ALPACA: A Lightweight Platform for Analyzing Claim Acceptability

Jeff King
College of Computing
Georgia Institute of Technology
Atlanta, GA, USA
peff@cc.gatech.edu

Jennifer Stoll
College of Computing
Georgia Institute of Technology
Atlanta, GA, USA
jstoll@gatech.edu

Michael T. Hunter
College of Computing
Georgia Institute of Technology
Atlanta, GA, USA
mhunter@cc.gatech.edu

Mustaque Ahamad
College of Computing
Georgia Institute of Technology
Atlanta, GA, USA
mustaq@cc.gatech.edu

ABSTRACT

Internet users face challenges in evaluating the validity of online information. Such evaluation is not adequately supported by current tools; we outline some of the shortcomings of these tools, including centralization, lack of automation, and lack of user-centrism. We propose a set of design principles to mitigate these shortcomings and introduce ALPACA, A Lightweight Platform for Analyzing Claim Acceptability, which adheres to these design principles. ALPACA provides a graphical means of organizing the user's trust with regard to information claims and sources, as well as tools for examining the trust assumptions of others.

Categories and Subject Descriptors

H.4.3 [**Communications Applications**]: Information browsers; H.1.2 [**User/Machine Systems**]: Human factors

General Terms

Design, Human Factors

1. INTRODUCTION AND MOTIVATION

Consider the case of trying to evaluate information on candidates in the current U.S. presidential election. In reading the news, voters may ask questions such as "Is Barack Obama a radical Muslim?"; or "does John McCain want to keep troops in Iraq for 100 years?" (the former is an outright falsehood, while the latter is a drastic oversimplification of McCain's statement). The Internet provides its users with an overwhelming abundance of confusing, contentious, and (un)true claims which the user must sift through in order to find answers. The confusion caused by claims such as these is real, and it has a real impact on election outcomes. In January of 2007, Insight magazine published an article claiming that Barack Obama is a Muslim[6]. Over a year later, a March 2008 poll indicated that one in ten respondents still believes that Obama is a Muslim[14], and a September 2007 poll had forty-five percent of respondents claiming that they would be less likely to vote for a Muslim candidate[12].

Although the presence of conflicting information online is an essential component of genuine human discourse, information consumers often lack effective tools for resolving conflicting information in a timely fashion. For example, to make conflicts in information more transparent or easier to view, the user may want to see the relationships between pieces of information. Which claims directly support or refute other claims? Who made which claims? What is the reputation of the claim makers, and what is the basis for that reputation?

Unfortunately, discovering the answers to these questions is currently a manual process where the user must identify the claim, sort out the supporting points and opposing counterpoints to those claims, and ferret out meta-data such as the provenance of the claim, all while keeping track themselves of the gathered pieces of information and their relationships. The problem, therefore, is not one of lack of information, but rather one of organizing the existing deluge of information available on the Internet in a meaningful way. Though there are expert tools available, such as Starlight[10], for supporting this task (among others) on a larger scale, these tools are not available to the average information consumer.

We ask, in the face of inaccurate or misleading information, what can users do to assess the quality of information they seek to use? How can we, as researchers, help users accomplish this goal? We approach this problem from the perspective of an individual user interested in determining the credibility of claims made on the Internet. We will focus in our examples on political discourse because it is an application domain which naturally creates contention and exposes opposing viewpoints; however, our points are meant to apply generally to Internet-based discussion.

Section 2 of this paper elaborates further on the problems with the current state of information quality and challenges in using the tools available to users. Section 3 proposes some design principles that address these challenges and introduces ALPACA, the authors' attempt at a system for determining information credibility that im-

WICOW'08, October 30, 2008, Napa Valley, California, USA.

plements these design principles. We discuss the preliminary design of ALPACA and outline future research directions.

