Looking for Genres on the World Wide Web: Content Analysis of American Author Web Sites

Timothy Hackman and Karen Patterson

University of Maryland Libraries

College Park, MD

Abstract

This project entails content analysis of over 200 websites for authors commonly taught in university-level American literature survey courses. Using a faceted classification scheme to categorize site content, the researchers explore the extent to which consistency has developed in the scope and content of Web sites for authors in the canon of American literature. The goal is to identify both the full range of unique content on these Web sites and; within that range, the core content elements that comprise the "American Author Web Site" genre. The authors also gathered biographical details on each author, to consider the question of whether Web site content varies by an author's gender or racial background. Ultimately, the goal is to identify a set of common and/or "best" practices for content and organization of author Web sites that will assist both undergraduate students and advanced researchers looking for author information on the Web, as well as site designers hoping to create quality resources for these users. In the choice of subject, size of the sample, and novelty of the approach, this promises to be a unique project that will interest librarians, web developers, and literature researchers alike.

Research Questions

By focusing on a particular genre of Web sites, this research seeks to answer the following questions:

- What is the range of content on free Web sites for American authors?
- What is the core content? That is, to what extent has consistency developed in the scope and content of Web sites for authors in the canon of American literature?
- What sub-genres exist in the overall genre of the American author Web site?
- Does content vary by:
  a) Relative status of each author within the canon?
  b) Gender of the author?
  c) Ethnic or racial background of the author?
- What are the common and "best" practices for content and organization for author Web sites?
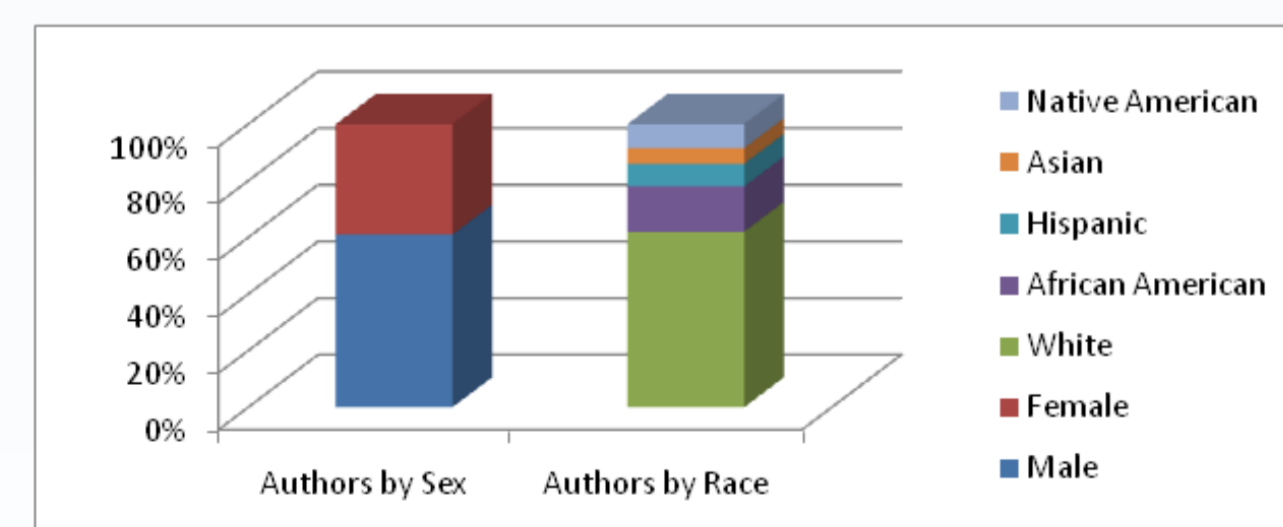

