Information is Power: Toward Intelligent Tools for Information Access and Evaluation

Michael Gasser (gasser@indiana.edu)
Associate Professor
Computer Science Department and Cognitive Science Program, Indiana University
Bloomington, Indiana 47405 USA
Tel: +1 812 855-7078, Fax: +1 812 855-4829

Stephen Hockema (shockema@indiana.edu)
Post-doctoral researcher
Psychology Department, Indiana University
Bloomington, Indiana 47405 USA
Tel: + 1 812 855-2012, Fax: 855-4691

Matthew Kane (markane@indiana.edu)
Doctoral student
Computer Science Department and Cognitive Science Program, Indiana University
Bloomington, Indiana 47405 USA
Tel: + 812 855-9930, Fax: +1 812 855-4829

January 8, 2006

Abstract

Despite the vast array of alternative sources of information brought by the advent of the Internet, comprehensible information on all but the most basic topics remains unavailable to most people, including those with access to ICT, because it is not in a language that the user can understand. At the same time, those who do have access to information because they can read English or one of the other major languages of the Internet face the problem of locating reliable information on a given topic. Much of the information is provided by unqualified agents or is designed to manipulate users through the use of deceptive arguments, exaggerations, and false claims.

We believe that research within artificial intelligence and cognitive science can be brought to bear on these problems in the development of tools that provide access to and evaluation of information, an intelligent Interface to the Information World. In this paper we discuss three specific projects that we have initiated:

1. a system for rudimentary automatic translation between documents in different languages, including languages with few available computational resources
2. a system that calculates believability scores for claims and trustworthiness rankings for sources of claims
3. a system that uses statistical techniques to analyze how particular words are used differently by different writers.

Keywords: artificial intelligence, computational linguistics, digital divide, framing, machine translation

Introduction

Despite the vast array of alternative sources of information brought by the advent of the Internet, comprehensible information on all but the most basic topics remains unavailable to most people even if they have access to ICT because it is not in a language they understand. At the same time, those who do have access to information are faced with the problem of locating information that is reliable and accurate. Much of the information is provided by unqualified agents or is designed to manipulate users through the use of deceptive arguments, exaggerations, and false claims.

In order to realize the full potential of the new information world, users will need access to intelligent tools that facilitate access to and evaluation of information. In this paper, we highlight several contributions that the fields of artificial intelligence and cognitive science can make to the development of such tools.

Access to information and the role of translation

The linguistic digital divide

Among the goals articulated in the Plan of Action resulting from the first phase of the World Summit on the Information Society (WSIS, 2003) is that of world-wide access to information and knowledge (point 10). Access to information in turn presupposes the availability of the information in a form that is comprehensible to the user, that is, in a language that the user can understand. Participation in the Information Society is not necessarily passive, however; a further goal of the WSIS Plan of Action is the fostering of a global dialogue among people of diverse regions and cultural groupings (point 23).

How serious a problem is this? Of the world's 7000 or so languages, about 400 are spoken as first languages by at least 1,000,000 people each, together making up about 90% of the world's population (Paolillo, 2004). If we take into account the large number of people who are fluent in a second language, it appears that the great majority of people could be reached with about 100 languages (though not necessarily through written documents since many people are not literate in their first or second language, if at all).

If we examine the availability of material on the Internet, however, we see an enormous skewing in favor of a small number of languages. The most recent study of linguistic diversity in Web pages (O'Neill, Lavoie & Bennett, 2003) shows that only twelve languages accounted for more than 99% of the pages, with English accounting for fully 72% and languages such as Hindi, Indonesian, and Arabic, each with hundreds of millions of speakers, not even represented among the twelve. In fact this distribution should not be surprising since it is in rough agreement with the languages represented in the library collections of the world (O'Neill, Lavoie & Bennett, 2003). And this distribution of languages is roughly reflected in the community of Internet users and the languages with which they are familiar (Paolillo, 2004). Thus, as measured by language, the Internet has not led to a more inclusive Information Society.

In fact the Internet is even having a negative impact on linguistic diversity. Because the use of some non-roman writing systems is still cumbersome on the Internet, some communities prefer a transcribed romanized version of the languages for informal communication. And because elites within many linguistic communities know English, they may even prefer to communicate in English (Paolillo, 2004). These practices only hinder attempts to integrate the languages concerned into the information world and reinforce the class divisions within these linguistic communities by making some documents inaccessible to those who are unfamiliar with the roman alphabet or with English.

A less important, but still significant, divide separates members of particular academic disciplines from those who are not trained in these disciplines and cannot easily read documents written by scholars in these fields, even though the documents are written in their native language. While some of the difficulty is obviously due to unfamiliarity with particular concepts, we believe that some of it is also due to the linguistic register used in scholarly documents, the word choice and complex syntactic constructions that define scholarly discourse. Many documents written in this register are available on the Internet but may be inaccessible to most users in part for linguistic reasons.

