

Estimating Ethos: Testing Heuristics for Gauging Authority

Paper Proposal for Computers and Writing 2008

Michael Wojcik, Michigan State University

10 January 2008

In this presentation I discuss an ongoing project to identify measurable proxies for ethos and implement software for analyzing corpora of digital text and determining which sources in those texts are likely to be high-ethos contributors, within the corpus at large or in relation to specific subjects.

One of the oldest and most prominent uses of the Internet has been to solicit information from the participants in various forums—bulletin-board sites, newsgroups, email lists, and so on. These are brief, casual, and usually unedited venues, so respondents often must establish the authority of their replies in part through Aristotelian ethos (“I know what I’m talking about”). In other words, online texts often incorporate explicit or implicit claims about the correctness of their claims based on the identity of the author (or the author’s sources), rather than on logical (logos) or emotional (pathos) arguments. Further, as these forums become established and acquire a history, frequent contributors often acquire reputations as generally-correct gurus, controversial mavericks, obstructionist trolls, etc; and these roles also affect their standing to make ethic arguments supporting their claims. Scholars such as Jim Porter and Fil Sapienza have explored many nuances of ethos in digital writing, demonstrating the significance of this area of inquiry.

Ideally, there should be some correlation between ethos and actual knowledge—we hope apparent experts really are experts. Certainly users appear to use ethos, among other factors, in deciding which sources are reliable. A source’s ethic standing may indicate competence, and probably correlates to persuasiveness, so heuristics for estimating ethos could be useful both for identifying high-value sources and for predicting user preference when choosing sources. Google’s PageRank algorithm, for example, uses document linking—in effect an indicator of popularity, an element of ethos—as one indicator of relevance in ranking search results. (In fact, PageRank weights links by its estimation of the “importance” of the linking page, which is itself an ethic factor.)¹ With the Estimating Ethos project, our hope is to implement heuristics that provide useful and interesting guesses about the ethic standing and authority of various contributors to a body of texts.

I will discuss some of the interesting theoretical aspects of the Estimating Ethos project, including my definition of ethos and proxies for it. I’ll explain key design elements and implementation choices for the software—a generic, open-source framework for analyzing documents using plug-in modules and using the results, suitable both for experimentation and use in actual writing environments. After a brief demonstration of the existing analysis modules and their output, I’ll conclude with some comments about possible applications, which I see as assisting (not replacing) sophisticated readers and writers.

1 See Google’s Technology Overview, <<http://www.google.com/intl/en/corporate/tech.html>>.