

# Estimating ἦθος

Michael Wojcik  
Rhetoric & Writing  
Michigan State University

1

Thanks, everybody, for coming here on a Saturday morning. Thanks also to D. Alexis Hart for moderating this panel and the other presenters for sharing it with me; Computers and Writing for providing this opportunity; and the University of Georgia for hosting it.

I'd also like to thank the faculty of the Rhetoric and Writing program at Michigan State University, particularly Bill Hart-Davidson, and my fellow graduate students, who provide—really—an amazing intellectual community.

And finally, my deepest gratitude to Malea Powell for her inspirational example and unfailing support.

With that out of the way: I'm going to show you a bit of a project, really just starting to get off the ground, called **Estimating Ethos**. It's an exploration, and its only firm goal is producing interesting results—though it does look like we may have something useful here as well. So I'll talk a bit about its origins and the theory behind it, and then show what's been implemented and what we can do with it.

An algorithm (or two, or three) that **facilitates formative assessment re: ethos as a complex function of interactions with others** in a social networking (broadly speaking) system. Of course, making it about ethos means that it requires some interesting approaches to identity/subjectivity to avoid simplistic reification. No rational subjects (tm) need apply!

2

I started this project last fall as an independent study suggested and supervised by Bill Hart-Davidson, in conjunction with work being done at the Writing in Digital Environments research center (or “WIDE”). Bill proposed a number of projects, and this was his original definition of what became Estimating Ethos. The basic question was: is there some algorithmic way to make judgments about writing (specifically in online social networking environments) that help us assess ethos?

if successful, this would become part of the

## Social Writing Application Platform (SWAP)

a writing platform being developed at WIDE

3

The eventual intent is to incorporate the results of the EE (Estimating Ethos) project into SWAP, which is a suite of online writing services that can be combined in various ways to produce social writing applications. SWAP services include things like content management, folksonomic tagging, social networking tools, and so on.

WIDE researchers are also working on incorporating analysis tools into SWAP. These would let groups of users derive interesting data about their own writing products and practices, like what passages readers highlight or copy for quotation, or who frequently edits shared documents.

EE would be one of these analytic tools, and it would be used to suggest answers to questions like “in this group, whose opinions seem to be considered valuable on topic X?”.

heuristic

contextual

probabilistic

“fuzzy”

proxies

4

Before I go further, let me qualify this. We knew from the start that we weren't going to identify some kind of meaningful computational measure of ethos. It's widely known that human judges can't agree on even such apparently simple matters as parsing natural languages, so obviously there isn't going to be some broadly acceptable metric for something as elusive and abstract as a whole category of persuasion.

What the EE project is looking to do is to propose and test heuristic algorithms that yield data that is constrained to a particular context, and that expresses some kind of fuzzy probability about ethos. That is, we'd like to get results like “in this conversation, with this degree of confidence we think this source is probably in a strong position to employ ethos”.