The role of translation

Overcoming the linguistic digital divide will depend crucially on translation. Documents appearing originally in one language are accessible only to those literate in that language until they are translated into other languages. Even with access to ICTs, a person who is functionally literate in only Khmer can only read documents or listen to sound files that are in Khmer and can only communicate with other users who know Khmer.

Given the general unavailability of most documents on the Internet in only one language and the desirability of ultimately having most documents available in a set of, say, 100 languages, the translation task is a formidable one: roughly 800 billion separate translations to make the pages searched by Google™ as of September 2005 accessible to the speakers of all of those languages. Clearly machine translation must be a part of a future Information Society for all.

There is no general agreement on criteria for evaluating MT systems, but a recent competition organized by the National Institute of Standards and Technology using unrestricted texts in two pairs of languages (NIST, 2005) gives a rough idea. For one pair of languages, the best results achieved were 50% accuracy, for the other pair 35%. Thus one conclusion that is obvious is that MT still has a long way to go if its goal is human-like translation of unrestricted texts. While performance improves when the texts are restricted to those in a particular domain, for the near future unconstrained MT will require human editors.

Most existing MT systems operate with particular source and target languages or with small sets of languages. Many of these systems rely on built-in lexical and grammatical resources for the languages that are covered, the result of years of linguistic and computational research on those languages. An alternative to this knowledge-based approach to machine translation is *statistical machine translation* (SMT), which relies on statistical techniques for learning the correspondences between the languages and relatively little built-in knowledge. Given the lack of computational resources for most of the languages in which we are interested, SMT is the only viable alternative.

SMT systems rely crucially on the availability of multilingual corpora on which the systems are trained, and performance improves with training. Obtaining such corpora is no trivial matter, but a recent experiment on developing a SMT system from scratch for translation from Tamil to English found that usable output could be produced with a month of work translating and training the system (Germann 2001).

Starting with insights from existing SMT system, we are currently developing a framework for training multiple language pairs simultaneously. We assume that the evolving database will reside on a central server and that it will be continually trained as more translation data become available. Performance in our system is improved greatly if the training data are sequenced in complexity; therefore we will begin by training on documents aimed at children, focusing first on health-related topics and folk tales. Output texts, at first largely incomprehensible, will be made available on the Internet, where "open editing" Wiki software (Leuf & Cunningham, 2001) will permit them to be edited by native speakers of the target languages who are familiar with the content or with the source languages. Corrections made by human editors are fed back to the machine translation system, which updates its knowledge of the correspondences between the languages accordingly. We envision a continually evolving system in which knowledge of particular languages within the system is bootstrapped off of what has already been learned for related languages. The steps in the process are illustrated here with an English-to-Swahili example:

- Original English text on health:
  … *How can people protect themselves against cholera?* …
- Initial automatic translation into Swahili entered on Wiki:
  … *Watu wanaweza kujikinga vipi dhidi ya maradhi ya kipindupindu?* …
- Translation as edited by Swahili speaker on Wiki:
  … *Watu wanaweza kujikinga vipi na maradhi ya kipindupindu?* …
- Feedback to system
  Association between English *against* and Swahili *na* is strengthened in the context of {protection *against* diseases}.

Translation within languages, from more formal to less formal registers, is a somewhat different problem since it is mainly a matter of replacing technical terms with more familiar ones and simplifying a relatively small set of complex grammatical constructions. Here we hope to collaborate with researchers at Hewlett-Packard Japan working on this problem in Japanese.

Quality of information

Access to information does not, in and of itself, solve the problem of maintaining an informed public. Any user of e-mail and of the World-Wide Web is familiar with the fraudulent claims that come with much of the advertising as well as with the bogus business offers, the phishing and the urban legends that seem to permeate the Internet. While it is important that users be protected from such blatant deception, we believe that the problem is more fundamental. More subtle forms of deception are well-known from conventional advertising and political persuasion, and the control over the sources of such information has always been in the hands of the powerful.

How does all of this change with the Internet? First, many more people become potential targets of advertisers and political agents. People who have not previously had access and are suddenly flooded with information may be particularly vulnerable, as they may have less experience with common tricks and their cultures may have different expectations and norms concerning persuasion and the presentation of information. (Note that here we do *not* mean to imply that groups of new internet users are in any way inferior in their critical or evaluative thinking!)