2. SHORTCOMINGS OF CURRENT APPROACHES

The Internet has given members of modern society access to contentious information at an unprecedented level, in terms of both broadcast and interpersonal communication. Instead of information being restricted to the traditional gatekeepers of newspapers, television, and radio, anyone with a web page can reach a large number of people. At the same time, the Internet has helped to connect people in a peer to peer fashion, sharing information and discussion via email, newsgroups, and other forums. Despite this wealth of information and options for connecting with others to share information, current approaches to helping users aggregate and disambiguate online information have several shortcomings. We focus on three such shortcomings: centralization, lack of automation, and lack of user-centrism.

2.1 Centralization

One way that users deal with the mass of information is to rely on aggregation services for organizing the information. A 2004 survey shows that television, radio, and newspapers all are more commonly used for receiving political claims than online sources[11]. And when online sources are used, users typically go to a few well-known sites, including "the news pages of AOL, Yahoo, or other Internet service providers,. . . the websites of the major broadcast and cable news networks,. . . [and] the website of their local paper."[11] Despite the availability of much more readily-available knowledge, users do not seem to necessarily know more or be more sophisticated consumers of information. A survey conducted as recently as 2007 indicates that little has changed with respect to voter knowledge: "On average, today's citizens are about as able to name their leaders, and are about as aware of major news events, as was the public nearly 20 years ago."[13]

Users rely on these organizations to provide a valuable service: they do the leg work of aggregating and interpreting a wide field of information to present to the audience. However, by their nature they also have drawbacks. First, centralized information sources provide an obvious target for manipulation. Because a great deal of work goes into locating, analyzing, and presenting information, there are necessarily a limited number of sources, each with a large audience. Furthermore, because the function of these sources is to simplify the flow of information to the audience, audiences are unlikely to carefully analyze the statements of these organizations or the evidence supporting their statements. Thus from a security perspective, centralized sources are a weak link in the information chain; manipulating a source (or creating a source with an agenda) is an easy way to manipulate an entire audience.

Second, centralized systems tend to accumulate polarizing bias that is unlikely to fully represent the opinion of a given voter. For example, a 2004 study looked at the overlap in "believable" news sources between different political affiliations. Of those identifying themselves as Republicans, Fox News was ranked as the "most believable", while not being in the top six for Democrats. Similarly, National Public Radio was in the Democrat top six, but not that of the Republicans[11]. This forces voters of different mindsets into predetermined niches, not all of whose opinions they necessarily share. The audience must either carefully vet any claim or analysis coming from an organization, or they must blindly accept the claim even though it may be based on assumptions that are not shared.

Some sites, such as Digg[3], try to democratize the filtering process, eschewing editors in favor of users voting on stories. However, such sites often end up acquiring their own set of biases through the self-selection of users. For example, there have been claims that Digg votes inordinately represent the support of Ron Paul for the Republican presidential nomination[8]. Though representing the views of many users, these sites show the same systematic bias as a centralized organization. Unfortunately, there is no way of escaping single source bias and achieving a more global view of the information presented on these sites; there are few if any tools for aggregating the information between multiple sites.

2.2 Lack of Automation

Some sites specialize in examining the bias of other media sites, or in examining the supporting evidence or origins of a particular claim. For example, factcheck.org[4] examines the validity of political claims by presenting a rigorously-sourced discussion of the facts supporting a claim. Another site, snopes.com[9] covers a wide range of topics from political claims to urban legends, presenting concise reasoned arguments with detailed sources. Both of these sites (and others like them) perform an invaluable service, but they are created manually, and they are consumed manually. That is, they are written by humans and for humans. The information they contain is not available in a parseable form which can be used computationally, but rather is presented in the form of natural language essays. This makes more work for writers, more work for readers, and prevents more advanced analysis of the information.

One drawback of manual discussion of claims is that the work of preparing an article must be performed by an expert human. This necessarily limits the number of articles which may be published to the most popular ones. We see analyses of claims made in the presidential race, but few for Congressional seats, and almost none for local races. An automated system which organizes the data without the help of humans, or with the help of non-expert humans (e.g., by aggregating votes) is capable of addressing a much larger variety of topics.

