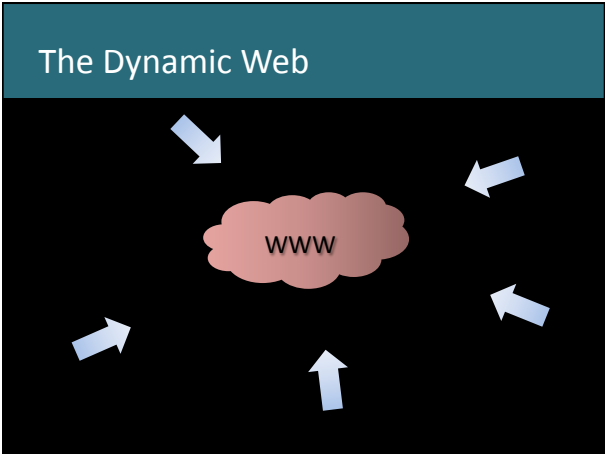


*Zoetrope:
Interacting with the Ephemeral Web*

Eytan Adar, Mira Dontcheva
James Fogarty, Dan Weld

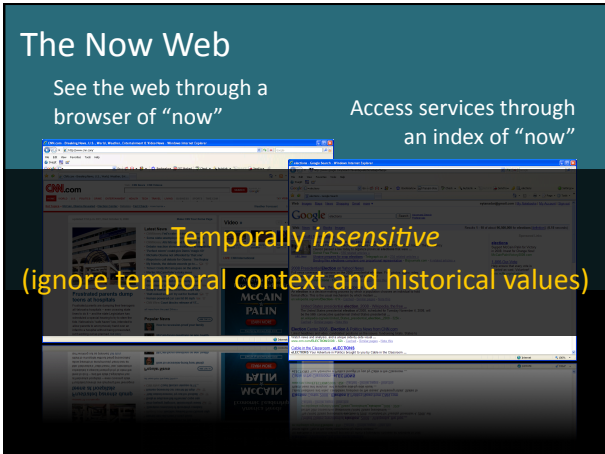
University of Washington & Adobe Systems



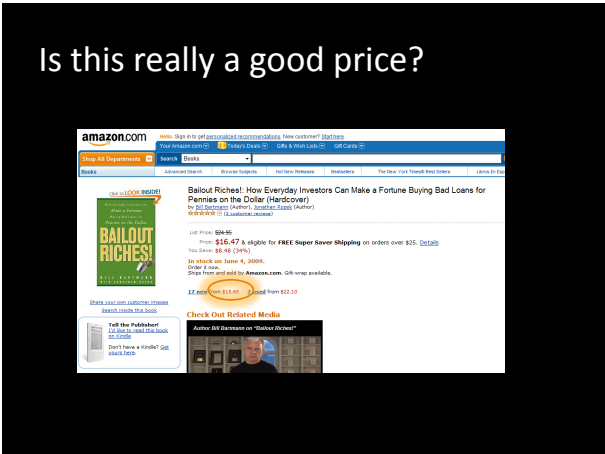
The Now Web

See the web through a browser of "now" Access services through an index of "now"

Temporally insensitive
(ignore temporal context and historical values)



Is this really a good price?



Is this really a good price?
When is really the right time to buy?

What is the relationship between weather and traffic in Seattle?

What is the relationship between traffic and temperature in Seattle?

State Route/Interstate	Route Description	Distance (miles)	Average Travel Time (minutes)	Current Travel Time (minutes)	Via HOV (min.)
98	Auburn to Renton	9.8	12	11	10
405	Bellevue to Bothell	9.7	25	21	11
405	Bellevue to Everett	22.6	52	37	25
405	Bellevue to Federal Way	24.7	60	46	29
520	Bellevue to

What is the relationship between traffic and temperature in Seattle?
an alternative...

Historical weather and climate statistics for Seattle area

State Route/Interstate	Route Description	Distance (miles)	Average Travel Time (minutes)	Current Travel Time (minutes)	Via HOV (min.)
98	Auburn to Renton	9.8	12	11	10
405	Bellevue to Bothell	9.7	25	21	11
405	Bellevue to Everett	22.6	52	37	25
405	Bellevue to Federal Way	24.7	60	46	29

Graph showing temperature (temp) and travel time (travel time) over time.