Second, while some believe that the Internet will make the distribution of information more equitable, it will continue to be the case that entities with resources can dominate because they will have access to servers and to technicians proficient in sophisticated presentation techniques as well as in techniques to undermine entities with competing views. More and more we may be seeing a kind of information war. One piece of evidence for this is an important recent document released by the United States Department of Defense (Defense Science Board, 2004), in which it is argued that "the United States is engaged in a generational and global struggle about ideas"; that "public diplomacy, public affairs, psychological operations and open military information operations must be coordinated and energized" in a new "strategic communication" policy that "support[s] the nation's interests". By "psychological operations", the report means "military activities that use selected information and indicators to influence the attitudes and behavior of foreign governments, organizations, groups, and individuals in support of military and national security objectives". In the context of the new information world, the report concludes that "political struggles" are no longer about control over "scarce information", but about "the creation and destruction of credibility".

In the face of such blatant uses of the new technology to influence people around the world on behalf of an entity such as the US Department of Defense, we believe participants in the Information Society need to be protected from deceptive uses of language. That is, the ability to *access* vast amounts of information needs to be offset with the ability to *filter* and *critically evaluate* this information. Thus, in this section we propose a tool suite to assist in the process of information evaluation.

Information comes to an information consumer in both **overt** and **covert** forms. By overt information, we mean explicit claims. For example, in his October 6, 2005 speech about US policy in Iraq, US President George Bush said, "these extremists want to end American and Western influence in the broader Middle East, because we stand for democracy and peace and stand in the way of their ambitions." Here he is making a claim about the motivation for the actions of a particular group of people. By covert information, we mean information that is implied by word choice or by the pattern of co-occurrence among words. For example, in Bush's statement, he uses the word "extremists" to refer to the people he is describing and the words "democracy" and "peace" to refer to the policies that they are claimed to oppose. Each of these nouns carries with it connotations about the entities that are referred to.

Evaluation of overt information: believability and trustworthiness

We have seen that as far as the US Defense Department is concerned, the "credibility" of information and information sources is at the heart of the "global struggle about ideas"; that is, what is at issue is whether a user believes particular claims. If we assume that claims can be distinguished from statements of opinion (not a trivial assumption), then for a given claim, we would like to be able to assess **believability**. Believability depends on the **trustworthiness** of the source or sources of the claim and whether a source claims to have direct experience with the reported fact or is citing a secondary source. The trustworthiness of a source in turn depends on its reliability, how believable its past claims have turned out to be.

We envision a database of propositions whose believability ratings are continually updated on the basis of new claims made about them and a database of information sources whose trustworthiness ratings are continually updated on the basis of new claims or new information about old claims. If, as we hope, the new information world will be one in which "everyone is a journalist", the sources in the database will include not only familiar ones such as CNN, Agence France-Presse, and the US State Department, but also ones that may be familiar only to the system itself. In fact we are interested in developing means by which the actual identity of a particular source is kept secret while that source's trustworthiness rating is maintained within the system.

While these ideas are only in their infancy, there is a body of relevant research in artificial intelligence within the areas of **belief logic** and **truth maintenance**. Belief logic provides a computational framework in which inferences are made about beliefs and about the truth of propositions based on a database of facts, including beliefs of agents. Truth maintenance theory is a framework within which a database of propositions keeps track of the dependencies between the propositions so that when new information about a proposition becomes available, the truth of related propositions can be updated appropriately.

Consider a simple example, the following statement from BBC News on October 29, 2005: "The US National Hurricane Center said maximum sustained winds had increased to nearly 120 km/h - making it [Hurricane Beta] a Category One hurricane." This is a claim made by BBC News citing as the source of the information another entity. The believability assigned to this claim by the system would depend both on the current trustworthiness of BBC News and the US National Hurricane Center, and independent evidence in favor of the truth of the claim would raise the trustworthiness of the US National Hurricane Center.

Claims in political or economic contexts would present greater difficulties. Consider the statement made by President Bush in his speech, cited above. We first have the problem of finding a referent for "these extremists". Equally seriously there is ambiguity in the claim: are the "extremists" opposed to policies that *they* would label (presumably in other languages) as "democracy" and "peace", or are these Bush's labels for policies that the "extremists" perceive otherwise? Given the frequency of similar sorts of claims, it seems that we will need to score them on a scale of vagueness as well as of believability.

Evaluation of covert information: bias and framing

Beyond overt information, the process of information evaluation must encourage readers to think more deeply about the messages they are receiving. For example, people are extremely susceptible to the implications of words. Within cognitive science there is a good deal of research dealing with the phenomenon of memory construction and suggestibility. For example, Loftus and colleagues have shown that eyewitness testimony is greatly influenced by the types of information to which the witness is exposed following the target event. In one experiment, after viewing a video of a car wreck, one group of witnesses were asked, "How fast were the cars going when they *bumped* into each other?" and another group of witnesses were asked, "How fast were the cars going when they *smashed* into each other?" Not only did the witnesses who were asked the second question have higher estimates of the cars' speed, but they also were more likely to subsequently report remembering seeing broken glass in the video when in fact there had been none (Loftus & Palmer, 1974). Thus, it appears that the choice of even a *single word* can greatly influence how people internalize and remember an event. Inaccurate news summaries or leading questions by authorities have been shown to be even *more* influential on witnesses' memories.