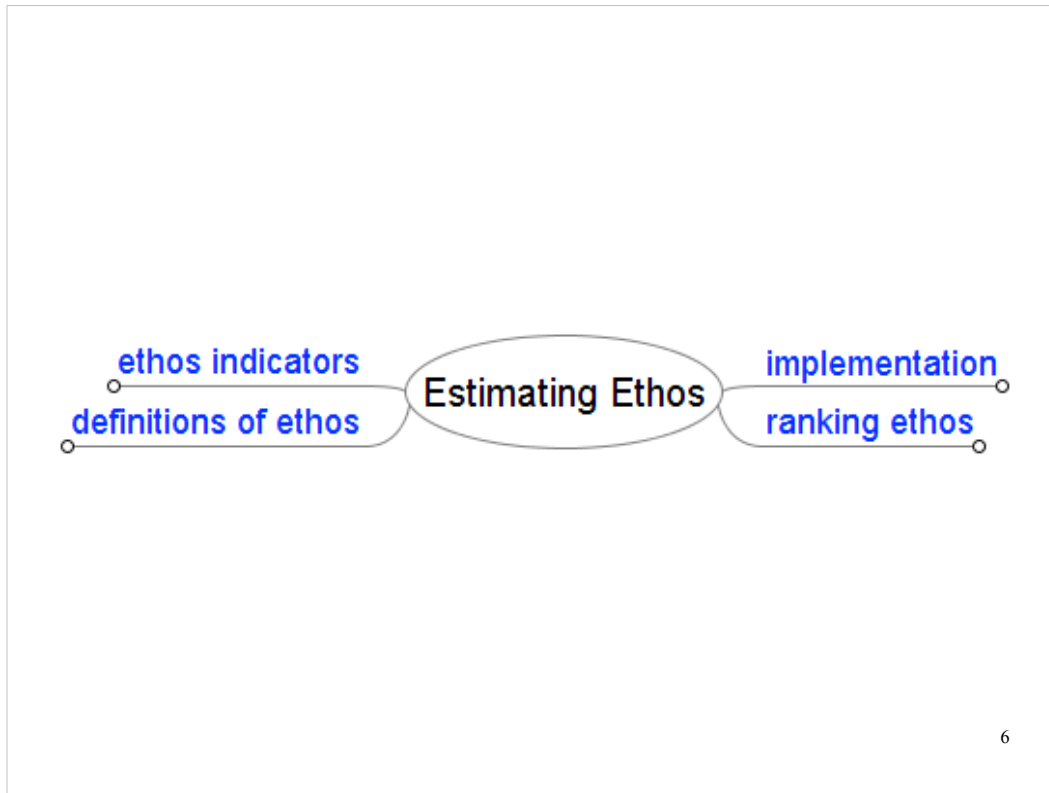
And since I don't know what it would mean to measure ethos directly, EE actually looks at measurable properties that I believe can serve as proxies for ethos, or even proxies for analogues of ethos—and I'll talk about one of those in a minute.

# “ethetic standing”

5

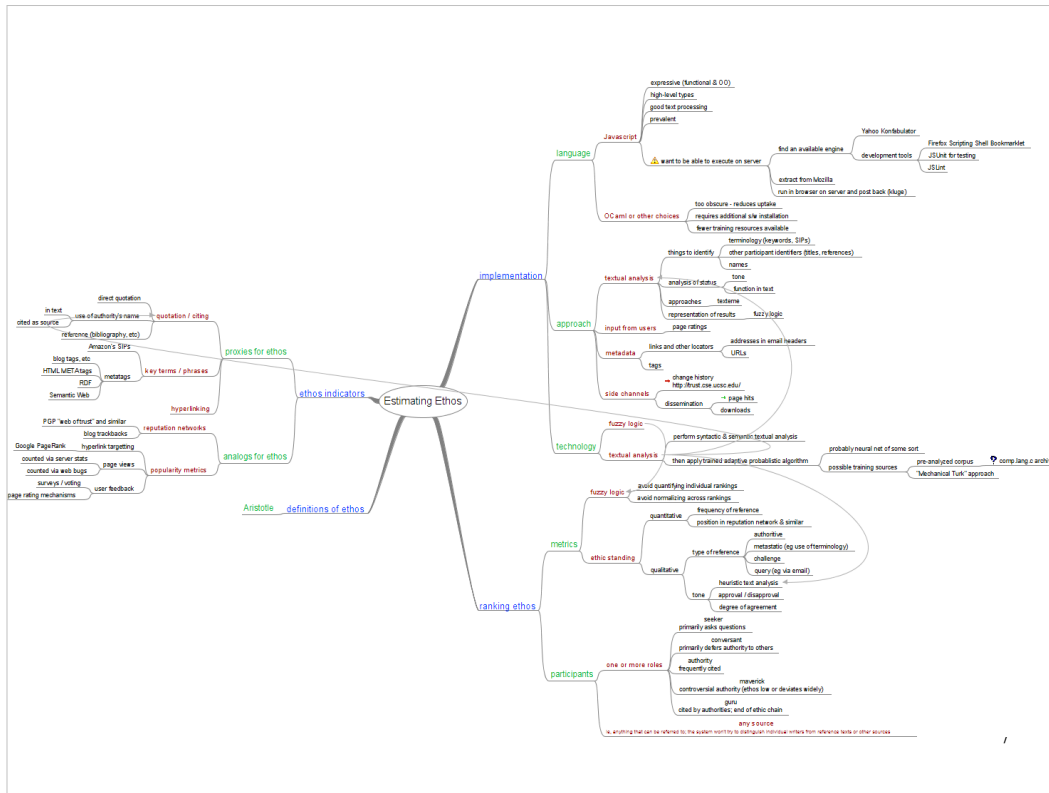
There's one other point I should clarify here. The EE project actually took a bit of a detour early in its development.

