

WPI

Implicit Interest Indicators

Mark Claypool
Phong Le
Makoto Waseda
David C. Brown

Computer Science Department
Worcester Polytechnic Institute
Worcester, MA 01609, USA.

The User's Intentions

- ❑ Intelligent interfaces should understand the intentions of the user.
 - ➔ e.g., by interpreting sequences of observable actions.

- ❑ Recommender systems require knowledge of user interests.

- ❑ Can we understand the “interest” the user has in some information?
 - ➔ e.g., in a web page.

- ❑ Can low level actions indicate interest?
 - ➔ e.g., mouse movement, scrolling, ...

Explicit Ratings

- ❑ User explicitly rates information.
 - ➔ Common & fairly precise.

- ❑ Can interrupt normal patterns of reading or action.

- ❑ Users may tire of providing them.
...and...
- ❑ Users need to be convinced of the benefit in order to make the effort.
...but...
- ❑ Many ratings are needed before Collaborative Filtering can provide accurate predictions.

Implicit Ratings

- ❑ Not obtained directly from user.
 - ↳ i.e., some inference needed.
- ❑ Removes cost of obtaining explicit rating.
- ❑ Every interaction could potentially contribute.
- ❑ Can be gathered at little/no cost.
- ❑ May be less accurate.
- ❑ Can combine many implicit ratings.
- ❑ Can combine with explicit ratings.

Research Overview

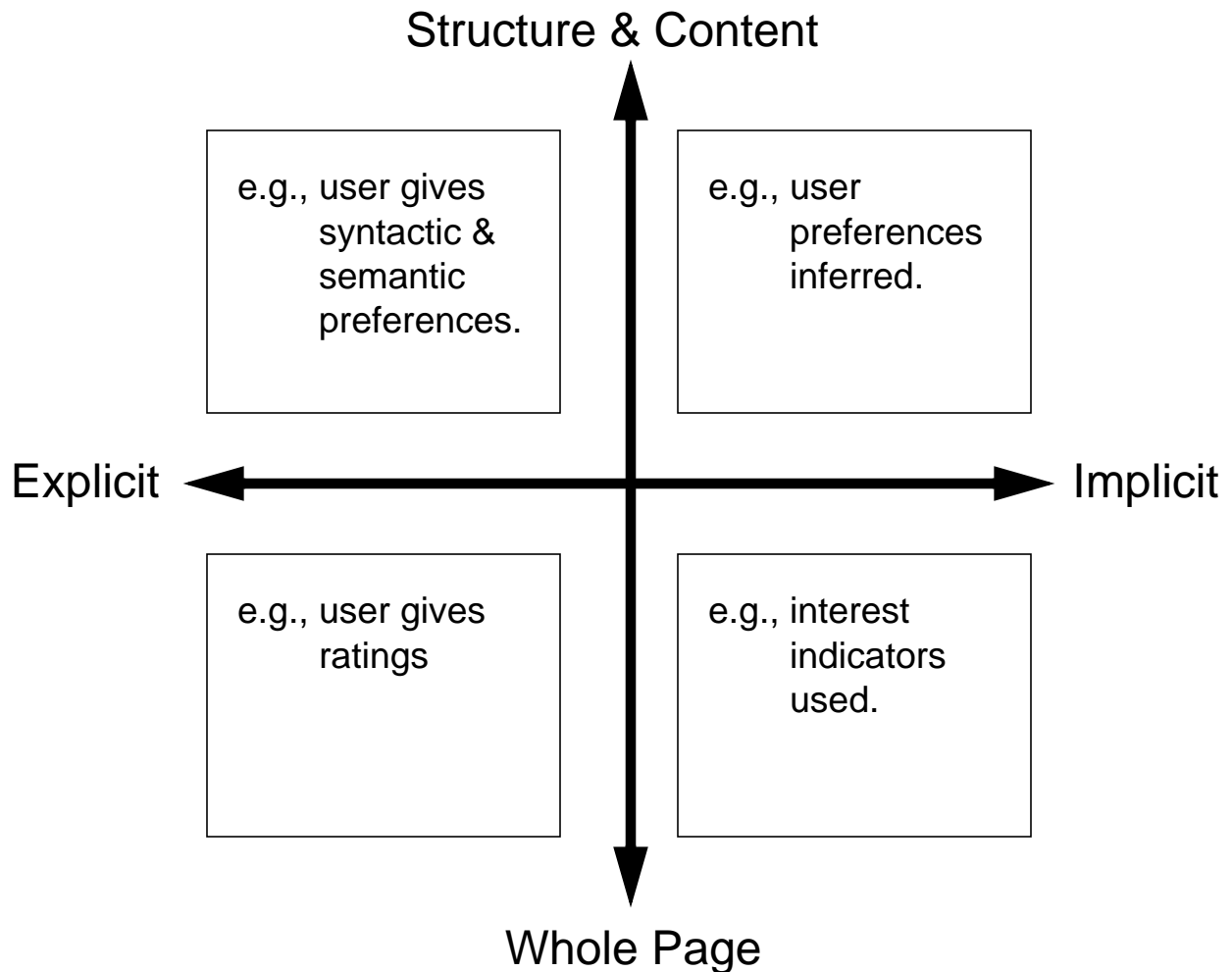
- ❑ Objective is to collect, measure, and evaluate the predictive power of **Implicit Interest Indicators** (i.e., of implicit ratings).
- ❑ Focus on prediction for single web page using a single indicator at a time.
- ❑ Developed web browser, **The Curious Browser**, that captured low level user actions.
- ❑ Used browser in user study of about 80 people browsing over 2,500 web pages.