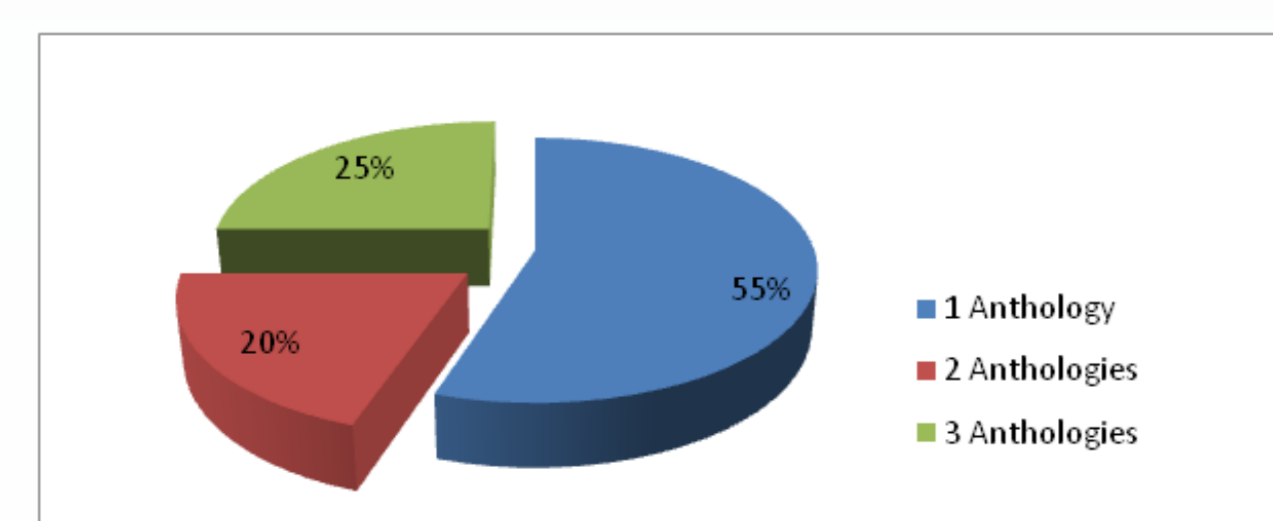Figure 1: Population and Sample Breakdown by Author Sex and Race

Figure 2: Population and Sample Breakdown by Number of Anthologies

Methods

1. Sample Selection: Researchers identified authors to be studied by listing all entries from three major anthologies of American literature. Data was collected on author sex, race, and the number of anthologies in which he or she appeared (one, two, or three). From the 447 unique authors included in the three anthologies, we selected a sample of 200 (44.5%), which was representative in terms of sex, race, and number of anthologies (see Figures 1 and 2).

2. Website Selection: One Web site was selected to be coded for each author. The primary criterion for selection was that the site be judged credible by the researchers. For this purpose we established a hierarchy of preferred sites. The first preference was for the author's official society (e.g., the James Fenimore Cooper Society) Web site, if one existed. Next were sites identified by the editors of the Chadwyck-Healey database Literature Online (LION), which includes Web resources in its "Criticism and Reference" section, or by the editors of the Librarians' Index to the Internet (LII). After LION and LII came sites selected by the editors of the Yahoo! or Google Web directories. Last of all came sites identified through open Web searches using Google or Yahoo! search engines. In addition, the researchers preferred Web sites that were devoted to a single author and included a wide variety of information (i.e., not just a biographical sketch or only a collection of author photos.) We were able to identify Web sites for all 200 of the authors in our sample.

3. Website Coding: Researchers visited all pages within each author Web site and coded every unique content element found therein. We began with a short list of content elements based on coding of ten test sites; from there we used a "constant comparison" method that allowed us to add new content elements as they were identified. We used a faceted classification scheme to describe Site content, with separate facets for:

- *Subject* (e.g., Author, Work (Fiction, Nonfiction, or All), Author's Family or Friends, Author's Home, Organization)
- *Form or Subform* (e.g., Bibliography, Biography, Criticism, Interview, Lesson Plan, Picture, Plot Summary, Quotation)
- *Completeness* (e.g., Comprehensive or Selected)
- *Arrangement* (e.g., Alphabetical by Author, Alphabetical by Title, By Genre, By Language, Chronological)
- *Audience* (e.g., Laypersons, Scholars, Children)
- *Full-Text Availability* (e.g., Excerpt, Full-Text, Link to Full-Text, Link to Media)

