Parallel Performance-Energy Predictive Modeling of Browsers: Case Study of Servo

Rohit Zambre*, Lars Bergstrom^, Laleh Aghababaie Beni*, Aparna Chandramowlishwaran*

*University of California, Irvine

^Mozilla Research

HPC Factory

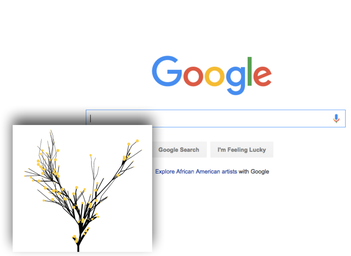
SC16 Salt Lake City, Utah | hpc matters.

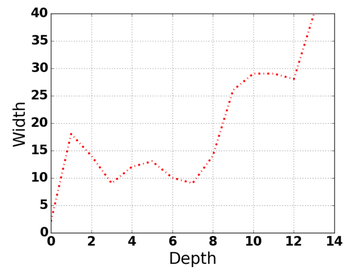
The Problem & Solution

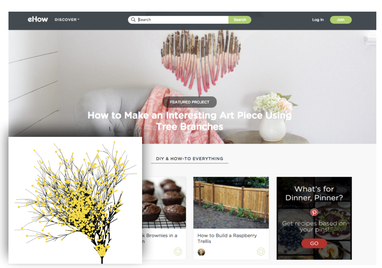
- *What is wrong with today's web browsers?*
  - None exploit multi-core parallelism benefits.
  - None were designed to handle today's highly complex web pages.
  - Slow page-load times are a growing concern, especially on mobile platforms.
- *What is a solution?*
  - A web browser engine that parallelizes its web rendering tasks.
- *But what about parallel overhead?*
  - Need to avoid parallelism when its overhead outweighs its benefits.
- *What is a **practical** solution?*
  - Smart parallelism!
  - Construct a model to predict degree of parallelism in a web page.
  - Use model to decide how many threads to spawn for rendering tasks.

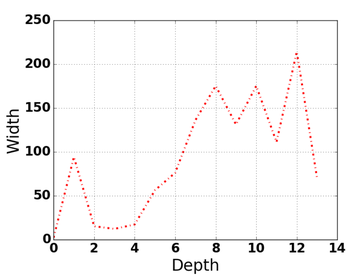
Web Page Characterization

- **The workload of a browser is dependent on the web page it is rendering.**
- Document Object Model (DOM) tree of a web page describes its structure.







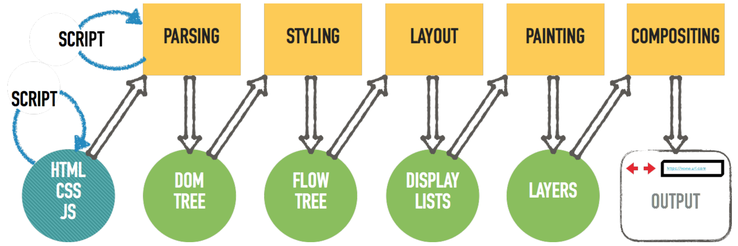


- DOM tree characteristics intuitively predictive of available parallelism:
  - **DOM-size**: total number of nodes in the DOM tree
  - **attribute-count**: total number of attributes in the HTML tags
  - **web-page-size**: size of the web pags' HTML in bytes
  - **number-of-leaves**: number of leaves in the DOM tree
  - **avg-tree-width**: average number of nodes at each level of the tree
  - **max-tree-width**: largest number of nodes at a level of the tree
  - **avg-work-per-level**: average work per level of the tree
- *Why characterize a web page?*
  - These DOM tree characteristics are the input features of the predictive model.

Servo Overview

- Mozilla Research's *Servo*
  - A new, parallel web rendering engine
  - Written in the *Rust* programming language.
  - Uses work-stealing scheduler.
  - Currently parallelism in *Styling* and *Layout* stages.

Servo in action!



