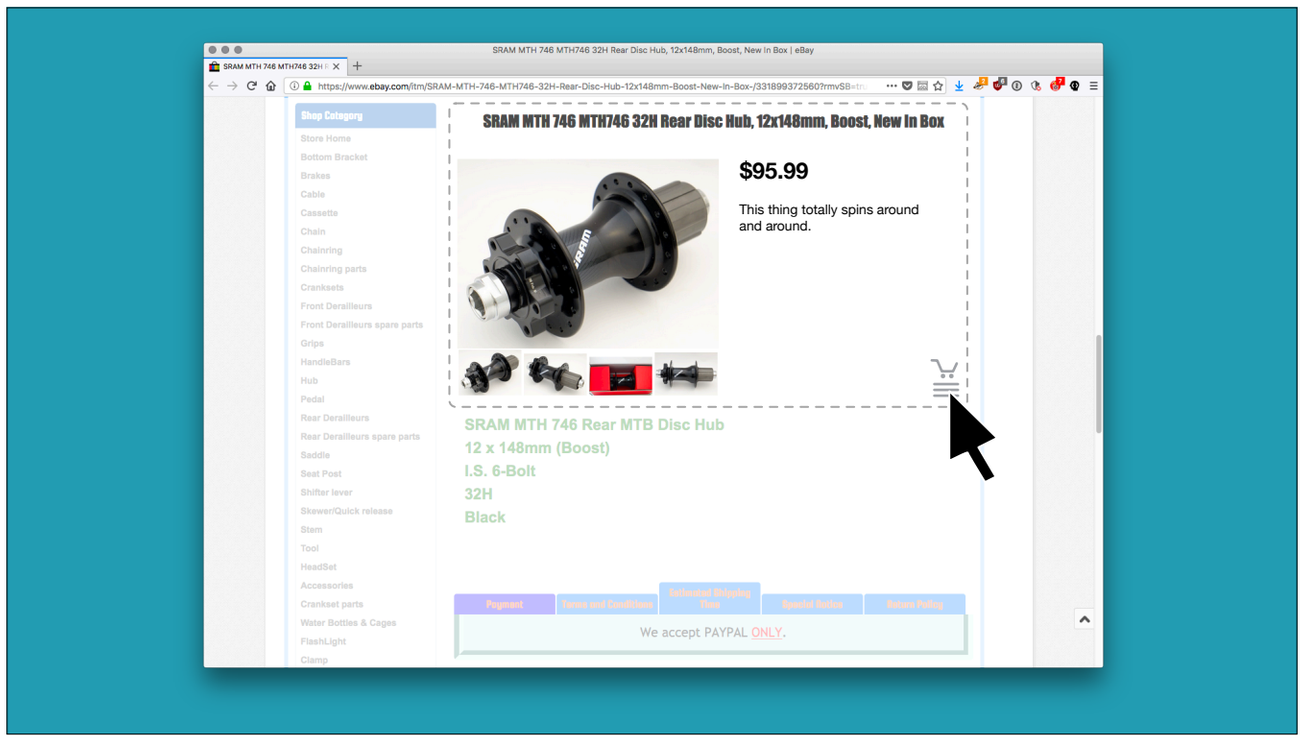

moz://a
all hands 2018
san francisco

About 10 mins.

- browsers
  - **render** well
  - **tabs, bookmarks**
- ~ mere renderer of the **page author's intent**.
  - ~ **PDFs**
  - **user's agent**
  - **next frontier** for that: **viewport**.

