Enabling the Transition to the Mobile Web with WebSieve

Michael Butkiewicz, Zhe Wu, Shunan Li, Pavithra Murali, Vagelis Hristidis, Harsha V. Madhyastha (UC Riverside), Vyas Sekar (Stony Brook University)

Presenter: Samarth V Shastri
Introduction

- Mobile platform constitutes a significant share of web traffic (> 20%)
- Performance is quite poor compared to its Desktop/Laptop counterpart
- Reasons for very less mobile-friendly versions of websites
  1. Mobile specific customizations are expensive thereby restricting its adoption to only high end websites
  2. Increase in the complexity of websites; rendering a single web page involves fetching several objects from multiple servers and administrative domains
- Vision: democratize the ability to generate mobile friendly websites through utility maximization problem
- Maximum Utilization problem: Problem of selecting a subset of high utility objects from the original website to render in a short span of time
The 3 high level requirements to be kept in sight before Maximum Utility approach could be seen

- Structure- Awareness
- Utility- Awareness
- Practical Optimization
Motivation

- Opportunity to reduce load times: Sites that have mobile optimized versions have significantly lower load times compared to those that do not.
- Website complexity causes high load times:
  Average website has to fetch over 50 objects over more than 10 server.
- Naïve approaches to tame complexity do not work.

![Graph showing CDF of Websites Page Load Time](image)
WebSieve’s Goal

State of the Art: Manual Reduction of Complexity
WebSieve’s Goal

State of the Art: Manual Reduction of Complexity
Goal: Automate Mobile-Friendly Optimization
WebSieve’s Goal

State of the Art: **Manual** Reduction of Complexity

Goal: **Automate** Mobile-Friendly Optimization
WebSieve’s Goal

State of the Art: **Manual** Reduction of Complexity

Goal: **Automate** Mobile-Friendly Optimization

– Block Low Importance Objects
WebSieve’s Goal

State of the Art: **Manual** Reduction of Complexity
Goal: **Automate** Mobile-Friendly Optimization

– Block Low Importance Objects
– Prioritize High Importance Objects
WebSieve’s Goal

State of the Art: **Manual** Reduction of Complexity

Goal: **Automate** Mobile-Friendly Optimization

– Block Low Importance Objects
– Prioritize High Importance Objects
Vision For WebSieve

Front-end

Back-end

foo.com/x
Our Vision For WebSieve

Front-end

Back-end

Dependency Extraction

foo.com/x
Our Vision For WebSieve

Front-end

User Utility Inference

Object Selection

Back-end

Dependency Extraction

O1  O2  O3  O4

foo.com/x
Our Vision For WebSieve

Front-end

User Utility Inference

Object Selection

Back-end

Dependency Extraction

Fetch Fingerprint

Fetch HTML

foo.com/x

O1  O2  O3  O4
Our Vision For WebSieve

**Front-end**

- **User Utility Inference**
- **Object Selection**
- \{O1, O3\}

**Back-end**

- **Dependency Extraction**

**Fetch HTML**

- foo.com/x
Our Vision For WebSieve

Front-end

User Utility Inference

Object Selection

Back-end

Dependency Extraction

1. HTML Received

2. Load \{O1, O3\}

foo.com/x
Challenge 1. Inferring Object Utility
Challenge 1. Inferring Object Utility
Challenge 1. Inferring Object Utility
Challenge 1. Inferring Object Utility
Challenge 2. Discovering Dependencies
Challenge 2. Discovering Dependencies

Compare prices on hundreds of flights. It’s that simple.

- Roundtrip
- One-way
- Multi-city

From: Los Angeles, CA - Los Angeles Intl (LAX)
To: Brunswick, GA - Glynn Co. Joplin (BQK)
Depart: 02/25/2013
Return: 02/28/2013
Travelers: 

Compare prices we find with (opens new windows):
- Hotwire
- Travelocity
- Priceline
- Orbitz
- Expedia

Also search for
- Hotels
- Cars

Search
Challenge 2. Discovering Dependencies

 Compare prices on hundreds of flights. It’s that simple.

- Roundtrip  
- One-way  
- Multi-city

From: Los Angeles, CA - Los Angeles Intl (LAX)
To: Brunswick, GA - Glynnco Jatport (BQK)
Depart: 02/25/2013  
Return: 02/28/2013  
Travelers:  

Compare prices we find with (opens new windows):
- Hotwire 
- Travelocity 
- Priceline 
- Orbitz 
- Expedia

Also search for
- Hotels 
- Cars
Challenge 3. Predicting New Load Times
Challenge 3. Predicting New Load Times
Challenge 3. Predicting New Load Times

Server Variability

Script Processing
Challenge 3. Predicting New Load Times

Compare prices on hundreds of flights. It's that simple.