The power of media to affect people's worldviews is well-known. And the stakes can be extremely high! For example, the Rwandan genocide of 1994 was preceded by years of build-up in the radio media in which Tutsis were framed as subhuman (for example, as *inyenzi* 'cockroaches') with ultimately horrific consequences. And the recent deception that led the majority of United States citizens to accept a war based on deception and even to re-elect the president that deceived them is well-known to the world. Less understood is how the media and information were manipulated to achieve this. However one strategy seems to have been to broaden the meaning of the word *terrorist* – a word that was already associated with strong emotions and fear in many U.S. citizens during the build-up for the war – so that it could be applied in Iraq, while narrowing the meaning of the word *Muslim* (and other related words) to become more associated with fundamentalist extremism. Obviously these issues are extremely complex and the system we propose herein cannot solve them – it can only attack pieces of the underlying problems. Nevertheless, we seek to create a mechanical system with as little bias as possible that will assist in evaluation of information from the media, both in high-stakes cases such as these and in more pedestrian and mundane cases, where the ramifications of word and phrasing choices, while not as potentially devastating, can be just as influential if accumulated over long periods.

As the above examples illustrate, it is often the case that, in addition to the information available overtly "on the surface" in text, there is *covert* information as well. The cognitive scientist and linguist George Lakoff has developed a theory of political language usage in which issues are *framed* according to underlying metaphors employed by the speaker/writer (1997, 2004, 2005). This framing is usually implicit and forms a sub-text to the actual surface message that can communicate and even change the audience's worldview. For example, when a politician refers to the *tax burden* and *tax relief*, in addition to the negative valence of the word *burden* that carries over to the word *tax*, these phrases also invoke a pre-established cognitive framework that includes assumed relationships between individuals and their government. (Lakoff has identified this particular case as part of a larger framework encoding the relationship between individuals and their government as familial, specifically with the government playing the role of a "strict father".) Or, to take another example from above, the use of the term *inyenzi* in Rwanda probably deliberately brought to mind a framework associated with cockroaches, vermin and other pests that invade your home (e.g., implying that Tutsis were non-native or not legitimate citizens of Rwanda), are unclean, and have to be actively exterminated by vigilant housekeepers, all implying the sorts of actions that were eventually taken in the genocide. If audiences accept these words and phrases uncritically or even begin to use them themselves (even if they are arguing that taxes are not too high or that the author's ideas about tax relief will not work, or that *inyenzi* should not be attacked), then they may also gradually come to adopt the metaphorical framework behind the phrase's usage. This potential makes understanding and recognizing framing very desirable, especially for political discourse.

It is noteworthy that people are remarkably naïve about the effects of such framing. Indeed, research from cognitive science in the realm of advertising and persuasion shows that the more confident people are about their ability to ignore such covert messages, the more influential they may be (Sagarin, Cialdini, Rice, & Serna, 2002)! In the case of Rwanda, this naiveté can be seen in the attitudes of both the French and the American ambassadors who opposed any action against Radio-Télévision Libre des Mille Collines (RTLM), the main purveyor of hate messages directed towards Tutsis. For example, the U.S. ambassador at the time claimed that it was the best radio for information and that its "euphemisms were subject to many interpretations" (Radio Netherlands Media Network, 2005).

Although Lakoff's ideas have been widely adopted and he has employed the theory of framing to analyze political discourse in several contexts, little work has yet been done on how a naïve reader/hearer might identify framings being employed in a message that could be covertly influencing their own perspectives on an issue. How does the metaphorical framework assumed by a source get realized in the message? The natural answer is in the words they chose. If speakers/writers conceptualize the role of government in their life as metaphorically like the role of a strict father, enforcing the rules and preventing them from having too much fun, for example, then they are likely to talk about giving money to the government in the form of taxes as something that is oppressive. For a conceptualization such as this, they might choose expressions like *tax burden* or *tax relief*. On the other hand, if speakers/writers conceptualize the role of government in their life as more like the management of a club in which they have active membership, they might choose to use words like *duty* or *dues* when discussing taxes.