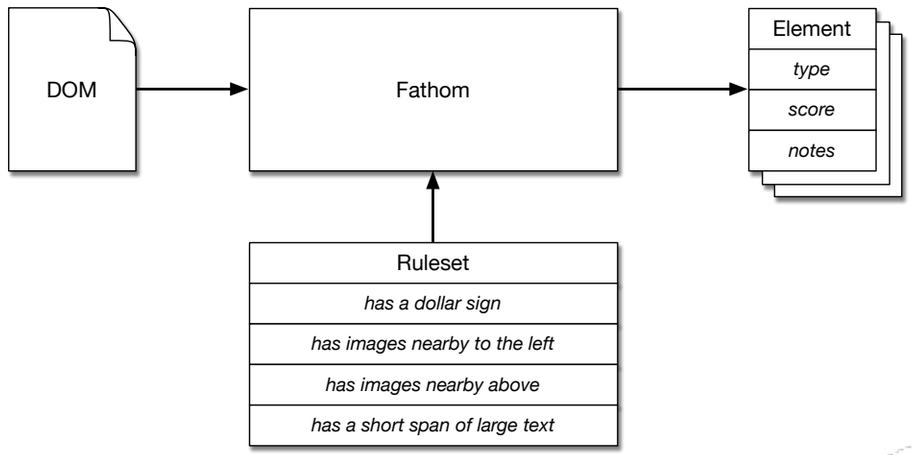
- What if
  - **Drag any product**


SRAM MTH 746 MTH746 32H Rear Disc Hub, 12x148mm, Boost, New In Box | eBay
https://www.ebay.com/itm/SRAM-MTH-746-MTH746-32H-Rear-Disc-Hub-12x148mm-Boost-New-In-Box-/331899372560?rmvSB=tru
Shop Category
Store Home
Bottom Bracket
Brakes
Cable
Cassette
Chain
Chainring
Chainring parts
Cranksets
Front Derailleurs
Front Derailleurs spare parts
Grips
HandleBars
Hub
Pedal
Rear Derailleurs
Rear Derailleurs spare parts
Saddle
Seat Post
Shifter lever
Skewer/Quick release
Stem
Tool
HeadSet
Accessories
Crankset parts
Water Bottles & Cages
FlashLight
Clamp
SRAM MTH 746 MTH746 32H Rear Disc Hub, 12x148mm, Boost, New In Box
$95.99
This thing totally spins around and around.
SRAM MTH 746 Rear MTB Disc Hub
12 x 148mm (Boost)
I.S. 6-Bolt
32H
Black
Payment
Terms and Conditions
Estimated Shipping Time
Special Notice
Return Policy
We accept PAYPAL ONLY.

- **pop-ups**

- Hard
  - HTML has not kept up
    - <**product**>
    - <**popup**>
  - If there was, **authors wouldn't use it**
    - ~ **semantic web**
    - ~ **scraped**
- So let's **understand pages the way humans do, with Fathom**
  - JS framework/mini-language that **embraces the mess**
    - like **search engines**
  - so we can bust into the viewport and bend page content to the user's will
- **key insight**: signal left on the table
  - **CSS** classes and IDs
    - It's where the **semantics** have gone.
  - Durable across **human languages** due to **software ecosystem momentum**
  - Another source of semantics: **Spatial positioning** and layering of elements
- Here's how it works
  - Fathom is the **engine**, providing…