Another drawback is that these sources must be consumed manually. While they do succeed in reducing the total volume of information which must be consumed to make a judgement, there is little facility for presenting their conclusions at varying levels of detail. While for many claims, reading a multi-page article is necessary to comprehend all of the issues, there is an abundance of claims for which the simple case of "there is no credible evidence to support this" will suffice. As stated previously, the problem of information credibility on the Internet is one of too much information in need of organization; automated systems should be capable not just of analyzing interesting claims, but also of culling the least interesting ones.

Finally, meta-analysis of claims is difficult if the relationships between claims are not in a computer-readable form. For example, the mentioned sites assume the end goal is determining whether a claim is true. But there are many interesting questions that can be asked beyond that by looking for patterns in claims, such as "Which media sources tend to give information that is credible?" or "What biases or inconsistencies do we see coming from particular sources?" Answering these questions through manual analysis is very difficult, but should be trivial once the information is in a computer-readable state.

2.3 Lack of User-Centrism

Many systems search for an "objective truth"; they assume that there exists a global credibility value for each claim that is the same for every user, and they serve as a black box, converting some input into a credibility value. This input often takes the form of mass user participation and produces a single credibility value for each claim (or item of interest). For example, Google's PageRank algorithm[1] takes into account incoming links to a page to indicate the "rank" of a page (though rank and credibility are not synonymous, rank can be a basis for determining credibility).

However, these sites neglect the fact that users have differing viewpoints, and that credibility values should reflect that. As an extreme case, consider a site that a user knows to be spam but that Google thinks is legitimate. The site's links still factor into PageRank, but the user's preference would be for the site to have no impact. The user has no way of recalculating the PageRank with this additional data. Moreover, trust in a source is not a binary attribute; there are many subtle levels that make a global credibility value inaccurate. It may be as simple as one user on a site trusting Fox News to a value of 30%, but another user trusting it to 70%. If a statement from Fox News has a significant bearing on the credibility of a claim, those users should perceive different credibility values for that claim.

When applied to "wisdom of the crowds" sites like Digg, this same principle holds. We can still use the wisdom of the crowds to determine credibility, but individual users should be able to say *which* crowds are used. That is, they should be able to filter results generally based on their own set of assumptions. Moreover, users should be free to change the filtering assumptions. There is value to the user in looking at multiple filters, even if she doesn't agree with them, to understand what informs the opinions of others. This opens up more opportunities for meta-analysis of the data; for example, what are the interesting questions that can be answered by looking at the sets of base assumptions which support particular claims?

Finally, a lack of user-centrism limits users in developing new ways of examining data. Sites generally provide as specific an answer as they can to a question. For example, Digg puts stories on the front page of their site. Google's search engine returns ranked results to a query. But a truly user-centric system should publish as much information as possible about what goes into those decisions, so that the users are free to perform their own alternate computation or meta-analysis.

3. PROPOSED APPROACH

It naturally follows from our critique of current systems in the previous section that our ideal system will have the opposite characteristics: it will be decentralized, automated, and user-centric. In particular, we propose that a system for evaluating the credibility of claims should adhere to the following principles:

- The system should work in a fully peer-to-peer manner. Peers should be able to share information with other peers. Computations carried out by one peer should be repeatable by other peers (provided they have access to the same information).
- The system should be mostly automated. It is acceptable for the system to use human-generated information as input, but the generation of such information should not require expert knowledge.
- The system should strive for the organization of information in a way that is comprehensible to both humans and computers.
- The system should, where possible, provide opportunities for the user to manipulate the input to the system, as well as parameters to any computations.

3.1 Computer Pundits

To implement a more concrete approach, we aim to mimic the positive function of traditional pundits in collecting, organizing, and presenting information without introducing the negative impact of bias and hidden assumptions. Fortunately, the organization of information is a task to which computers are well-suited, and the faithful execution of their input programs is a given. Note that this means computers will not eliminate bias; but they can be trusted to convert bias into results in a deterministic and open manner. In other words, rather than having hidden assumptions, they can openly note and refine explicit assumptions.