There is nothing profound here: the words people choose obviously reflect their attitudes, beliefs, and views. However, what we would like to suggest is that this simple fact can be put to effective use. Information sources can be automatically clustered into groups of authors who use words in similar ways. For example, conservative and liberal authors, who presumably have differing views on the proper role of government, would write about taxes in very different ways. In this grouping process, certain words, phrases, word co-occurrences and word usage patterns will stand out as being especially useful for discriminating the groups and forming the clusters. For example, the phrases *tax burden* and *tax relief* might very reliably identify conservative authors. But, we can go a step deeper than just identifying surface-level phrases that distinguish groups of authors. It is likely that there will be words and phrases that on the surface appear identical across the groups but when viewed in context can be seen to have very different meanings. The single word *tax* in this example is such a word. Both liberals and conservatives alike would probably claim to share the same *definition* of the word *tax*, but from the contextual patterns in which it occurs in their writings (for example, very frequently preceding *burden* or *relief*), it can be inferred that the two groups have radically different underlying metaphorical conceptualizations of it. These different conceptualizations are similar to different *senses* of a word, but subtler. Importantly, though, they will have similar effects on sentence and discourse structure as variations among more-traditional word senses have. For example, text about the sense of the word *space* typical of the phrase *outer space* will differ from text about the sense of the word *space* typical of the phrase *space bar* in the types and meanings of surrounding words and probably the syntactic structure of the sentences in which they are found in much the same way as liberal and conservative usages of the word *tax* differ.

In this section we propose a tool that operates just as the previous paragraph describes, where the author clusters are discovered automatically using a tool based on a word-sense disambiguation algorithm. The point is *not* to be able to associate the discovered uses ("senses") of words like *tax* with simple labels like "conservative" or "liberal", rather it is simply to identify words and phrases that are *polarizing*: words that are consistently used only in certain contexts by one group of authors and only in certain other contexts by another, distinct group of authors.

How might this be useful to naïve users? If polarizing words can be highlighted in real-time as they read a new text, perhaps by an unknown author, users might use this as an additional source of information about both the author and degree of bias of the information they are reading. Many highlighted words concentrated in a passage could place the user on guard and elevate their attentiveness when reading a section, encouraging them to think more deeply about both the overt and the covert messages they are absorbing. In the context of the above example, the system might highlight a word like *tax* in a new document (even if in the current context the words *burden* and/or *relief* do not directly appear), indicating that there were two relatively distinct sets of authors using the word *tax* in two relatively different and distinct ways (contexts). At a minimum, this might be helpful already by encouraging the reader to pay better attention to the context surrounding the instance of the word and even to wonder why the word *tax* might be polarizing. Furthermore, the system would allow the curious reader to then dig deeper if so desired to discover 1) examples of different contexts for the word *tax* by authors in the different polarized groups, and 2) which of the polarized groups of authors the current usage pattern most closely resembles. In general, the fact that the author chose to use the word *tax*, in and of itself, is probably not that revealing or interesting to readers (unless they want to know the answer to question 2), but if the author consistently uses words of this nature, it is a sign for readers to become more active in critical evaluation of the covert message and framing of what they are reading. Because of the statistical nature of this type of analysis, it is unlikely that all of the words that get identified will be useful; highlighted words and phrases will probably be sprinkled throughout all documents. We hope readers will become used to this and begin to "see through" the coloring system (much as we have become used to seeing underlined text when it is hyperlinked so the underlining does little to distract us from the content of the underlined text). However, we hope that the presence of an inordinate number of highlighted words in any particular passage will be enough to attract the reader's notice and elevate attentiveness.

Front end interface

The primary motivation behind our system's user interface is simplicity. We want accessibility to transcend social class and education level and the technology to be accessible to users regardless of skill level. Therefore our goal is a front end that is as unobtrusive as possible, while still being customizable to fit a particular user's preferences. This will require that the system to be deployable on today's most common web browsers and that no significant changes will become mandatory for the user. Furthermore, no obligatory registration processes should exist, which could deter users. Clients should not need to make any changes or even be aware if alterations are made on the back-end servers. Ultimately, we aim to have a demand-based system that only affects changes to a user's screen when desired. We accomplish this through JavaScript bookmarklet additions to web browsers.

Bookmarklets appear to be a computationally friendly method for interacting with the user. These are JavaScript programs contained within a bookmark in the user's web browser and use the javascript protocol. That is, while most areas on the World Wide Web accessed by clients use the hypertext transfer protocol (i.e. http), bookmarklets consist of web site addresses that begin with "javascript" instead of "http." In fact, bookmarklets can be saved and used in the exact same way as normal bookmarks. Bookmarklets are most commonly used to alter the way a client sees someone else's web page or to extract data from a web page. As we will see, our bookmarklet takes advantage of both of these usages.

Since the popularity of JavaScript has soared in the last decade, numerous sites implement JavaScript's capabilities. The vast majority of Internet users today have JavaScript enabled on their web browsers; thus security flags will be avoided when a client attempts to use our system. As a result, bookmarklets are safe to use and do not require software installation. They live entirely within the web browser and will not cause conflicts on a user's system, a positive consequence of JavaScript's inherent security restrictions. Finally, perhaps the greatest benefit of bookmarklets is that users have the freedom to implement its functionality with ease whenever they choose, while at other times the bookmarklet remains in the web browser bookmark toolbar nearly out of sight.