Taken together, a string of facets is capable of fully describing a single Web site content element. Figure 3 shows a typical Web site with content elements identified by their initial facets, Subject and Form/Subform.
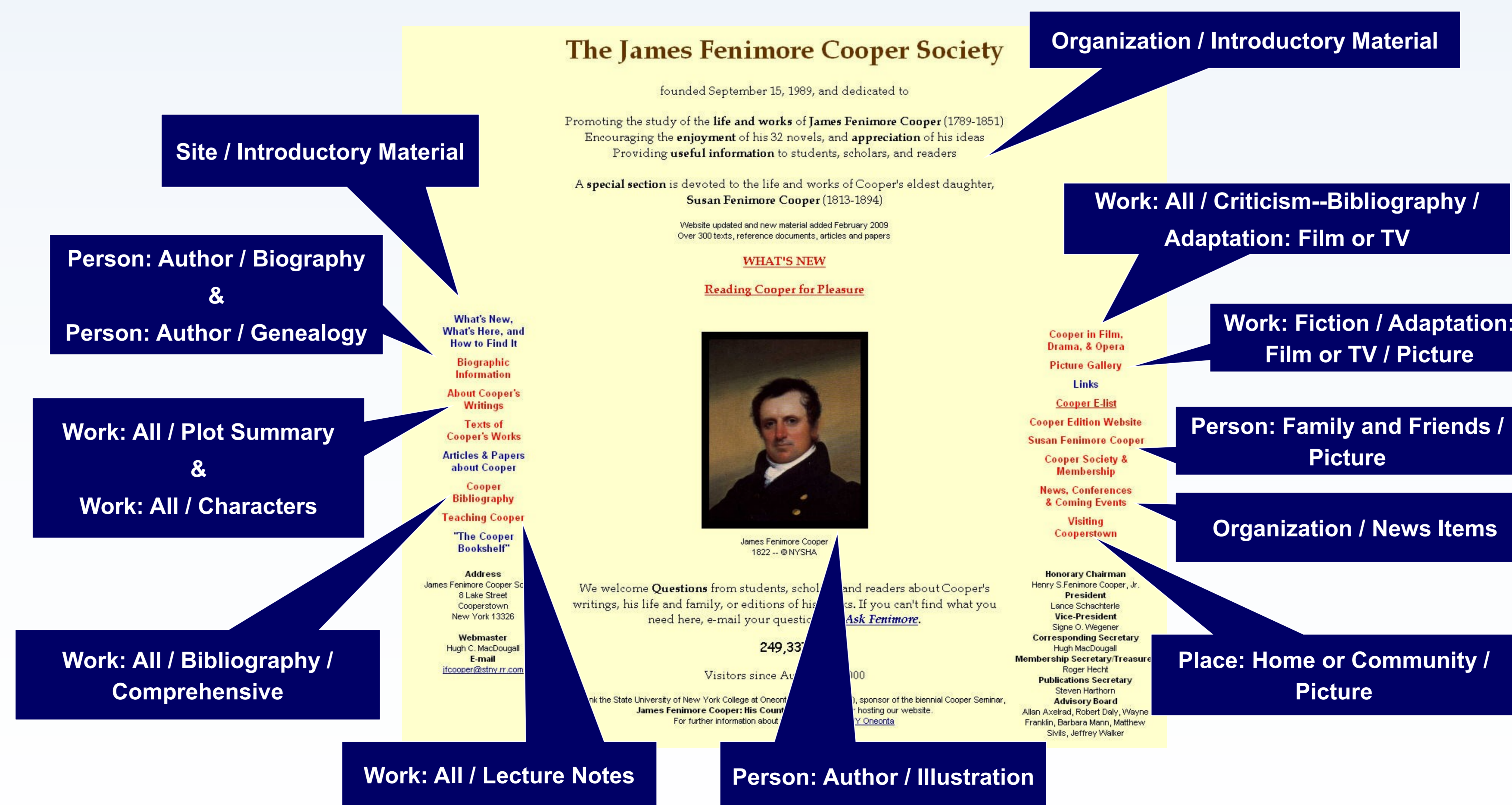