Dimension of Interest



- ❑ **Explicit:** current user action to express interest; no inference.
- ❑ **Mixed:** past user action (e.g., keywords); some inference.
- ❑ **Implicit:** no user action; inference (e.g., from reading time).

Categorizing Indicators



Indicator Types

- ❑ *Explicit*: user selects from scale.
- ❑ *Marking*: bookmark, save, print, ...
- ❑ *Manipulation*: cut/paste, scroll, search, ...
- ❑ *Navigation*: follow link, read page, ...
- ❑ *External*: eye movement, heart rate, ...
- ❑ *Repetition*: repeated visits, ...
- ❑ *Negative*: not following a link, ...

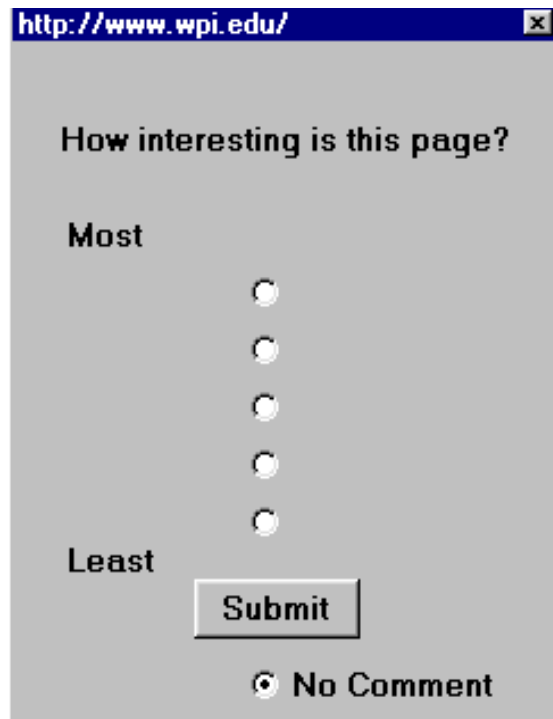
The Curious Browser

- ❑ Familiar GUI.
- ❑ Captures mouse and keyboard actions, and times, to a database, for each page and user.
- ❑ Used Visual Basic, with Internet Explorer version 5.0 html layout engine.

Browser Interface



Evaluation Window



The image shows a screenshot of a web browser window with the address bar displaying "http://www.wpi.edu/". The main content area of the browser contains a form with the following elements:

- Text: "How interesting is this page?"
- Text: "Most" (positioned to the left of the top radio button)
- Five vertically aligned radio buttons.
- Text: "Least" (positioned to the left of the bottom radio button)
- Text: "Submit" (positioned below the radio buttons)
- Text: "No Comment" (positioned below the "Submit" button, with a selected radio button next to it)

- Prompts user for an Explicit Rating when leaving a web page.
- "No Comment" is default.

Activities Captured

- ❑ Mouse:
 - Number of clicks.
 - Time spent moving cursor.

- ❑ Scrollbar:
 - Clicks on scroll bars.
 - Time spent Scrolling.