We therefore propose to create a computer pundit: a framework for communicating, organizing, and presenting metadata on the validity of information. Computer pundits perform the useful tasks of human pundits (managing the overflow of information) while minimizing the drawbacks. A computer pundit organizes information with a bias known to the user, and should allow individual users to manipulate the system's bias to receive information that matches their assumptions and preferences. As a framework (as opposed to a single piece of software), it must encompass not only algorithms for organizing and refining information, but also methods for collecting information, including the standardization of data forms so that users can communicate with each other about claims, assumptions, and so forth.

The remainder of this section describes ALPACA, a computer pundit that is being developed by the authors.

3.2 ALPACA: A Lightweight Platform for Analyzing Claim Acceptability

ALPACA is a distributed framework for analyzing the validity of claims. By validity, we mean not an objective truth, but a subjective belief by a particular user in the truth of a claim. ALPACA is built on several core principles:

1. **Everything is a claim.** All information can be represented as a set of distinct claims. Each claim by itself carries no information on its source or validity. It is merely an unsupported assertion.
2. **Validity of claims is in the eye of the individual.** Because claims are distinct from their validity, many users may communicate about the same claim, but have different opinions of its truthfulness.
3. **Some claims have semantic relationships, independent of their validity.** Given two claims, "the sky is blue" and "the sky is not blue", they always function as logical opposites. That is, no matter whether you believe one claim or the other, the relationship between the two remains unchanged: the validities have an inverse relationship.
4. **Every claim can be the subject of another claim.** If user A tells user B that the sky is blue, then we have not only the claim "the sky is blue," but also the claim that user A has asserted "the sky is blue," i.e., "user A claims the sky is blue." The idea that user A has told user B can in turn be in question, and user B might tell user C that user A told user B that the sky was blue ("user B claims user A claims the sky is blue"), which constitutes yet another claim. Claims can be recursively made about claims ad infinitum.

3.2.1 ALPACA Design

ALPACA, as a framework, has several components: a conceptual model, a data format, and a software component.

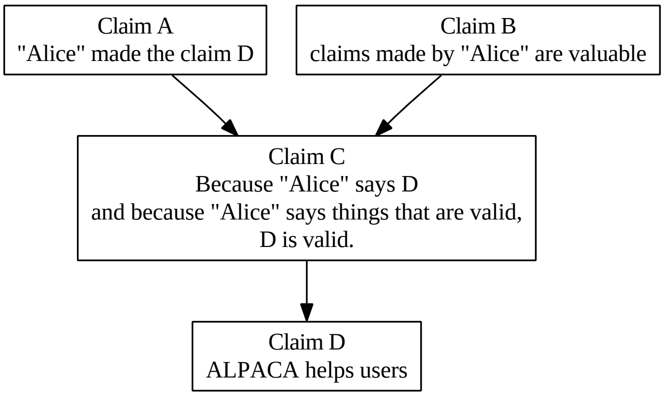

Figure 1: A simple ALPACA graph.

ALPACA's conceptual model is that claims can be represented as nodes in a directed graph, with edges implying the flow of validity from one claim to another. Figure 1 shows a single example. Claims *A* and *B* imply claim *D*. The combining relationship is represented by claim *C* (since *A* or *B* by themselves do not imply *D*).

Each user maintains a personalized validity value for some subset of claims. The validity of a node is either set explicitly by the user, or is gained by "inheriting" validity from a node which points to it. Each user therefore seeds the validity of a graph by specifying some assumption nodes (or obtaining validity values of those nodes from an alternate source), from which the validity of other claims can be computed.

ALPACA can only infer claim relationships if it understands the semantics of each claim. Therefore, each ALPACA claim can be thought of as a claim *type*, coupled with arguments that instantiate a particular claim. In this case, *D* has an opaque claim type; ALPACA doesn't know what it means, but it is the subject of other claims. However, *A* may be said to have an "attribution" type with arguments "Alice" and "*D*".

