Content or Graphics?
An Empirical Analysis of Criteria for Award-Winning Websites

Rashmi Sinha, Marti Hearst & Melody Ivory
University of California, Berkeley

Maya Draisin
International Academy of Digital Arts and Sciences

For more information: http://sims.berkeley.edu/~sinha/WebbyAwards

Content or Graphics

Question: How can we identify characteristics of good websites on a large scale?

Question: How can we turn these characteristics into empirically validated guidelines?

Content or Graphics

- •Conduct Usability Studies:
  - •Hard to do on a large scale

- •Find a corpus of websites already identified as good!

Use the WebbyAwards database

Content or Graphics

Talk Plan

- Details about WebbyAwards 2000
- Qualities of highly rated websites
- Relative importance of Content & Graphics

- Empirically validated design guidelines
- Future Plans: WebbyAwards 2001, user studies etc.

Content or Graphics

Criteria for submission to the WebbyAwards

- Anyone who has a current, live website
- Should be accessible to the general public
- Should be predominantly in English
- No limit to the number of entries that each person can make

Content or Graphics

Site Category

- Sites must fit into at least one of 27 categories. For example:
  - •Arts
  - •Activism
  - •Fashion
  - •Health
  - •News
  - •Radio
  - •Sports
  - •Music
  - •News
  - •Personal Websites
  - •Travel
  - •Weird

Sites can be listed in multiple categories
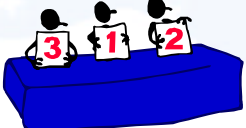
Content or Graphics

3 Stage Judging Process

- **Review Stage: From 3000 to 400 sites**
  - 3 judges rate each site on 6 criteria, and cast a vote if it will go to the next stage
- **Nominating Stage: From 400 to 135 sites**
  - 3 judges rate each site on 6 criteria, and cast a vote if it will go to the next stage

- **Final Stage: From 135 to 27 sites**
  Judges casts vote for favorite site
- **People's Voice: 135 to 27 sites**
  - **Anyone can vote for their favorite site among the ones in the final stage**

Criteria for judging

- 6 criteria
  - Overall Site Experience
  - Five specific Criteria
    - Content
    - Structure and Navigation
    - Visual Design
    - Interactivity
    - Functionality
- Scale: 1-10 (highest)

- Content:
  is the information provided on the site.

  Good content is engaging, relevant, appropriate for the audience-you can tell it's been developed for the Web because it's clear and concise and it works in the medium ...

- Structure & Navigation:
  is the organization of information and navigation.

  Sites with good structure and navigation are consistent and effective. They allow you to form a mental model of the information provided ...

- Visual Design:
  is the appearance of the site.

  Good visual design is high quality, appropriate, and relevant for the audience and the message it is supporting ...

- Interactivity:

  is the way a site allows a user to do something.

  Good interactivity is more than sound effects, and a Flash animation. It allows the user to give and receive. Its input/output in searches, chat rooms, ecommerce etc....

- Functionality:

  is the use of technology on the site.

  Good functionality means the site loads quickly, has live links, and any new technology used is functional and relevant for the intended audience ...

- Overall Experience:

  The overall experience encompasses content, structure and navigation, visual design, functionality, and interactivity, but it also encompasses the intangibles that make one stay or leave...

Webby Judges

– Internet professionals who work with and on the internet: new media journalists, editors, web developers, and other Internet professionals

– have clearly demonstrable familiarity with the category which they review

---

Quick Review of Judging Process

- Review Stage: 3000 to 400 sites, each site judged by 3 judges
- Nominating Stage: 400 to 135 sites, each sites judged by 3 judges

Criteria for judging:

- Content
- Navigation
- Visual Design
- Functionality
- Interactivity
- Overall Experience

Are we sampling from the whole range of websites (good to bad) in the Review Stage?