What EE is currently looking to measure isn't ethos itself, but a precursor to it: what I'm calling “ethetic standing”. Ethetic standing is the leverage a source has for mounting ethos arguments. It's the force a rhetor can apply through ethos, through appeals to the audience's willingness to believe, in a given persuasive situation.



So I started exploring this problem by gathering notes in four basic areas. This is a mind map I started from my initial notes. The four areas are:

- how to define ethos for the project
- what in the data would indicate ethos (proxies, and evidence of them)
- how to rank those indicators, to determine the relative ethos of sources
- if any algorithmic approaches emerged, how would they be implemented?



It quickly became clear that this could go all sorts of places...

...which is why this will always be a work in progress, and the system I've designed accommodates that by being modular and extensible. We'll see that in a moment.

an appeal to the audience's  
inclination to believe the source

8

For my working definition of ethos I went back to Aristotle. Over the centuries “ethos” has been used to mean various things—often today it's used for something like a combination of tone, stance, and motive—but for this project I'm using the classical one: ethos is an attempt to persuade by appealing to the audience's inclination to believe the writer (or whatever the source of the claim).



an appeal to the audience's  
inclination to believe the source

the “good source”  
(*arete*)

the “informed source”  
(*phronesis*)

the “benevolent source”  
(*eunoia*)

9

Ethos might take the form of a claim about the author's goodness, or in general about the audience's good impression of the author. Or it might appeal to the audience's belief that the author is a subject-matter expert, or to their inclination to trust a perceived authority figure.

“You know I know what I'm talking about”

10

In any case, what ethos boils down to is an implicit or explicit claim:  
trust me, because you know I know what I'm talking about.



11

What are some possible proxies for ethetic standing? That is, what are some clues an algorithm might be able to identify that a source is likely to be believed—just based on who or what that source is?

Well, quotation and citation could be a proxy: if you're quoted a lot, there's a decent chance that it's because you're considered an expert. Nomenclature is another; lots of scholars using the term “deconstruction” suggests that Derrida has ethetic standing in their community. Or hyperlinking, which is the key attribute measured by Google's PageRank algorithm, for example. Obviously, those are all members of the same family, what we might call “memetic attention markers”.

Analogues for ethetic standing could be popularity metrics (page views, user feedback), or more broadly reputation networks like PGP's “web of trust” or blog trackbacks.

*Reputation* is one way of gaining ethetic standing.