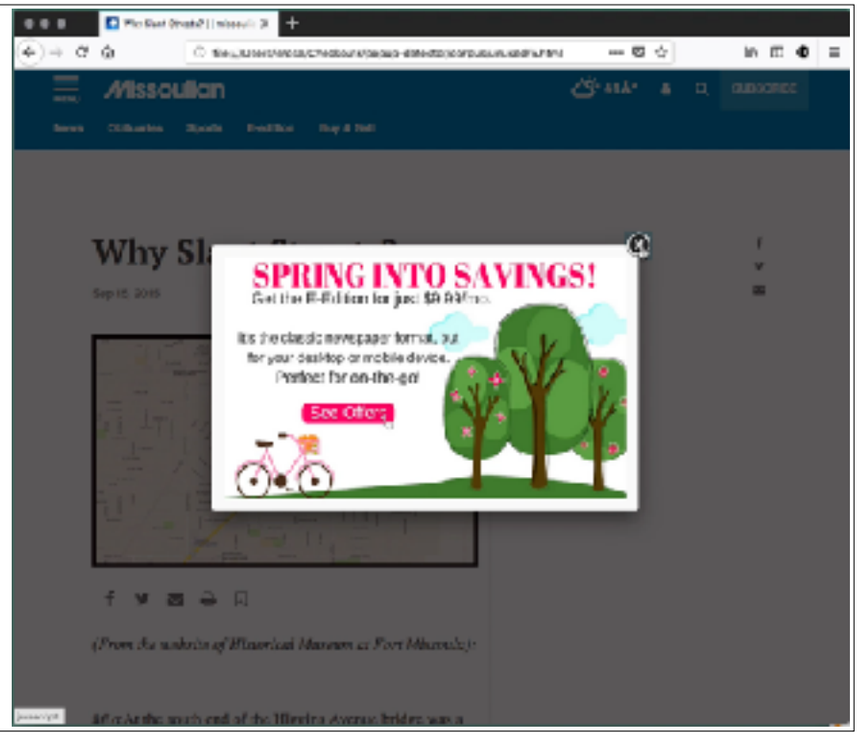
Fathom as Engine



- **data-flow** language
- **categorization** primitives
- **DOM traversal**

- You write **rulesets**, which are the programs it runs.
    - **fuzzy**
    - bags of **characteristics** that, say, products on a page **tend** to have.
    - But you don't have to **get them right**, because
    - After that, you collect maybe 100 **examples**

https://addons.mozilla.org/en-US/firefox/addon/fathomfox/

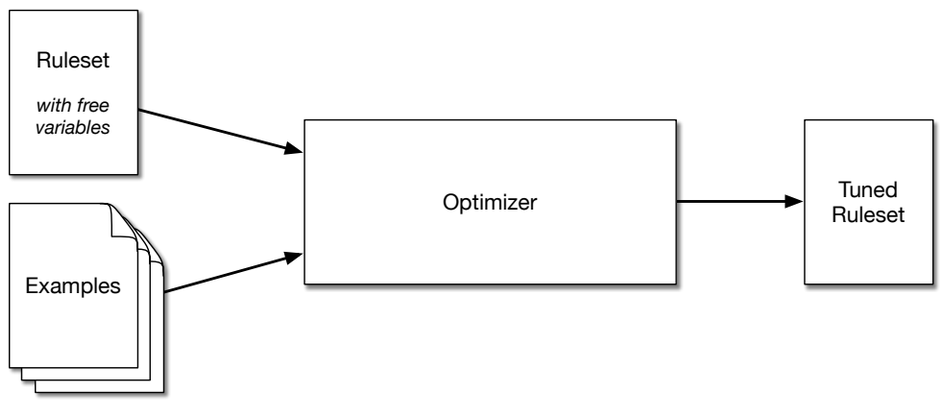


using the **FathomFox** webext

- lets you **label**
- serializes web pages, inlining images, CSS, other resources
- stripping out scripts
    - things are the same every time

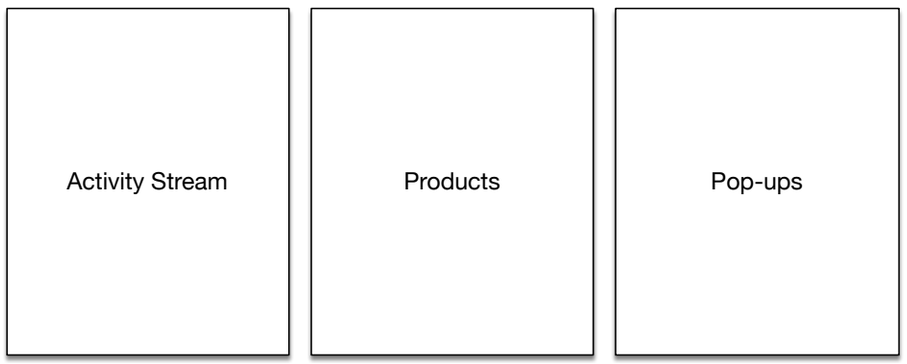
You feed **examples** + your **ruleset** to Fathom's **optimizer**.