- **Roundtrip**  
- **One-way**  
- **Multi-city**

**From:** Los Angeles, CA - Los Angeles Intl (LAX)

**To:** Brunswick, GA - Glycine Jetport (BQK)

**Depart:** 02/25/2013  
**Return:** 02/28/2013  
**Travelers:**

Parallel Connection Limits

Script Processing

Server Variability
Challenge 1. Inferring Utility

Approach: User Study + Machine Learning

STATEMENT: Would removing the 'Object of Interest' greatly impact a user's experience on the website?

- STRONG YES
- YES
- NEUTRAL
- NO
- STRONG NO

Object of Interest

(Has Link 🔗)

Full Website (Object Highlighted)
Challenge 1. Inferring Utility

Approach: User Study + Machine Learning

STATEMENT: Would removing the 'Object of Interest' greatly impact a user's experience on the website?

Object of Interest

(Has Link)

Full Website (Object Highlighted)

Simple Machine Learning Not Enough
Personalization Is Key
Challenge 1. Inferring Utility

Approach: User Study + Machine Learning

- Build a predictive model that takes object attributes as input and estimates the potential utility
- Usage of Amazon terk to carry out survey
- Need for personalization: As variable inputs from different users
- Websieve needs to learn and use classifiers customized to specific users. The user is given freedom to mark objects which are important to him. Based on user response, classifier can be built over time.
- Functional dependencies should be accounted
- Provider utilities should be accounted
Challenge 2. Discover Dependencies

Approach: Block & Infer

- Let $O^w$ be the set of objects of original webpage and $O^{w-oi}$ be the set of objects after blocking $oi$ object from the webpage.
- Assumption: Number of objects of web page does not change over the period of time it takes to infer the dependency structure of the webpage.
- Infer a one to one relationship between $O^{w-oi} - O^w$ and $O^w - O^{w-oi}$ objects.
- Some objects in the original webpage might be replaced by content provider. These can be found in $O^{w-oi} - O^w$.
- True set of objects is the ones that appear in $O^w - O^{w-oi}$ but do not match up with any object in $O^{w-oi} - O^w$.
- Example: $O^w = \{a,b,c,d\}$ $O^w-a = \{c,d\}$ $O^w - O^{w-oi} = \{a,b\}$
  
  $O^w = \{a,b,c,e\}$ $O^w-a - O^w = \{c,d,e\}$
  
  $(O^w - O^{w-oi}) - (O^{w-oi} - O^w) = \{a,b\}$
Challenge 2. Discover Dependencies

Approach: Block & Infer

Original Load

- HTM
- JPG a
- JS 1
- JS 2
- JPG b
- JPG c
Challenge 2. Discover Dependencies

Approach: Block & Infer

Original Load

Block JS 1

HTM L

JPG a

JS 1

JPG b

JPG c

HTM L

JPG a
Challenge 2. Discover Dependencies

Approach: Block & Infer
Challenge 2. Discover Dependencies

Approach: Block & Infer

Original Load

Block JS 1

Block JS 2
Challenge 3. Object Selection

Approach: Page Tree Cut

- Cut Dependency DAG
- Maximize Utility
- Respect Load Time Budget
Challenge 3. Object Selection

Approach: Page Tree Cut

- Cut Dependency DAG
- Maximize Utility
- Respect Load Time Budget
Comments

- Very good approach in terms of selecting maximum utility objects from a large object set which gives importance to dependencies.
- User has the freedom to select objects important to him, Also model builds upon his pattern of usage.
- Decrease in load time will increase popularity of mobile web browsing.
- Positive impact on mobile battery life and network usage.
Wrap Up

• Websieve Stability:
  How often the fingerprint of the individual webpage has to be updated?
• Extrapolating across clients:
  Load time of website is dependent on device. Algorithm to extrapolate load time across clients with a reference device type in particular network setting and characterization of target device and network
• Balancing user-provider Utilities
• Implementation details unclear. Future work has to be carried out
• Lot of work to be carried out in the backend by website provider (read fingerprint)

Discussion
• Use of Internet based Applications instead of mobile websites is viable or not
• Dynamic loading of objects is not taken into account
Thank you
References

• Some slides borrowed from “Enabling the Transition to the Mobile Web with WebSieve”- slideshare.com