Through JavaScript, we are able to connect to a back-end server, potentially shared by many users, that provides the client with a list of keywords to highlight. Accompanying each of the keywords will be an example context from each major sense cluster that the back-end algorithm identified, thus revealing ambiguity among superficially-similar words or phrases. Additionally, each of these words that appears in a user's current document will become uniquely colorized, hence drawing immediate attention to consistent word repetition or heavy use of polarizing words. For example, consider a word currently discussed (and disputed) in the media today, such as *interrogation*. While both ends of the political spectrum possess strong feelings concerning the proper role of and limits for interrogation, they may naïvely believe that all share the same underlying meaning of the word. However, judging from the widely-differing contexts surrounding the actual usages of this word, we can conclude it is likely that the meanings for this word diverge across authors of different political beliefs. For a pro-war activist who is familiar with the interrogation tactics used by the United States government, the idea of interrogation may refer to harsh but justifiable treatment of a terrorist for the sake of extracting information vital to the security of the world. In contrast, a citizen who vehemently opposes both war and severe treatment of prisoners probably views current United States interrogation tactics to be torturous human rights abuses. Obviously, neither of these world views can be declared correct or incorrect by an automated system; rather, the purpose here is to discover how meanings for a word such as *interrogation* contrast across the ideological continuum. Given a skillful writer attempting to advance a biased agenda, a system like ours might help a reader to more clearly determine where on the spectrum such a writer sits.

We hope that by providing more information about the semantic contexts behind such divisive words, users will focus more attention on distinctive usages of repetition, metaphor, and framing, enabling them to become more critical readers. However, we believe simply identifying dissimilar senses of a word may not help users become fully aware of the nature of political rhetoric today; rather, it appears necessary to identify clusters of authors who tend to use a polarizing words in a similar manner. An inherent assumption here is that not only do there exist contrasting senses of important words, but specific groups of people tend to use a particular sense consistently. For example, many users might be surprised to realize that a certain group of authors use the word *taxation* in a sense very different from another group of authors.

In keeping with our design goal of unobtrusiveness, users will not receive any supplemental information about any of the highlighted words until they decide to make a mouse-click on one of them. When they make a selection, a new web browser layer will appear next to the word in question displaying examples of the multiple senses associated with this word. Finally, machine learning techniques will enable the system to learn user preferences and subsequently ignore certain categories or distinctions that users do not find helpful. For example, if a user consistently fails to investigate colorized words related to the "taxation" frame or the "marriage" frame, then occurrences of these words will no longer be called to the user's attention; rather, they will be ignored entirely for this user. Of course, any user will have the ability to view current keywords that are ignored and opt to revert back to viewing the original set of keywords.

Back end algorithm and server

The specific implementation details for the above features are quite simple and do not require unreasonable technical overhead. The current approach for delivering the set of keywords and contexts to the user is through browser cookies, which commonly can hold 4KB of data. With the help of Java Servlets, the bookmarklet can connect and receive a temporary cookie containing the essential data, which will in turn alter the JavaScript with a new set of important words for the next time the system is called by the user. Initially, this cookie will reside in the browser's cache. However, to maintain synchronization with the back end, the cookie will expire after a short period of time, causing the bookmarklet to again contact the Servlet for an updated set of words.

From the point of view of the Servlet, the most important task is document retrieval and analysis. The Servlet must mine the web nightly and retrieve popular news articles and key weblog postings through, for example, the help of services like Intelliseek's BlogPulse, a service that specializes in locating daily lists of key persons, key phrases, and key paragraphs from thousands of weblogs. Once a set of documents has been retrieved, the pattern of word usage across documents is analyzed to determine words and phrases that are highly discriminative of clusters of authors.

We are presently implementing Schütze's (1998) Automatic Word Sense Discrimination algorithm followed by a clustering algorithm to accomplish this task automatically. The unsupervised word sense discrimination algorithm addresses the problem of distinguishing senses among multiple occurrences of a word that is used in an ambiguous manner. Lexical ambiguity is a common problem for natural language processing systems. For example, among other things, the word *space* could denote a blank character on a typed page or the void region surrounding our planet. In an algorithm like Schütze's, the context in which the word is used specifies the intended sense. We hope to be able to apply the same technique to the different "senses" of a word such as *tax*, and we hope to find that authors can be distinguished by the patterns of word senses that they use for such words. The words that are most discriminative of author clusters are the ones that will be highlighted by our system. Schütze (1998) was originally concerned with sense discrimination, rather than sense labeling. Similarly, our goal is to merely call attention to divergent senses while leaving the user to determine exactly how to characterize such senses.