Review Stage: The whole range of sites, good to bad

Figure 1a. Review Stage: Overall Rating

Overall Rating

OVERALL

Mean = 6.01
SD = 1.59
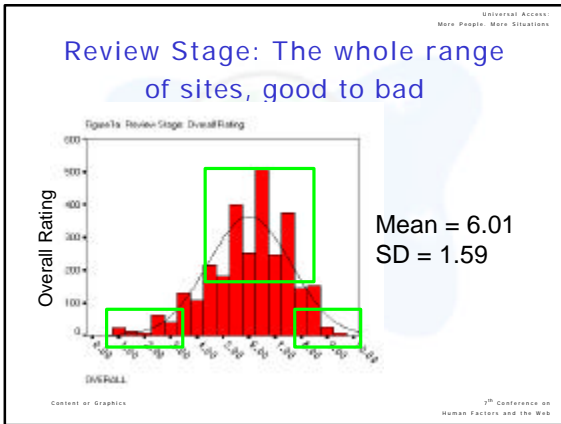
Can overall rating be predicted by specific criteria?

Statistical Technique: Regression analysis
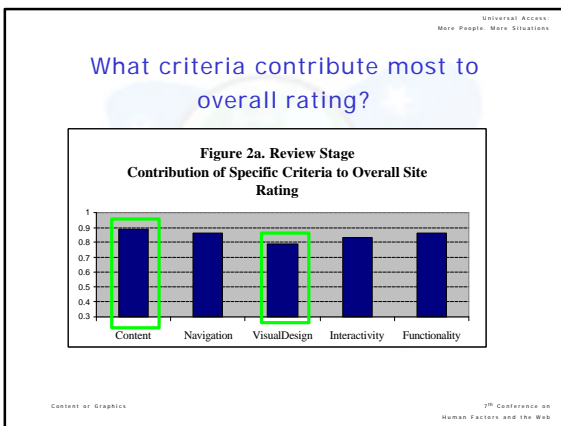Question: What % variance is explained by 5 criteria

Percentage variance explained = 89%

Can votes be predicted by specific criteria?

Statistical Technique: Discriminant analysis
Question: Can we predict the votes from the 5 specific criteria?

Classification Accuracy for Sites = 91%

What criteria contribute most to overall rating?

**Figure 2a. Review Stage
Contribution of Specific Criteria to Overall Site Rating**

Content | Navigation | VisualDesign | Interactivity | Functionality

---

Nominating Stage Analysis

- 6 criteria
  - Content, Structure & Navigation, Visual Design, Functionality & Interactivity
  - Overall experience
- 400 sites
- 3 judges rated each site

Nominating Stage: Top sites for each category

Overall Rating

overall

Mean = 7.6
SD = 1.66

What criteria contribute to overall rating at Nominating Stage?

77% variance explained in overall rating

Contribution of Criteria (Correlation)
Unique Contribution of Criteria (Partial Correlation)

Content | Navigation | VisualDesign | Interactivity | Functionality

Unique Contribution of Content and Visual Design

Partial Correlations — Content / Visual Design

- People's Voice Ratings also indicate that people vote for sites with better content rather than better visual design