12

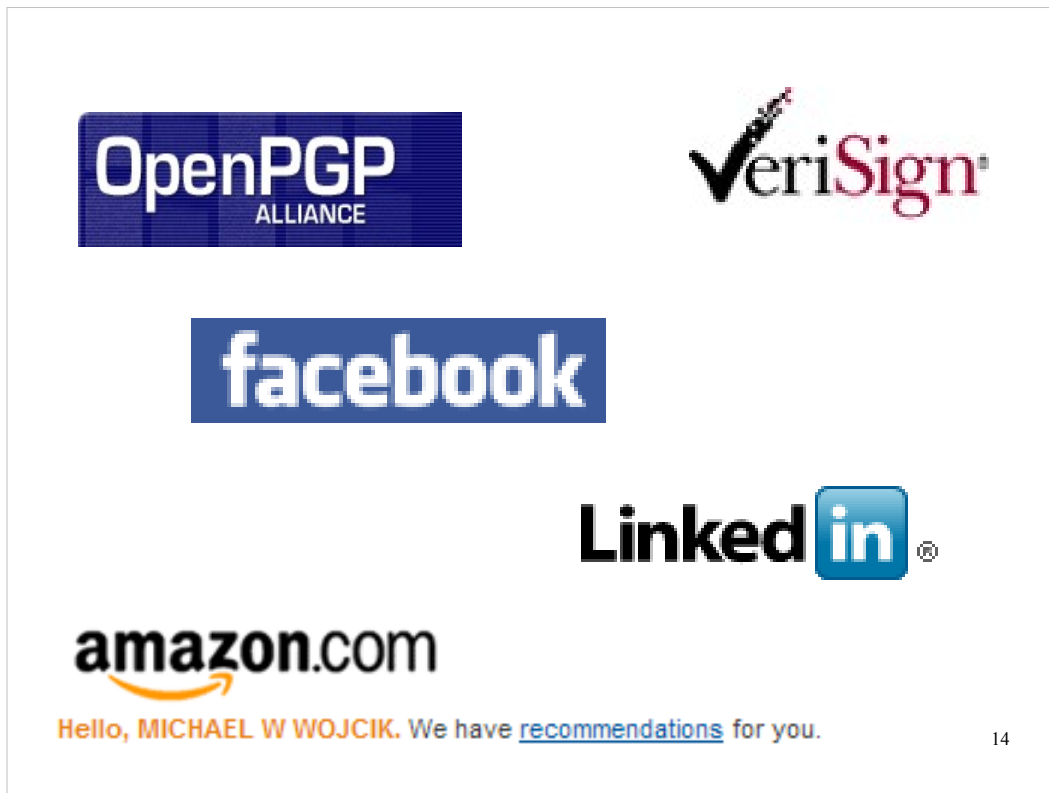
It should be pretty obvious that reputation is one way of gaining ethetic standing.

The relations of ethetic standing in a discourse community are an informal reputation network.

13

In fact, I'll go so far as to say that the relations of ethetic standing in a discourse community are an informal reputation network, and we can model ethetic standing as a reputation network.

Which is good, because there's a lot of reputation network theory to draw on.

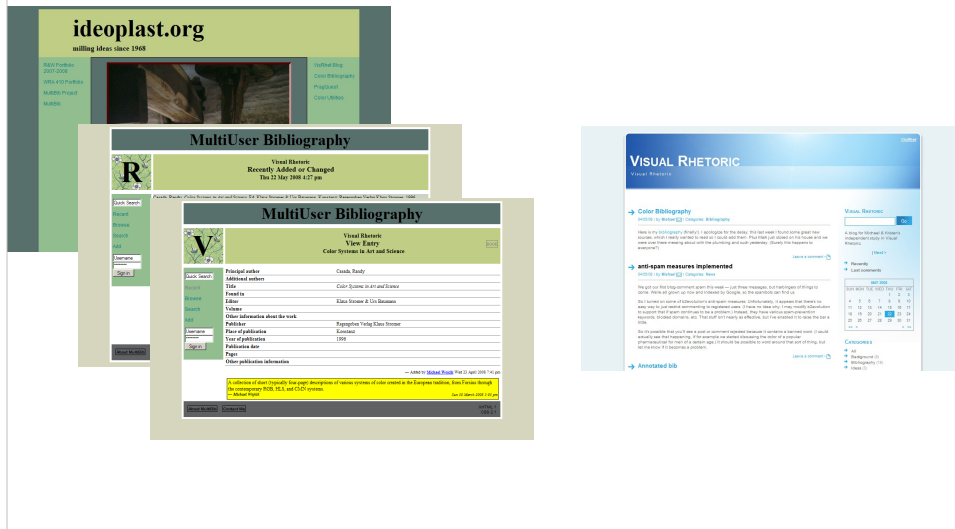


There's a lot of research on reputation networks because reputation networks are everywhere. There are decentralized networks like PGP's web of trust, where anyone can indicate they trust anyone else. There are hierarchical networks like the X.509 certificate tree used for SSL in e-commerce, which is rooted at VeriSign: we all trust whomever VeriSign trusts.

There are social networks.

There are even fully-automated reputation networks, like recommendation systems.

# Google Page Rank



15

Google PageRank is a particularly interesting example. You know how PageRank works, right? Say you have two pages that have similar content. Initially they're ranked the same.

Now say another page links to one of them...

[click]

...that increases its rank. It's an endorsement; it increases the target page's reputation.

And any pages that are linked to by the target page...