Related Work

- Evolution of the Web
[Fetterly et al., WWW '03]
[Ntoulas et al., WWW '04]
Internet Archives



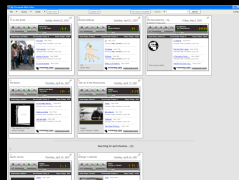
Related Work

- Evolution of the Web
- Monitoring
[Sugiura & Koseki, UIST '98]
[Greenberg & Boyle, GI '06]

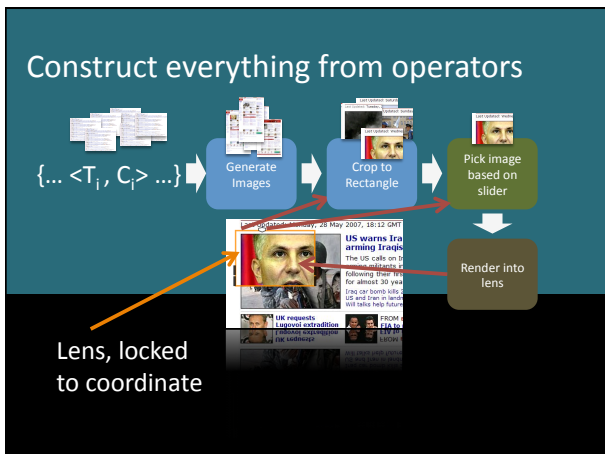
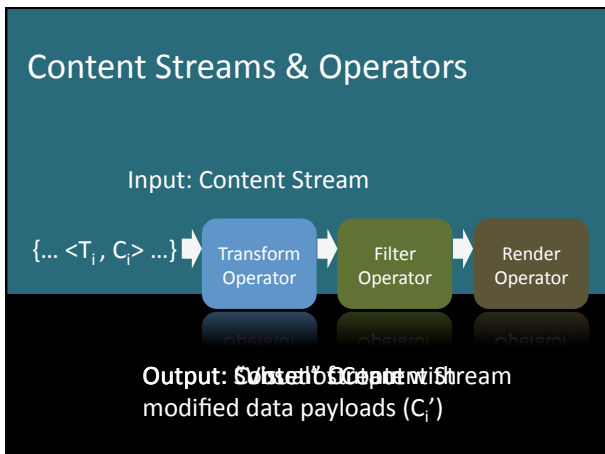
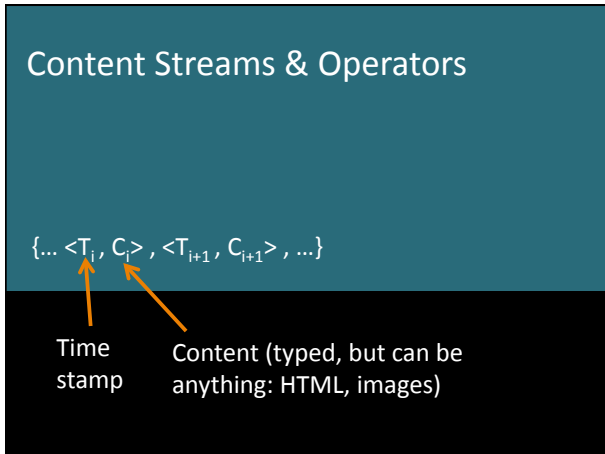
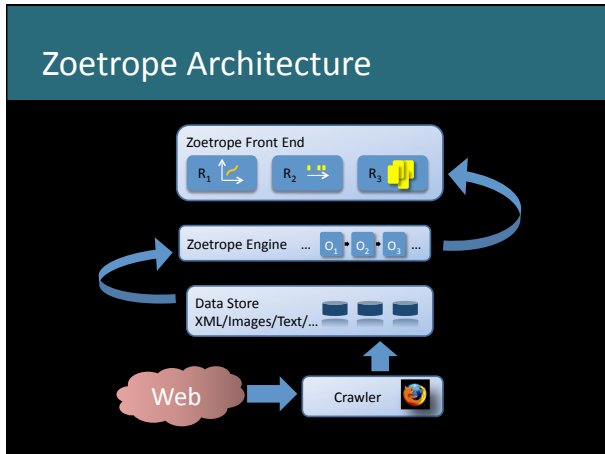


Related Work

- Evolution of the Web
- Monitoring
- Clipping & Aggregating
[schraefel et al., WWW '02]
[Dontcheva et al., UIST '07]



Our Approach



Construct everything from operators

The diagram illustrates a workflow for data extraction from a webpage. It starts with a set of operators $\{... <T_i, C_i> \dots\}$ which are used to select a DOM forest. From this forest, numerical values are extracted, such as 44.10, 49.10, 43.10, and 50.20. These values are then rendered into a time series plot. The background image shows a webpage for 'Planet Earth - The Complete' with a price of \$29.98 and a line graph showing price fluctuations over time.