Figure 3: Sample Coding for the James Fenimore Cooper Society Web Site, http://external.oneonta.edu/cooper/

Results

In all, the researchers coded a total of 1,611 content elements for the 200 author sites, and identified 247 unique content elements. The average number of content elements per Web site for all authors was 8.06 (median = 6). There were some differences in this number by sex, race, and the number of anthologies in which the author appeared, as shown in Figures 4, 5 and 6. The most notable finding was the extraordinarily high average of 12.7 features per site for authors who appeared in all three anthologies.

Figure 4: Average Number of Content Elements per Site by Author Sex

Figure 5: Average Number of Content Elements per Site by Author Race

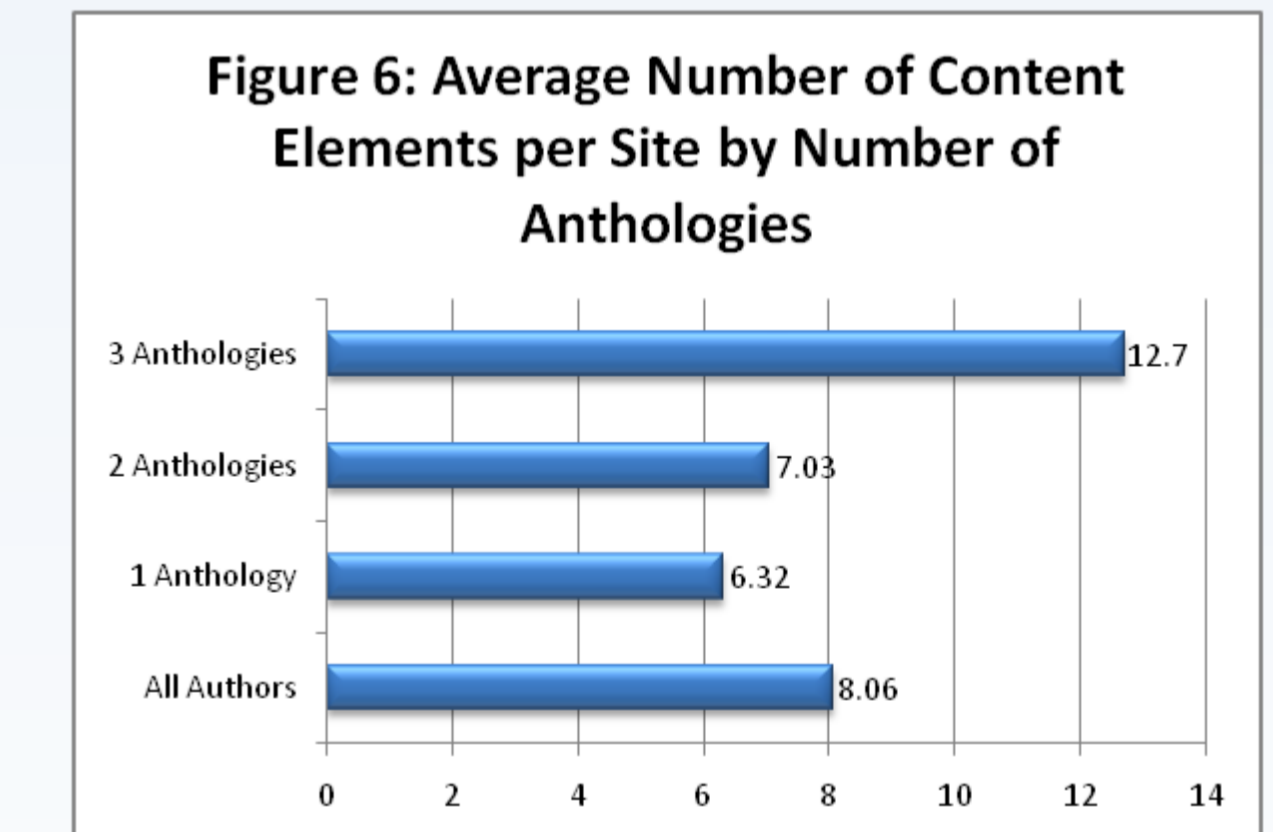Figure 6: Average Number of Content Elements per Site by Number of Anthologies