[click]

...also get a higher rank, because an endorsement from a higher-ranked page counts for more.

PageRank uses linking as a proxy to discern an implicit reputation network, which it treats as an analogue for ethos.

16

PageRank says, you should look at *this* page because a lot of other pages say you should look at that page; and those pages are trustworthy because a lot of pages say you should look at them. And so on.

So I think you can reasonably argue that PageRank uses linking as a proxy to discern an implicit reputation network, which it treats as an analogue for ethos.

And that's my starting point too, except that I'm trying to extend the idea beyond web pages and links.



Subject: Re: Lack of end tag messes up form styling - a known bug?  
From: Michael Wojcik <mwojcik@newsguy.com>  
Date: Fri, 11 Apr 2008 16:45:42 -0400  
Newsgroups: comp.infosystems.www.authoring.stylesheets  
References: <5gmLj.324043\$eT7.60746@reader1.news.suuna.fi>  
In-Reply-To: <uWwLj.324342\$mS3.196779@reader1.news.suuna.fi>

Jukka K. Korpela wrote:

>  
> (P was never officially a paragraph break. The first HTML  
> specification was HTML 2.0, and it clearly defines P as an element  
> indicating a paragraph. I don't think "VK" even ever read an HTML  
> specification.)

P was a paragraph break in HTML 1.0, but I don't believe that ever  
moved past IETF internet-draft status,  
so I'd agree that doesn't count as "official".

From the 1993 HTML 1.0 internet-draft [1]:

```
<!ELEMENT P      - O EMPTY -- separates paragraphs -->
```

The "EMPTY" indicates that the P element has no content. Also, the body  
of the i-d states: "The empty P element indicates a paragraph break".

[1] <http://www.w3.org/MarkUp/draft-ietf-iiir-html-01.txt>

17

So in trying to estimate ethos, we really want to look at all the available clues, both direct textual data and the metadata that accompanies it in the communications channel. This is a Usenet post, and you can see...

[click]

... how it includes both data and metadata.

[click]

In this case pretty neatly divided up into header and body.

[click]

In the header we have things like the source of this message and references that let us situate it in a conversation. In the body we have data like quotations and citations in various forms, and metadata like these quotation markers and footnote number.

All of that is grist for the mill, but it requires different analytic approaches. Metadata is usually pretty easy to process because it's usually machine-readable. Analysis of text data usually means natural language processing, which is difficult. So we want a technology that lets us use multiple approaches.

# email

Subject: Re: Lack of end tag messes up form styling - a known bug?  
From: Michael Wojcik <mwoj@news.guy.com>  
Date: Fri, 11 Apr 2008 16:45:42 -0400  
Newsgroups: comp.infosystems.www.authoring.stylesheets  
References: <5gmLj.324043\$T7.60746@reader1.news.saunalahti.fi>  
In-Reply-To: <uWwLj.324342\$T7.60746@reader1.news.saunalahti.fi>

Jukka K. Korpela wrote:  
>  
> (P was \_never\_ officiall  
> specification was HTML 2  
> indicating a paragraph.  
> specification.)  
  
P was a paragraph break in  
moved past IETF internet-d  
so I'd agree that doesn't  
  
From the 1993 HTML 1.0 int  
  
<!ELEMENT P - 0 EM  
  
The "EMPTY" indicates that  
of the i-d states: "The em  
  
[1] <http://www.w3.org/Mark>

Jukka K. Korpela wrote:  
>  
> (P was \_never\_ officiall  
> specification was HTML 2  
> indicating a paragraph.  
> specification.)  
  
P was a paragraph break in HTML 1.0, but I don't believe that ever  
moved past IETF internet-draft status,  
so I'd agree that doesn't count as "official".  
  
From the 1993 HTML 1.0 internet-draft [1]:  
  
<!ELEMENT P - 0 EMPTY -- separates paragraphs -->  
  
The "EMPTY" indicates that the P element has no content. Also, the body  
of the i-d states: "The empty P element indicates a paragraph break".  
  
[1] <http://www.w3.org/Markup/draft-ietf-iiir-html-01.txt>

plain text

That also means that we really want to look at each text in different forms. An email metadata analyzer wants to see email, but a text data analyzer just wants...