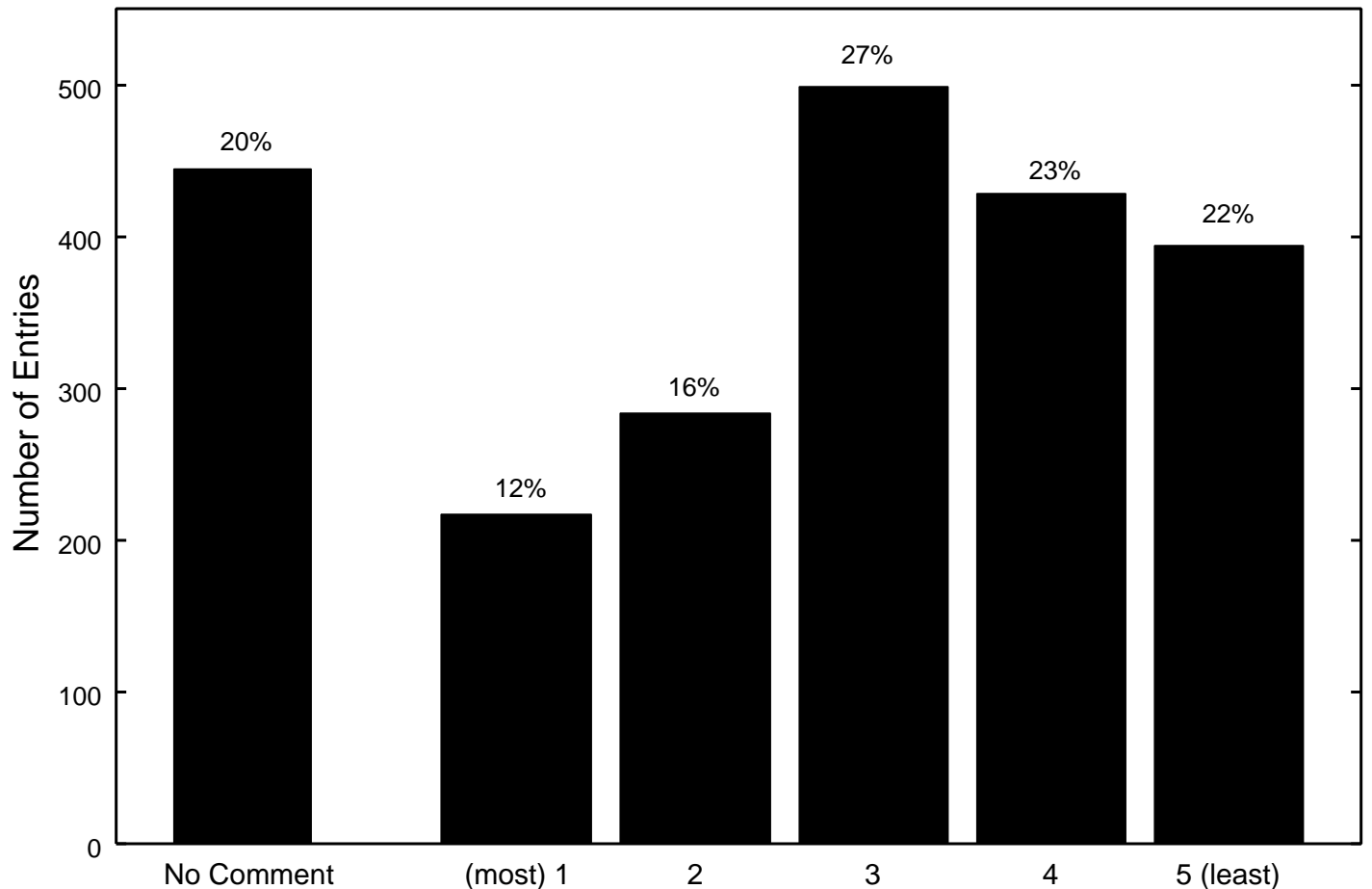
- ❑ Keyboard:
 - Page Up/Down.
 - Up/Down Arrow.
 - Time spent holding down key.

- ❑ Rating:
 - Explicit.

Experiments

- Browser installed on about 40 PCs running Windows 98 in two WPI Labs for about 2 weeks.
- Users told to use it for “browsing”, with no additional task instructions.
- Users were not told the purpose of the experiments.

Explicit Rating Histogram



** Note error in figure: 5 is 'most'.

- 80% of URLs were rated.
- Mean explicit rating was 3.3

Analysis

- ❑ Filtered extreme outliers
 - ↳ (e.g., >20 minutes).