- *What is Styling?*
  - Process of applying CSS styles to DOM nodes.
- *What is Layout?*
  - Process of assigning geometric dimensions to web page elements.
- *Where/How is the parallelism?*
  - Multiple DOM nodes can be styled and assigned height/width simultaneously.
  - Executed using parallel, consecutive top-down, bottom-up tree traversals.
- *Why Servo?*
  - The only publicly available parallel web rendering engine.

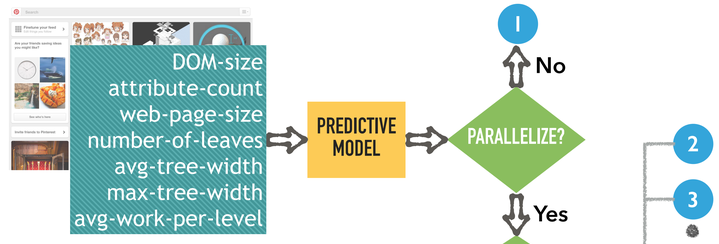
Automated Labeling

- *Why automated labeling?*
  - Eliminates need for domain expert to manually label training data.
- Three cost models
  - Performance Cost Model
  - Energy Cost Model
  - Performance & Energy Cost Model
- Labeling flow
  1. Collect performance & energy data for each web page with 1, 2, 4, 8, … threads.
  2. For each web page, compute speedups and greenups for each thread.
  3. Using a tuned cost model, label the web page.
- *How to label using the Performance/Energy Cost Model?*
  - Using thread that achieves optimal performance/energy.
- *How to label using the Performance & Energy Cost Model?*
  - Using Performance-Energy Tuple (PET) buckets.
  - Each PET bucket has a greenup limit.
  - Choose thread from highest PET bucket that satisfies greenup limit.

Our Labeling Settings	
X_1X_2	$X_1 = 1.1; X_2 = 1.3$
X_2X_3	$X_2 = 1.3; X_3 = 12.87$
Y_1	0.9
Y_2	0.85

Our Experimental Data			
Samples	535 (Alexa Top 500 & 2012 Fortune 1000)	*Threads*	1, 2, 4
Platform	Quad-core Intel Ivy Bridge MacBook Pro	*Performance & Energy Labels*	**1: 59.25%** **2: 9.34%** **4: 31.40%**

Modeling & Results



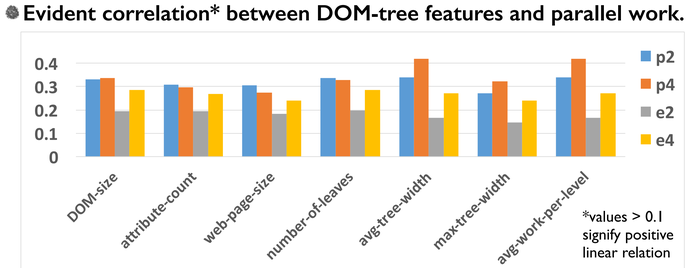
- *Off-the-shelf supervised learning.*
- *Implementation-blind feature space.*
- *Within complete browser execution.*
- *Confident accuracies & high savings!*

Supervised Learning Algorithm	*Evaluation Method*	*Average Accuracy*	*Maximum Accuracy*
Multinomial Logistic Regression (MNR)	*Cross-validation*	*72.22%*	***87.27%***
Ensemble Learning	*Cross-validation*	*71.27%*	***83.63%***
Neural Network	*Holdout-set*	***77.8%***	**-**

- *Why is accuracy not > 90% ?*
  - Heavy class imbalance.
  - Feasible training data set is small.
  - Servo is a young software.
- *How much are we **saving**?*
  - Up to ***91.39%*** on performance.
  - Up to ***46.32%*** on energy.

Conclusion & Vision

- **Evident correlation* between DOM-tree features and parallel work.**



- **Guided, work-load aware scheduling**
  - Use model to guide scheduling and maintain load-balancing.
- **big.LITTLE for parallel web browsers**
  - Exploit and model big.LITTLE parallelism for web rendering tasks.
- **App-clutter-free smartphones**
  - Demonstrate **greener** and **higher** performance than native apps.
- **Revive the universal Web platform**
  - Shift focus of app development to the Web, accessible by all device-types.