[click]

...the plain text extracted from the body.

So the technology should support multiple transformations of each textual object.

## fuzzy results from fuzzy inputs

19

And the third important technological requirement I want to mention, returning to the idea of fuzziness, is that we don't want to require analyzers to make hard evaluations when they probably can't.

A metadata analyzer may be able to say “I'm sure this text came from source X”, or “this text seems to be a response to text Y”. But even there it's starting to get a bit fuzzy—it's not completely confident of those determinations. And it's worse for textual analysis: “this section appears to be quoting source Z”.

certainty

guess

suspect

assume

sure

suggests

indicates

determines

strength

“I **suspect** this **indicates** this subject has ethetic standing”

20

Fuzziness means letting analyzers return results that represent an approximate degree of confidence and strength. An analyzer should be able to say, for example, that it's found something that might indicate ethetic standing.

This also means we can combine results from different sorts of analyzers without having to normalize them.

Textmill



- a general data-flow system for text processing
- uses plug-in modules
- knows how to connect things
- provides services

EE

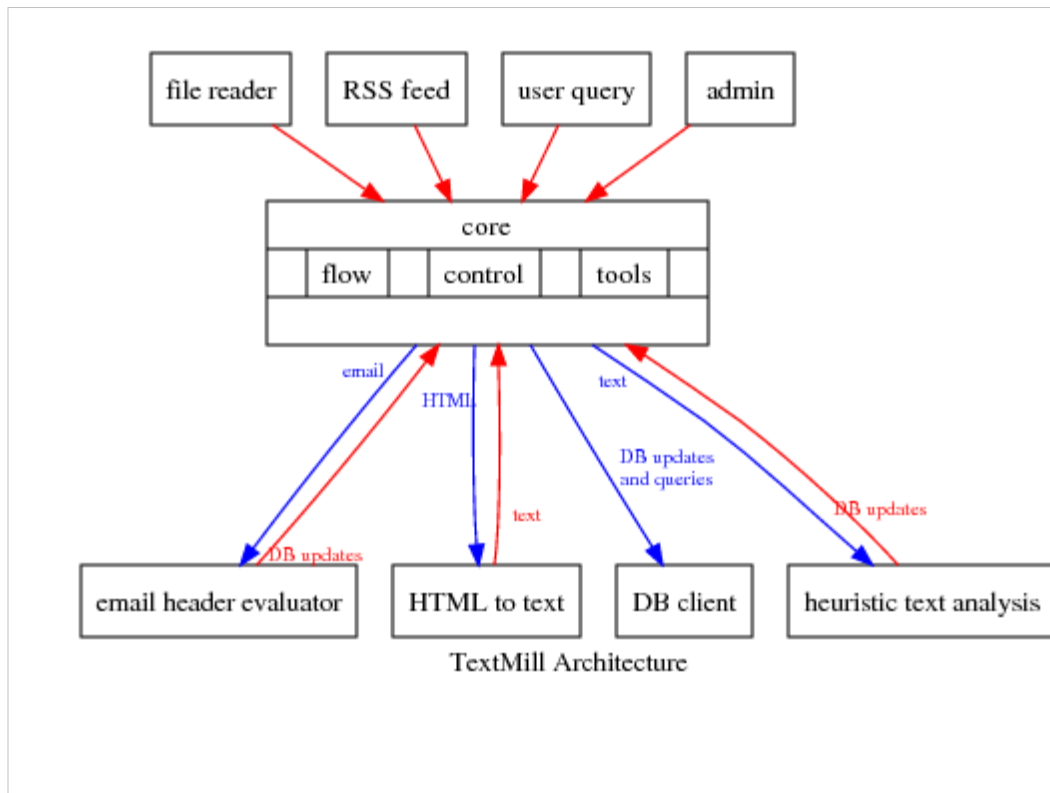


- the Ethos Estimator
- a Textmill application
- modules for loading and transforming texts
- analysis modules

21

To satisfy all these requirements in a general way, I'm building a text-processing system I call "Textmill". Textmill makes data flow through an arbitrary collection of plug-in modules. It connects modules based on what they use for input and output, and it provides them with services like storing and retrieving data and handling fuzzy logic calculations.