Although the full details of the analysis algorithms are beyond the scope of this paper, we will give an overview here. Keep in mind that these analyses happen automatically without intervention or supervision from a human. The part-of-speech of each word is first determined (via a standard probabilistic technique using a Hidden Markov Model) and common inflections are removed using basic rules to allow lumping different forms of a word stem together in the analyses. (Although word stems are the unit of analysis, in what follows we will refer to "words" for the sake of simplicity.) A compact representation of the co-occurrences of a word with other words is created for each word in the set of documents under consideration. This representation takes the form of a vector and may be thought of as the *semantic representation* for each word. For every occurrence of an ambiguous word, a corresponding *context representation* vector is created. The context representation for any word is generated by considering the semantic representations of each word that co-occurs with the word in question and essentially computing an "average" meaning among all of these vectors. After the system has recorded occurrences of a word in multiple contexts, the observed context vectors can be automatically clustered so that each cluster defines a word sense.

For example, everyone is familiar with the different senses for the word *space*: outer space, office space, dimensional space, etc. When the word *space* is encountered in a document, a context vector is created from the words around it, which might include *gas*, *vacuum*, and *gravity*. At this point, the context for this occurrence of *space* can be determined based on how similar its context representation is to other clusters of contexts for *space*. While multiple senses for words like *space* or *bank* are not debatable, we hope to make users better aware of multiple covert "senses" of words being used differently by different groups of authors. In other words, there probably does not exist a subpopulation of authors that has a pattern of usage for the word *space* that is different from the general population's pattern of usage as a whole. However, we believe words such as *interrogation*, *freedom*, or *tax* are in fact used in contrasting ways by different groups of authors.

As another example, consider the word *suit*. Two common senses for *suit* include those associated with legal contexts and clothing contexts. Suppose the following words co-occur with a particular usage of *suit*: *law*, *judge*, and *statute*. Now the compact semantic representations for each co-occurring word can be extracted and the "average" location among all of these representations can be taken to indicate the context representation of this usage of *suit*. To make these ideas more concrete, consider the following figure.

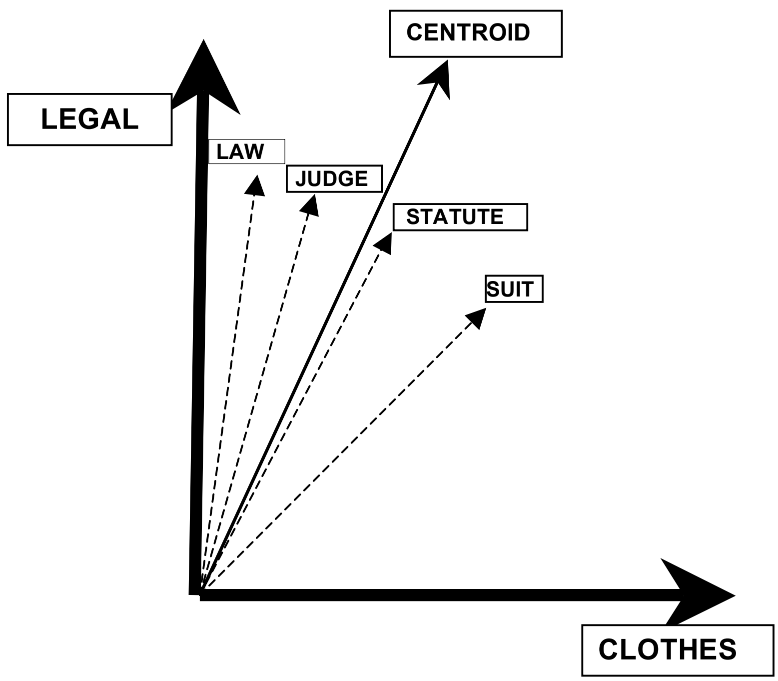

Figure 1: Simplified semantic representation and context representation for an instance of the word *suit*

Here, the horizontal axis indicates the clothes context and the vertical axis the legal context. Each word that co-occurs with *suit* in the usage in question is indicated by a dashed arrow and the angle of its placement is determined by its semantic representation as described above. Thus the word *law* has a semantic representation much closer to the legal context than to the clothes context; in contrast, the word *suit* in general appears equally often in a legal context and a clothes context. Finally, the solid arrow marked "centroid" indicates the average meaning among all of the co-occurring words for this particular usage of *suit*. Since the centroid arrow is closer to the axis corresponding to the legal context, we can conclude that this usage of the word *suit* is in fact referring to its legal sense. Note that the system does not provide a label such as "legal" for the sense. The algorithm is designed to merely locate *differences* in senses; the precise semantics behind each sense is not identified.