Optimizing a Ruleset



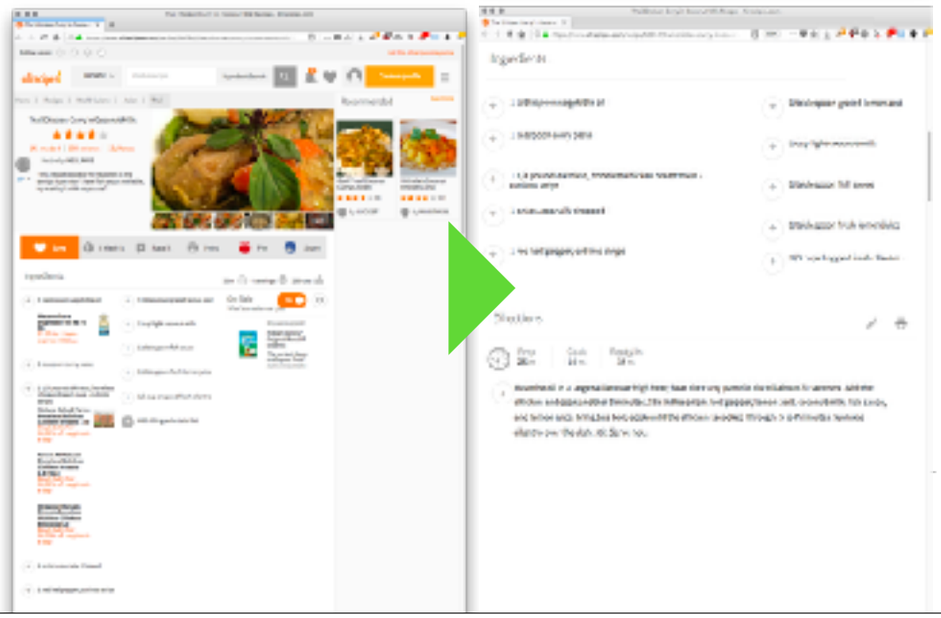
  - Uses **simulated annealing** to come up with **optimal weights** for your rules
  - and **pitches out** ones that don't amount to much.
- At the end, you have a lovely little **encapsulated pearl** that takes in a DOM and labels the entities on it
  - ads
  - slideshows
  - recipes
- And it also tags each with a **confidence**, so you can act accordingly.

Current Uses



- **So far**
    - Powers **Activity Stream**
        - where we replaced a monthly subscription to parsely with 70 lines of code.
    - Wrote PoC **product** recognizer, with 90% accuracy on images, 100% on titles
    - Writing **pop-up recognizer** as we speak
- But I want to see us go **even further**

It's Hard to Scroll with Egg on My Hands



- So many pages are
  - Abusive
  - Ugly
  - Hard to hard to remix
    - **monolithic**. Stuck with **site's own tools**.
    - **Address**→Google Maps. ~ **NASCAR** buttons.
  - Or just inconsistent
    - mystery meat UI
    - hover nav
- Let's use Fathom to make Firefox the browser in which every page
  - **looks beautiful**
  - acts civilized
  - respects the **user's best interest.**

The Future

- Tuning in FF
- Better optimizer math
- Arbitrary predicates
- Optimal decision trees
- OCaml

- I can talk more about
  - **Prolog**-like DSL
  - **Tooling** for example collection

But I also want to know:

- What kind of things are you looking to recognize in pages?
- How would you like to reshape pages?
- How do you wish your experience of the web were different?

I'm here to make that possible.

- privacy of collected study data: just send characteristics
- local rules to add to rulesets: appendability is nice for that