The actual ethos estimator is a Textmill application called "EE". It's a set of Textmill modules that load and transform texts, and another set that perform analysis for various ethos proxies.



A quick diagram of the Textmill architecture—just to show that there are modules feeding stuff in, and other modules transforming it or analyzing it. The basic idea is that Textmill pushes the data through modules as it gets converted into an appropriate form, so your text-citation-analyzer doesn't need to understand HTML, for example.

## why Javascript?

expressive (functional, OO, flexible)

well-known

widely available,  
good for experimentation

run in a browser, as Yahoo Widget, on a server...

23

The whole thing's written in Javascript. There are some good reasons for that, but I'm going to skip over them now so I can show a quick demonstration.

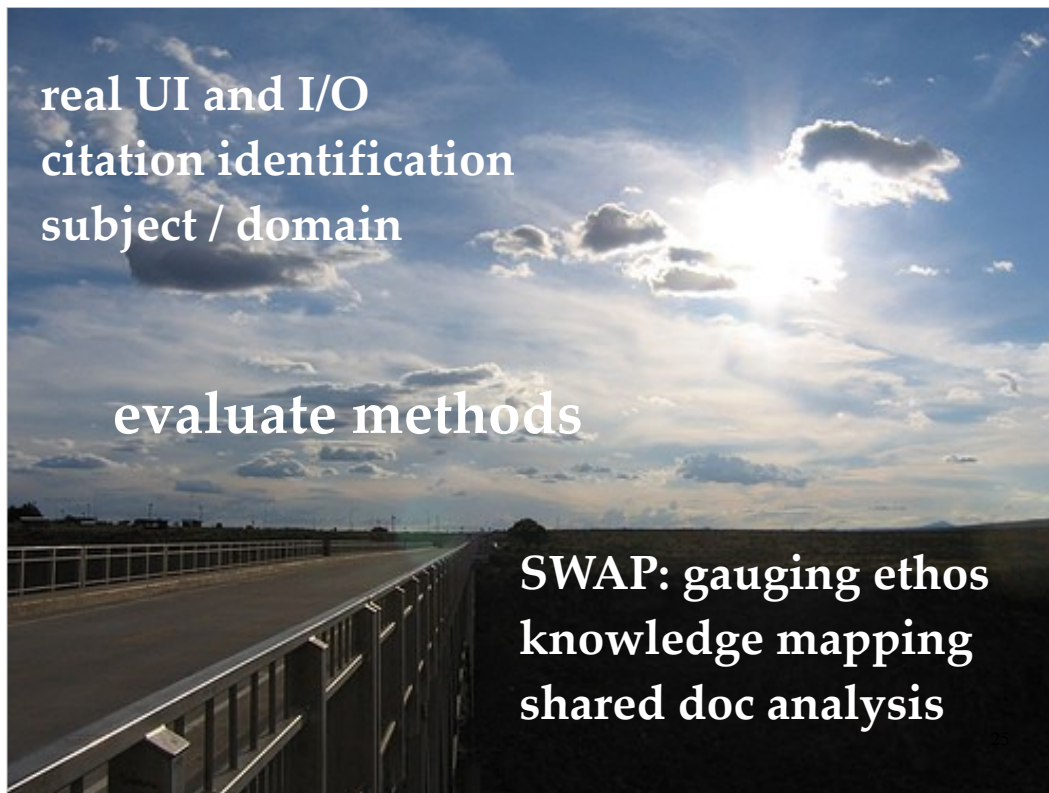
Bear in mind that this is still a prototype.

The demonstration I performed at this point in the original presentation is not yet publicly available. I hope to have an automated or interactive demonstration, or a link to one, on my site:

[Ideoplast.org](http://Ideoplast.org)

in the near future.





So what's ahead?

Work on the prototype:

- it needs a real UI and real I/O capabilities
- modules that analyze text for citations and quotations
- subject / domain identification

Theoretical application: evaluating methods for estimating ethos or ethetic standing.

Practical applications:

- for SWAP: gauging ethos to see how persuasive efficiency changes over time
- knowledge mapping: find your local experts
- shared document analysis: whose changes are durable?