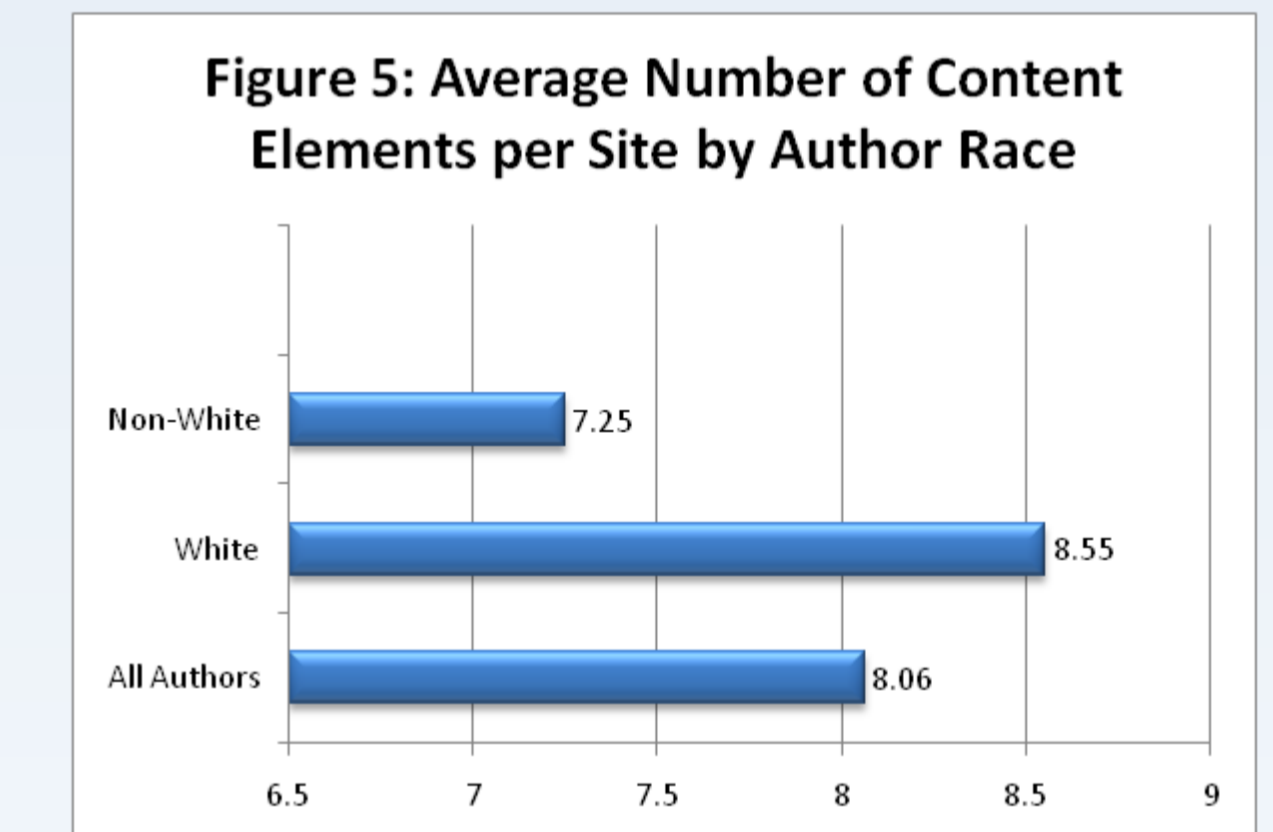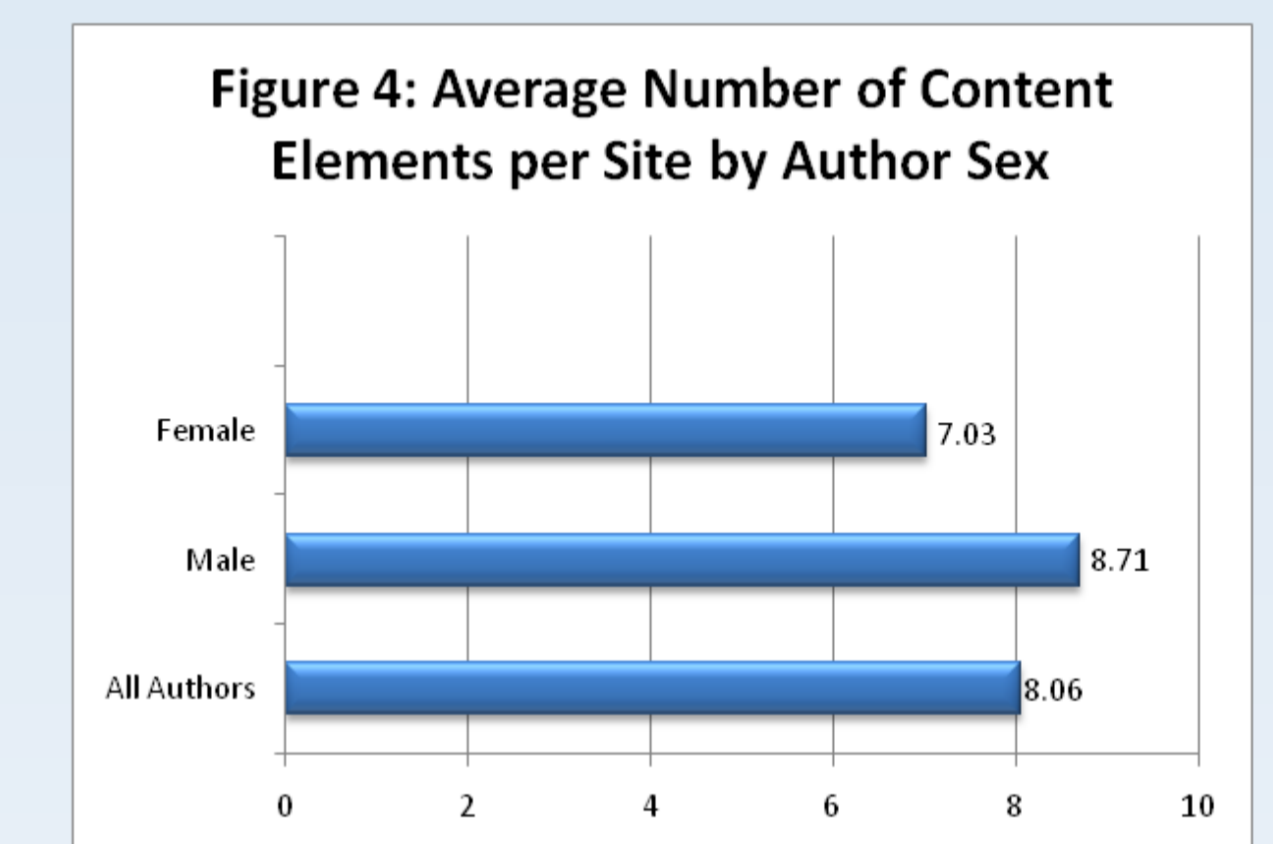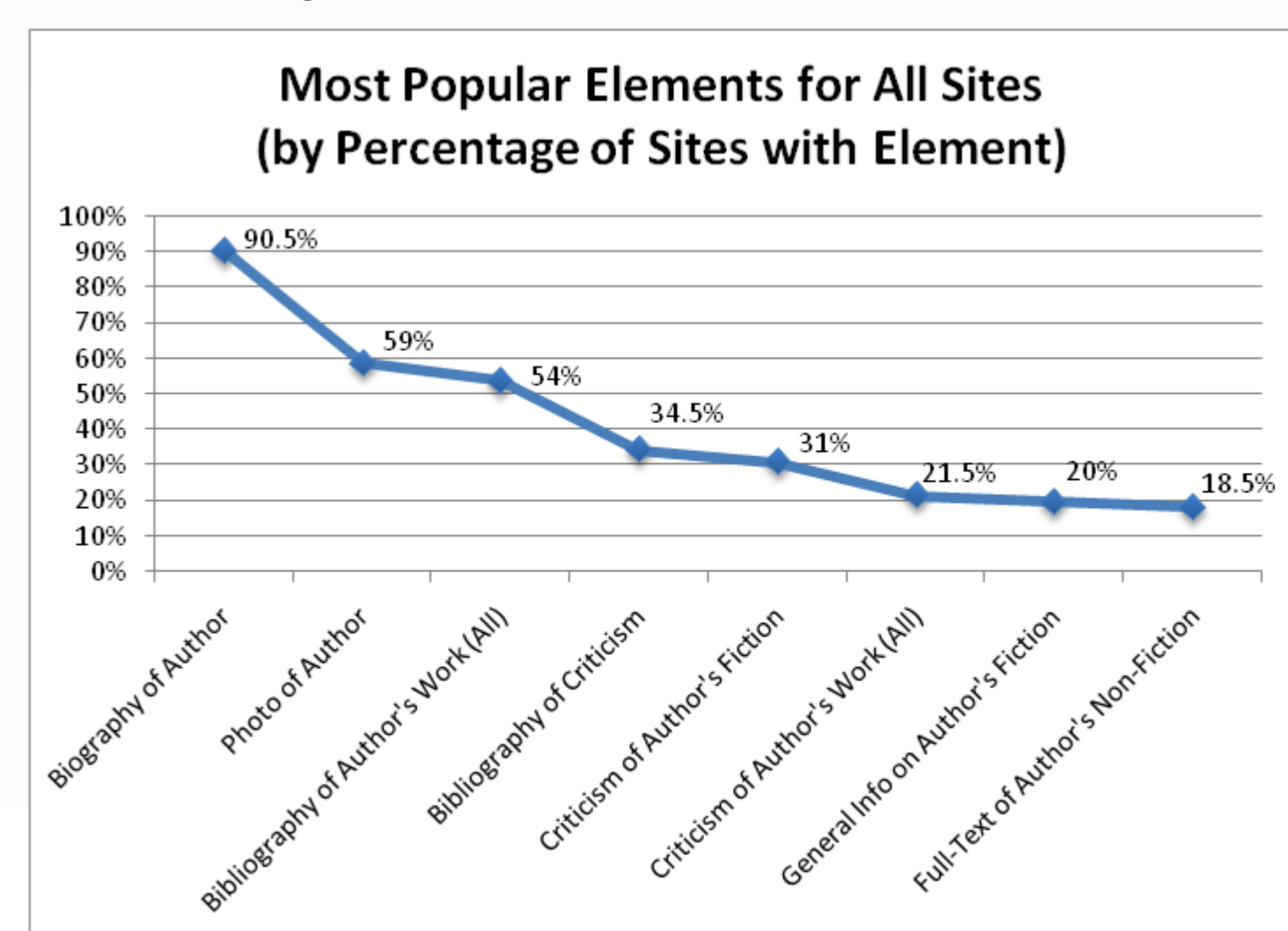
Figure 7 shows the eight most common content elements for all sites by the percentage of sites that included each element. A biography of the author was by far the most common (90.5% of sites had this feature), followed by a photo of the author and a bibliography of the author's work (59% and 54%, respectively). Beyond these top three, however, the percentage of sites with each element drops off quickly—the fourth and fifth most popular elements only appeared in about one-third of coded sites—indicating that there is greater diversity than expected in the content of these free author Web sites.

Most Popular Elements for All Sites (by Percentage of Sites with Element)