Once a "word sense" has been determined for each word in a set of texts, the next step is to determine groups of authors who use words in similar ways, that is, use similar senses of a word in conjunction with other word senses. Like the identification of word senses from clusters of context vectors, this can also be achieved using an automatic ("unsupervised") clustering algorithm. Finally, the words or phrases that best discriminate the author clusters can be identified and selectively highlighted by the front-end interface. Again, the ultimate purpose here is to call the user's attention to words and phrases that are used to mean different things by distinct subpopulations of authors. The more such polarizing words there are in a particular text, the less generally applicable its claims are likely to be. We believe just highlighting this fact can help a user to make a critical assessment of texts, especially if the texts are written by an author with whom they are not yet familiar.

Conclusions

From its beginnings, research in artificial intelligence and computational linguistics has been tied closely to the goals of the military establishment in the United States. There are many obvious military applications for intelligent systems, including ones that use the kinds of computational approaches we have discussed in this paper. In recent years, corporate research and development in these fields in the US, Europe, and Japan has also been an important source of funding. We see the results of these efforts in the increasingly sophisticated tools available to Internet users who have access to them.

Cognitive science has a more theoretical orientation than artificial intelligence, but when it has been applied, the projects have again had either military or industrial orientations. Both the military and corporations are interested in influencing people's belief systems, and the science of the mind has obvious relevance here.

We would argue, however, that little of the progress that has been made in understanding the nature of information, of intelligence, of concept formation, of natural language has benefited humankind in general. In this paper we have outlined a few ways in which these fields *could* contribute to a more informed public and ultimately a more just and democratic world. It is clear that such contributions can only be realized, however, if there is a radical rethinking of how research in these fields is funded and how the fields are presented to the students who will be conducting the research in the future.

References

Brown, P.F., Cocke, J., Della Pietra, S.A., Della Pietra, V.J., Jelinek, F., Lafferty, J.D., Mercer, R.L., Roossin, P.R. 1990, 'A statistical approach to machine translation', *Computational Linguistics*, v.16, pp.79-85.

Defense Science Board 2004, 'Report of the Defense Science Board Task Force on Strategic Communication (September 2004)', Washington, DC, Department of Defense, Available at: http://www.acq.osd.mil/dsb/reports/2004-09-Strategic_Communication.pdf.

Germann, U. 2001, ‘Building a statistical machine translation from scratch: how much bang for the buck can we expect?’, ACL 2001 Workshop on Data-Driven Machine Translation, Toulouse, France, Available at: http://www.isi.edu/natural-language/people/germann/publications/germann-acl-2001-DDMT-workshop.pdf.

Intelliseek’s BlogPulse, Available at: http://www.blogpulse.com.

Lakoff, G. 1997, *Moral politics: how liberals and conservatives think,* University of Chicago Press.

Lakoff, G. 2004, *Don't think of an elephant: know your values and frame the debate – the essential guide for progressives,* Chelsea Green Publishing.

Lakoff, G. 2005, ‘Simple Framing: An introduction to framing and its uses in politics’, Available at: http://www.rockridgeinstitute.org/projects/strategic/simple_framing/view.

Leuf, B. and Cunningham, W. 2001, *The Wiki way: collaboration and sharing on the Internet*, Addison-Wesley Professional, Indianapolis, IN.

Loftus, E.F. & Palmer, J.C. 1974, ‘Reconstruction of automobile destruction: An example of the interaction between language and memory’, *Journal of Verbal Learning and Verbal Behavior*, v.13, 585-589.

National Institute of Standards and Technology 2005, ‘NIST 2005 Machine Translation Evaluation’, Washington, DC, NIST, Available at: http://www.nist.gov/speech/tests/summaries/2005/mt05.htm.

Radio Netherlands Media Network 2005, ‘Hate Radio: Rwanda’, Available at: http://www2.rnw.nl/rnw/en/features/media/dossiers/rwanda-h.html.

Paolillo, J. 2004, ‘Language diversity on the Internet: Examining linguistic bias’, UNESCO Institute of Statistics, Available at: http://www.uis.unesco.org/ev.php?ID=5648_201&ID2=DO_TOPIC.

Sagarin, B. J., Cialdini, R. B., Rice, W. E., & Serna, S. B. 2002, ‘Dispelling the illusion of invulnerability: The motivations and mechanisms of resistance to persuasion’, *Journal of Personality and Social Psychology, 83*, 526-541.

Schütze, H. 1998, ‘Automatic word sense discrimination’, *Computational Linguistics*, v. 24, pp.97-124.

World Summit on the Information Society 2003, ‘Plan of Action’, Geneva, Available at: http://www.itu.int/wsis/docs/geneva/official/poa.html