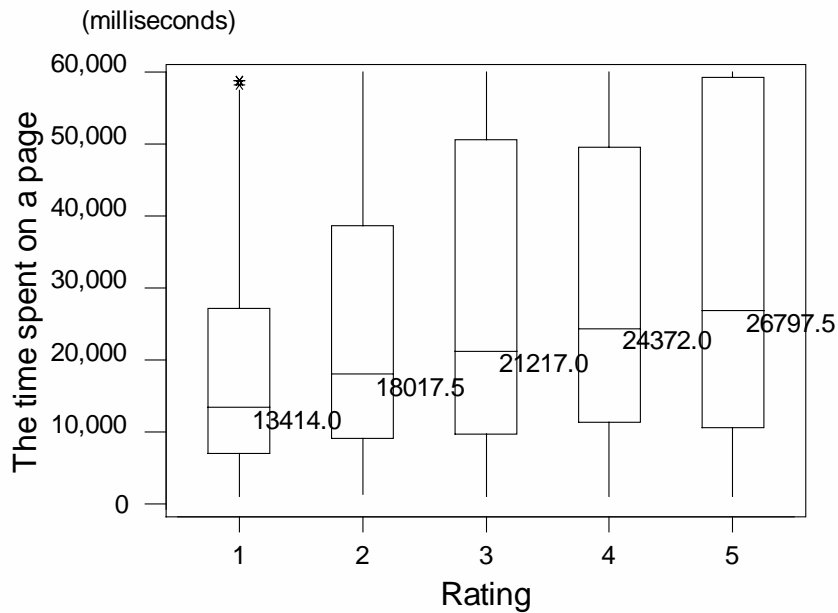
- ❑ Examined Explicit Rating vs. Indicator.

- ❑ Kruskal-Wallis test:
 - ↳ the degree of independence of the medians for each rating.

- ❑ Box plots:
 - ↳ line shows median.
 - ↳ shows 25% to 75% quartiles.

Time on Page

The time spent on a page vs. The explicit rating

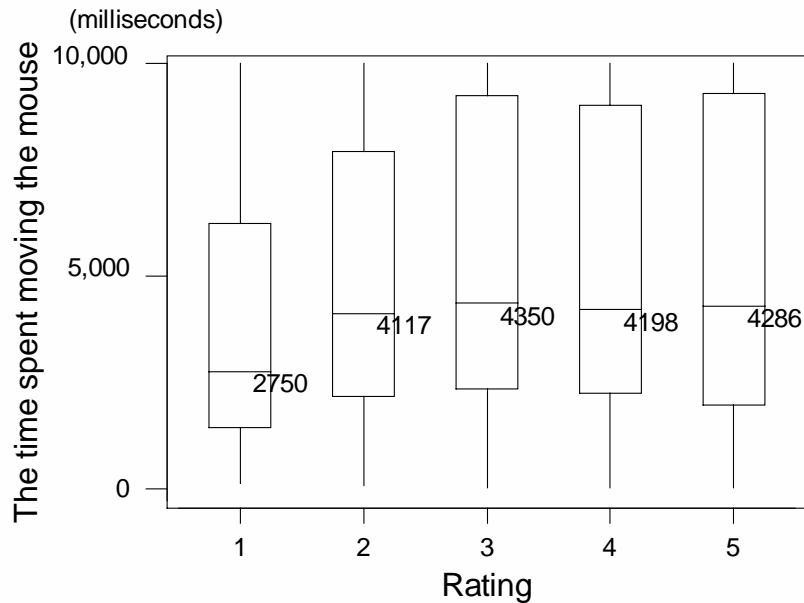


Y-max: 60,000 msec, *: outlier

- median values different.
- appears to be a good interest indicator.

Time Moving Mouse

The time spent moving the mouse vs. The explicit rating

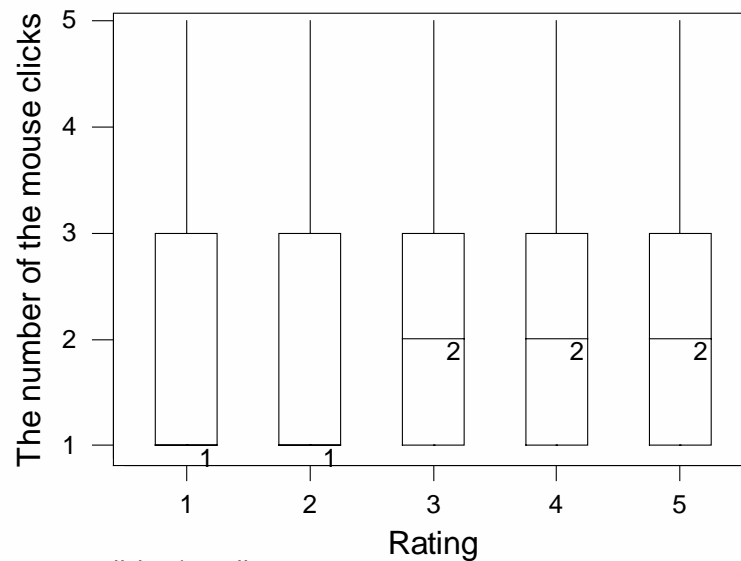


Y-max: 10,000 msec, *: outlier

- median values different.
- appears to be a weak interest indicator.