---
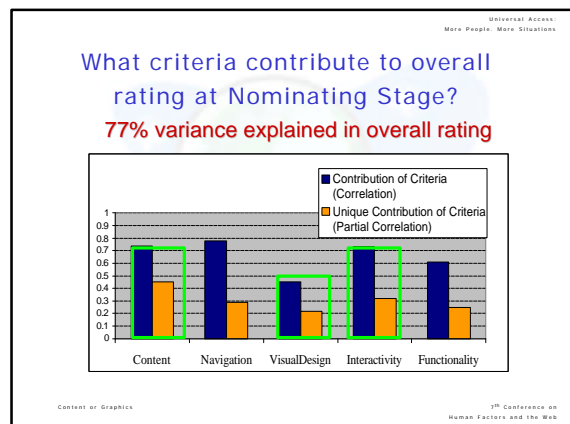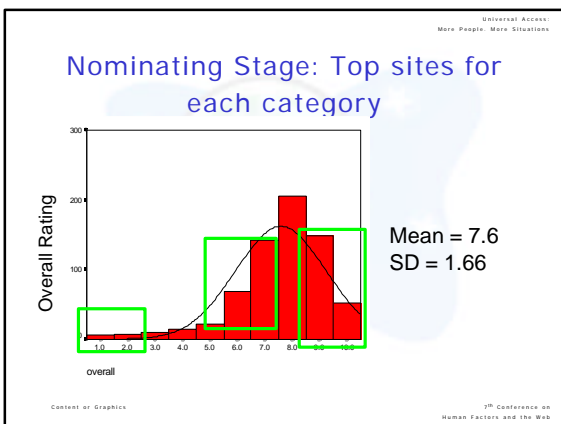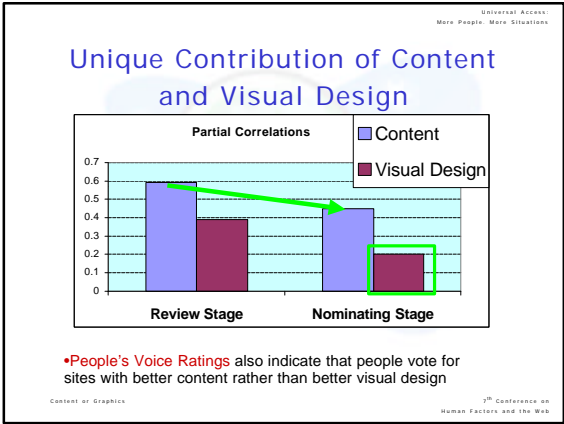
Summary of Findings

- The specific ratings do explain overall experience.
- The best predictor of overall score is content.
- The second best predictor is interactivity.
- The worst predictor is visual design

Are there differences between categories?

- Arts
- Activism
- Fashion
- Health
- News
- Sports
- Music
- News
- Personal Websites
- Travel

Focus on a few Categories
Art, Commerce & Radio

Art

Arts: Contribution of criteria to overall rating
Variance explained = 93%

Content, Navigation, VisualDesign, Interactivity, Functionality

---

Commerce Sites

Commerce Sites: Contribution of criteria to overall rating
Variance explained = 87%

Content, Navigation, VisualDesign, Interactivity, Functionality

---

Radio Sites

Radio Sites: Contribution of criteria to overall rating
Variance explained = 90%

Content, Navigation, VisualDesign, Interactivity, Functionality
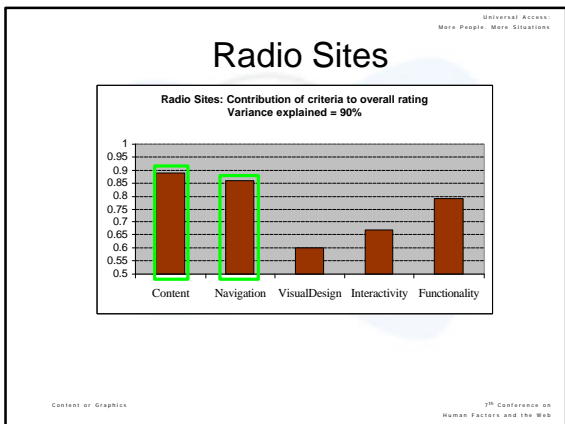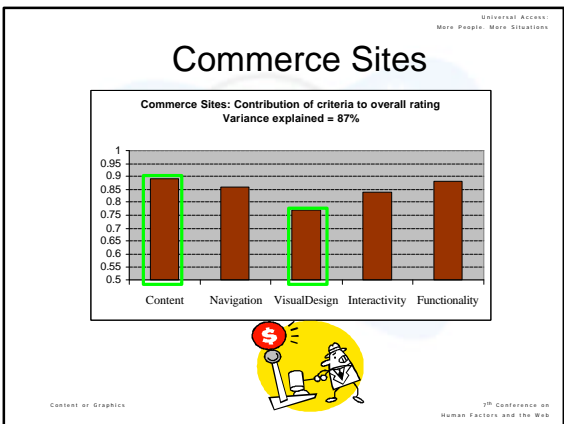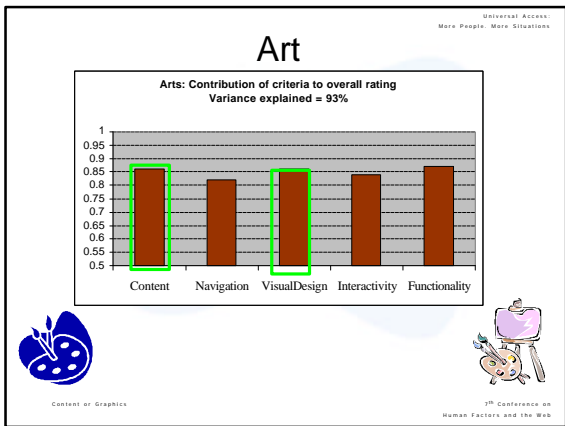
Slide 1 — Conclusions