LENSES

Visual (same place on screen)

A screenshot of a news website interface. A red box highlights a specific article titled 'US warns Iran over arming Iraqis'. The article text includes: 'The US calls on Iran to stop arming militants in Iraq following their first direct t... for almost 20 years. Iraq car bomb kills 20 US and Iran in landmark talks Will talks help future of Iraq?'.

Structural (same place in HTML)

A screenshot of a weather website interface. A tree diagram labeled 'DOM Structure' is overlaid on the right side of the page, showing the hierarchical structure of the HTML elements. The tree has a root node (blue) with three children (blue), and the rightmost child has two children (red).

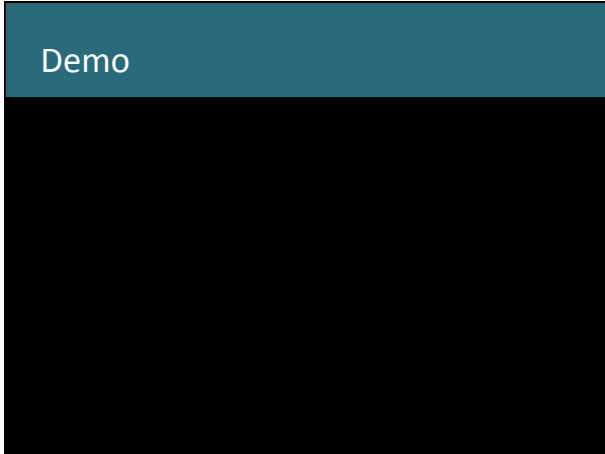
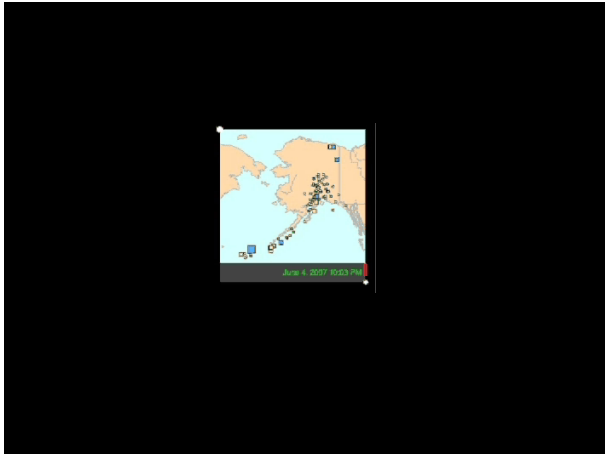
Content (same text)

2007 Standings						
American League						
Team	W	L	Pct	GB	Home	Road
Boston Red Sox	41	22	651	--	20-10	21-12
New York Yankees	31	31	500	9.5	16-13	15-18
Cleveland Indians	31	34	477	11.0	18-15	13-19
Tampa Bay Devil Rays	29	33	468	11.5	17-17	12-15
Baltimore Orioles	29	35	453	12.5	16-14	13-21

Filters

RENDERERS





Contributions

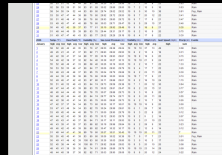
- *Visual programming toolkit* for temporal Web content
- *Semantics* for temporal data streams
- *Operators* for fast processing of temporal data streams
- *Indexing structures* for fast processing and interaction with Web content over time

Future Work

- Evaluation
 - User Study
 - Performance

Future Work

- Evaluation
 - User Study
 - Performance
- Data gathering



Information Arbitrage

- Combine observations to improve granularity



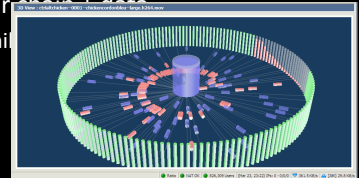
Future Work

- Evaluation
 - User Study
 - Performance
- Data gathering
- Scaling up



Possible Distributed Architecture

- Server collects snapshots
- Client collects finer resolution
 - Gaps?
- DHT for collected data (good for intervals)
- Query: operator chains data
 - Release on simi
 - Repeat



Thanks!

Jaime Teevan, Susan Dumais, Rob Miller , GRAIL, DUB

NSF Graduate Fellowship, ARCS, WRF/TJ Cable Professorship