Number of Mouse Clicks

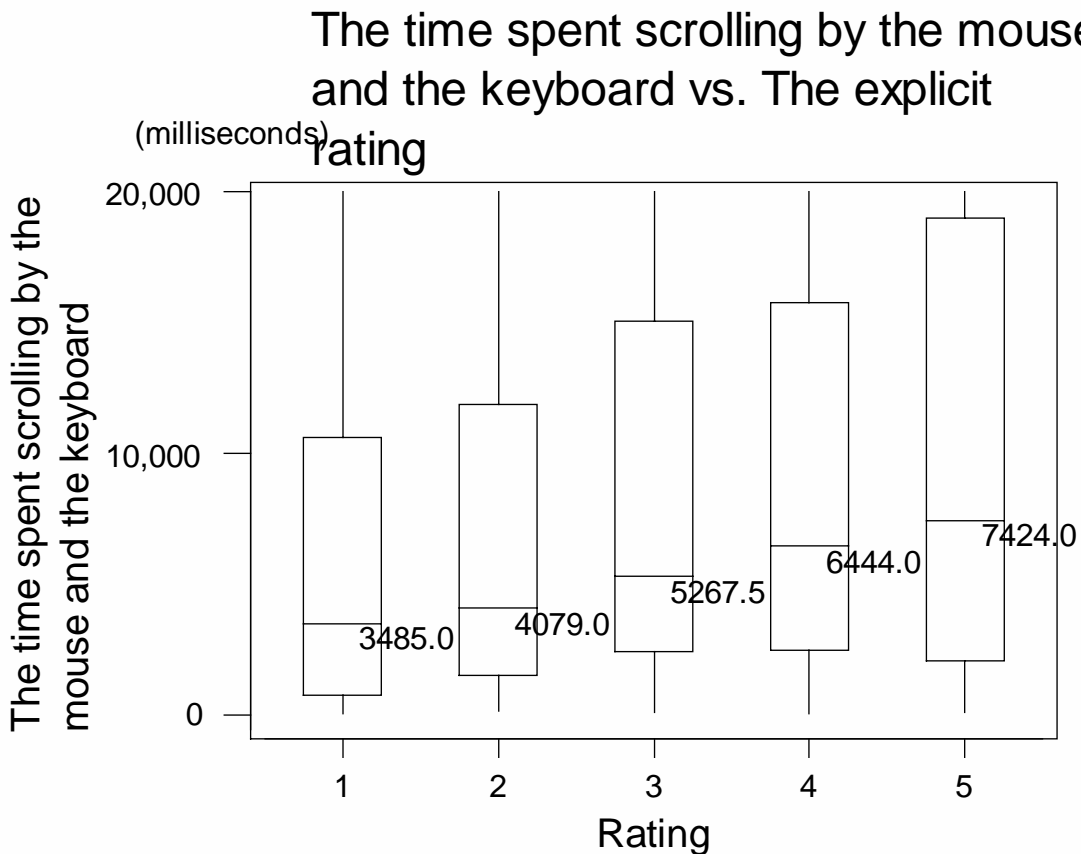
The number of the mouse clicks vs. The explicit rating



Y-max: 5 mouse clicks, *: outlier

- median values not different.
- appears not to be an interest indicator.

Combined Scrolling Time



Y-max: 20,000 msec, *: outlier

- median values different.
- appears to be a good interest indicator.

Rough Accuracy

- ❑ Assume explicit rating is accurate.
- ❑ Assume a “false” prediction is off by >2 wrt explicit interest value.
- ❑ Considering only “true” predictions, time and scrolling each provide about 70% accuracy.
- ❑ In our experiment, explicit rating provided 80% accurate coverage, while implicit interest indicators could provide about 70% accurate coverage.

Contributions

- ❑ correlated with explicit interest:
 - ↳ time spent on page.
 - ↳ amount of scrolling.

- ❑ not well correlated with explicit interest:
 - ↳ number of mouse clicks

- ❑ categories of implicit indicators.

- ❑ the Curious Browser itself.

- ❑ the dataset from the user experiments.

Future Work

- ❑ Combinations of Interest Indicators:
 - e.g., time spent + amount of scrolling.

- ❑ General and personal interest prediction functions.

- ❑ Task dependent interpretation of Interest Indicators.

- ❑ Task determination from Interest Indicators.

- ❑ Additional Interest Indicators:
 - e.g., bookmarking, printing,...