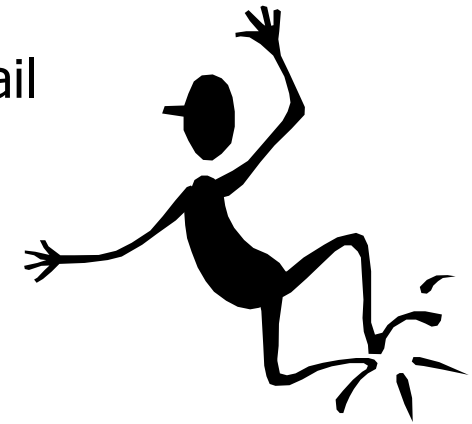
Thanks for your attention...

http://machinelearning.inginf.units.it/

Heuristics

❑ **findElementBySrc**

    ❑ Retrieves the first element with the same tag name and src attribute value

❑ **findElementByInput**

    ❑ Same type of form input and same value for certain attributes

❑ **findElementByGrid**

    ❑ Same tag name and coordinates in a specified grid

Uniqueness of «id»

Website	Href		Id		Name		Src		XPath	
	Precision	Recall	Precision	Recall	Precision	Recall	Precision	Recall	Precision	Recall
Amazon	67.2	89.9	100.0	3.0	100.00	2.9	81.0	7.0	4.5	100.0
Corriere	56.6	82.8	99.2	49.4	69.6	89.8	66.7	12.2	1.9	100.0
Google	78.2	78.0	96.6	29.5	34.6	13.0	100.0	1.0	11.5	100.0
Monclick	59.6	82.2	100.0	2.9	40.0	2.2	47.1	15.6	2.2	100.0

Experiments...
(just one of many)

- Repeat every 6 hours
  - Google login
  - Search «SD 8 GB»
  - Click «shopping» link
  - Select first product
  - Click «compare prices»
  - Select Monclick shop
  - Activate «+1» on product details page
  - Deactivate «+1»

Experiments...
(just another)

- ❑ Repeat every 6 hours
    - ❑ Select «Book» product category on Amazon
    - ❑ Search «Landover»
    - ❑ Select first result
    - ❑ Click «Share»
    - ❑ Execute Amazon login
    - ❑ Fill «Share» form
    - ❑ Click «Send»