The data format of ALPACA follows directly from this definition. A "claim file" is simply a textual representation of the claim's type and arguments. Because claims must refer to other claims, we need a global, unique way to refer to each claim. To accomplish this, we name each claim by the cryptographic hash of its contents. Thus claims are self-verifying; if two users talk about claim "1234abcd", they can be sure that, no matter how they might have heard about the claim originally, they are talking about the same thing. Users can publish claims, create new claims, and make statements about existing claims that will be understandable by other users. Because ALPACA claims have a serialized data format, they can be published or transmitted over existing systems such as HTTP or via email.

The software component of ALPACA helps the user perform these tasks; it also allows the user to view and analyze the resulting graphs. The software collects claims from a variety of sources (e.g., websites, email, social networks) constructs a graph; combines it with the user's assumptions; and analyzes and presents the data in a meaningful way, including:

- For a claim of interest, visualizing the supporting evidence; i.e., answering the question "why might I believe this claim?"
- Performing meta-analysis of the graph structure to find root causes of certain outcomes, logical inconsistencies in the user's assumptions, or other patterns of note

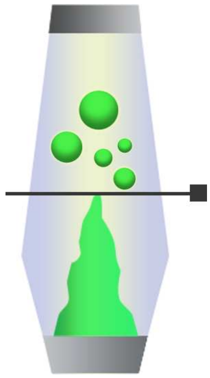

Figure 2: ALPACA GUI basic claim graph view

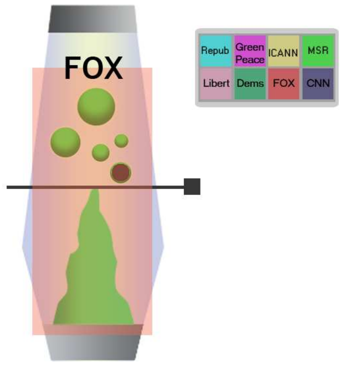

Figure 3: ALPACA GUI single-lens view

- Providing a "lens" effect, where a graph can be viewed not only with the user's assumptions, but with the assumptions of others. This would enable comparative analysis of trust preferences between the user and other claimants.

3.2.2 *ALPACA Interface*

Figures 2 through 4 show a possible graphical user interface design for the ALPACA system. This particular design relies on the "lava lamp" metaphor to easily manage the complexity of the claim graph that might be presented to the user via the ALPACA system as well as considering different perspectives on the graph. Bubbles in the lamp represent claims, with claims of lesser interest aggregated in the lava blob at the bottom. More specifically, figure 2 shows the initial user view of a claim graph as represented by the "lava" in the display. A rudimentary zoom filter is provided as a slider which the user can adjust upwards or downwards to reveal more or less details of the claim graph. The node at the top of the display represents the claim being analyzed by the user. All other nodes below represent related claims which support the primary claim in varying degrees.

Figure 3 displays how a user can utilize the lens palette in analyzing a claim graph. To view the claims from a different perspective, the user can select a lens using a drag-and-drop motion with the mouse over a graph. The resultant view is the perspective of the graph according to the entity represented by the lens. In figure 3, the user has selected the Fox News lens perspective which shows that one of the nodes displayed is not accepted under the assumptions of Fox News as being reliable.

Figure 4 shows basic comparisons between other perspectives

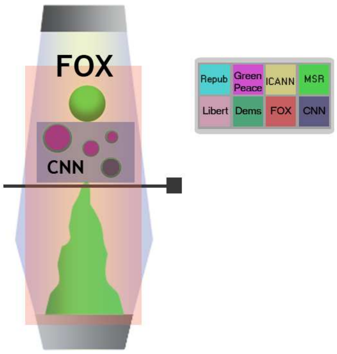

Figure 4: ALPACA GUI double-lens view

using the lens palette. In this example, a user can use both the Fox News lens and the CNN lens to compare which nodes are not accepted by either or both entities.