Conclusions

- The importance of criteria varies by category.

- Content is by far the best predictor of overall site experience. Interactivity comes next.

- Visual Design does not have as much predictive power except in specific categories

  Can we develop design guidelines by identifying characteristics of good web pages?

Slide 2 — Study 2

Study 2

- An empirical bottom-up approach to developing design guidelines

  –Challenge: How to go use Webby criteria to inform web page design?

  –Answer: Identify quantitative measures that characterize pages

Slide 3

Quantitative Measures
- Page Composition
  - words, links, images, …
- Page Formatting
  - fonts, lists, colors, …
- Overall Characteristics
  - information & layout quality

Slide 4 — Quantitative page measures

Quantitative page measures

- Word Count
- Body Text %
- Emphasized Body Text %
- Text Cluster Count
- Link Count
- Page Size
- Graphic %
- Color Count
- Font Count

Slide 5 — Quantitative Measures: Word Count

Quantitative Measures:
Word Count

Slide 6 — Study Design

Study Design

Quantitative Page Metrics	Webby Ratings	Model Accuracy Across Categories	Within Categories
•Word Count, Body Text %, Text Cluster Count, Link Count etc.	Highly Rated Sites Top 33%	67%	76%
	Low Rated Sites Bottom 33%	63%	83%

6

Classification Accuracy

- Comparing Top vs. bottom
- Accuracy higher for within categories

			Classification Accuracy	
		N	Top	Bottom
	Overall	1286	67%	63%
Within Categories	Community	305	83%	92%
	Education	368	76%	73%
	Finance	142	77%	93%
	Health	165	93%	87%
	Living	106	42%	76%
	Services	208	86%	75%
	Cat. Avg.		76%	83%

What page metrics predict site quality

- **All metrics played a role**
  - However their role differed for various categories of pages (small, medium & large)
- **Summary**
  - Across all pages in the sample
    - Good pages had significantly smaller graphics percentage
    - Good pages had less emphasized body text
    - Good pages had more colors (on text)

Role of Metrics for Medium Pages
(230 words on average)

- **Good medium pages**
  - Emphasize less of the body text
  - Appear to organize text into clusters (e.g., lists and shaded table areas)
  - Use colors to distinguish headings from body text
- **Suggests that these pages**
  - Are easier to scan

Low Rated Page

No Text Clustering

No Selective Highlighting

High Rated Page

Selective Highlighting

Text Clustering

Why does this approach work?

- Superficial page metrics reflect deeper aspects of information architecture, interactivity etc.

7

Future work

- Do similar analysis for Webby2001
- Add more metrics
  - Aspects of info, navigation, and graphic design
- Category-based profiles
  - Use clustering to identify different kinds of good and poor sites
  - These can be used to suggest alternative designs
- Do user studies to verify guidelines

Content or Graphics

7th Conference on
Human Factors and the Web

In Summary

- Developing an understanding of characteristics of good pages and also generating empirically validated design guidelines

- Laying the foundation for a new methodology
  - Empirical, bottom up

- Can predict if a page is good or not with some accuracy

Content or Graphics

7th Conference on
Human Factors and the Web

More information:

http://sims.berkeley.edu/~sinha/WebbyAwards

Content or Graphics

7th Conference on
Human Factors and the Web