3.3 Future work

ALPACA is still in its formative stage, and there are many research questions to be answered:

- What claim types must be supported to represent a useful set of statements by users? The set must be broad enough to allow a wide variety of statements, but narrow enough to implement sensible credibility semantics.
- How should validity be represented, and how will values flow through the graph? The flow of credibility through the graph can be considered similar to transitive trust in authentication systems. This topic has been explored before[2, 7, 15] but ALPACA represents a much more generalized form of trust.
- Where do claims come from? Claims may be created explicitly by users, but they may also be created based on observations of the user's computer (e.g., a claim which comes from a website may be attributed). What other data sources can be converted into claims? Some of these issues have been explored in the creation of the FilmTrust system, which mines social networking relationships to build trust graphs[5].
- What novel visualizations of the claim graphs can be used to help users understand the data? Allowing users to "walk" the graph, exploring the supporting evidence for each claim, is one simple interface. The lava lamp metaphor provides another, more intuitive, view of the same data. But there may be other ways of representing the data that highlight patterns larger than a single node.
- What meta-analyses can be performed on the graph structure and the validity values? By seeing what the graphs tell us about our basic assumptions, we can then help the user refine those assumptions to create new graphs. These graphs in turn will allow new information and new refinements, creating a feedback loop.

We plan to develop ALPACA further and address these questions in future papers.

4. CONCLUSION

In this paper, we have enumerated some shortcomings of current systems for looking at information credibility on the Internet. In response, we have proposed a set of design principles and described the basics of ALPACA, a system for examining claim credibility that attempts to follow those principles. ALPACA uses a graph structure to organize claims and references and relies on a simple set of semantic relationships and operations to render and manipulate complex information relationships. ALPACA aims to offer a user interface that allows users to give their own weight to claims and to use powerful tools to cull and accentuate lines of reasoning based on trusted sources.

5. REFERENCES

[1] David Austin. How Google Finds Your Needle in the Web's Haystack. American Mathematical Society. `http://www.ams.org/featurecolumn/archive/pagerank.html`.

[2] T. Beth, M. Borcherding, and B. Klein. Valuation of trust in open networks. In *Proc. 3rd European Symposium on Research in Computer Security – ESORICS '94*, pages 3–18, 1994.

[3] Digg. Web site. `http://digg.com`.

[4] Factcheck.org. Web site. `http://factcheck.org`.

[5] Jennifer Golbeck. *Computing and Applying Trust in Web-based Social Networks*. PhD thesis, 2005.

[6] Insight Magazine. Hillary's team has questions about Obama's Muslim background, January 2007. `http://www.insightmag.com/Media/MediaManager/Obama_2.htm`.

[7] M.K. Reiter and S.G. Stubblebine. Toward acceptable metrics of authentication. In *Proceedings of IEEE Symposium on Security and Privacy*, pages 10–20, 1997.

[8] The Bivings Report. Are Ron Paul Supporters Gaming Digg? Blog entry. `http://www.bivingsreport.com/2007/ron-paul-and-digg-2008-elections-section/`.

[9] Snopes.com. Web site. `http://snopes.com`.

[10] PNNL: Starlight Information Visualization Technologies. Web site. `http://starlight.pnl.gov/`.

[11] The Pew Research Center for the People and the Press. News Audiences Increasingly Politicized, June 2004. `http://people-press.org/reports/pdf/215.pdf`.

[12] The Pew Research Center for the People and the Press. Clinton and Giuliani Seen as Not Highly Religious; Romney's Religion Raises Concerns, September 2007. `http://people-press.org/reports/pdf/353.pdf`.

[13] The Pew Research Center for the People and the Press. Public Knowledge of Current Affairs Little Changed by News and Information Revolutions, April 2007. `http://people-press.org/reports/pdf/319.pdf`.

[14] The Pew Research Center for the People and the Press. Obama Weathers the Wright Storm, Clinton Faces Credibility Problem, March 2008. `http://people-press.org/reports/pdf/407.pdf`.

[15] R. Yahalom, B. Klein, and T. Beth. Trust relationships in secure systems–A distributed authentication perspective. In *RSP: IEEE Computer Society Symposium on Research in Security and Privacy*, 1993.