UNIVERSITY OF SOUTHAMPTON

FACULTY OF PHYSICAL SCIENCES AND ENGINEERING

Electronics and Computer Science

Semantically Representing Informal Argumentation from the Social Web

by

Tom Blount

A progress report submitted for continuation towards a PhD

Supervisor: Dr. David Millard

Co-Supervisor: Dr. Mark J. Weal

Examiner: Dr. Nicholas Gibbins

July 9, 2014

UNIVERSITY OF SOUTHAMPTON

ABSTRACT

FACULTY OF PHYSICAL SCIENCES AND ENGINEERING
Electronics and Computer Science

A progress report submitted for continuation towards a PhD

SEMANTICALLY REPRESENTING INFORMAL ARGUMENTATION FROM THE SOCIAL WEB

by Tom Blount

Argumentation, debate and discussion are key tools within human communication, shaping the way people think and form new ideas. As the social web becomes more prevalent, the study of how users engage with one-another becomes increasingly important. Antisocial behaviour, which often stems from argumentation, can have a negative impact on online communities, driving away new users and stifling participation. To encourage and stimulate respectful debate, the notion of "reflecting" an argument back to the participants is proposed, to provide an overview of the argumentation structure and to highlight key features (such as relevance of points or credibility of participants). However, preliminary work indicates that existing techniques for modelling argumentation are insufficient to capture the structure and dynamic of argumentation taking place on the social web. Future work will focus on augmenting these systems to better capture the available data, and developing the means to represent it back to users in a useful and supportive manner.

Contents

List of Figures

List of Tables

Acknowledgements

Thanks to my supervisors David Millard and Mark Weal for their help and guidance, and to the EPSRC for providing the funding for this work.

Chapter 1

Introduction

1.1 Problem Space & Motivation

Argumentation is fundamental to human communication – it is how people share new information and new ideas, and propose courses of action that will have a favourable outcome (at least, for the proponent) (Hahn et al., 2005). As a result, there is a large amount of research from a wide variety of disciplines including philosophy (Johnstone, 1952), sociology (Pólos and Hannan, 2002), law (Bertea, 2004) and artificial intelligence (Bench-Capon and Dunne, 2007).

However, as Van Eemeren and Grootendorst (2004) note, *"perhaps out of fear of metaphysics or of 'psychologizing,' present-day logicians tend to concentrate exclusively on formalized arguments that lack any direct relation with how argumentation is conducted in practice."* Social argumentation, or the way people argue day-to-day, often has a very different structure to formalised models. In these instances, the aim of a proponent is not to prove himself right through irrefutable logic, but simply to make others believe that he has proved himself right. Schopenhauer, in his satirical work *The Art of Always Being Right*, emphasises that *"A man may be objectively in the right, and nevertheless in the eyes of by-standers, and sometimes in his own, he may come off worst"* (Schopenhauer, 2009).

This is particularly relevant when applied to the social web. As a network of social relationships, that are created, formed and maintained through the world wide web, the social web (and the social media presented across it) are rife with discussion, debate, and argumentation (Rowe et al., 2011). The social web and social media present a number of challenges for extracting and analysing argument. This is due to the often informal language and slang used, the differing "genres" and audience types and the lack of structure or argument "markers" (Schneider et al., 2012). As the use of social media grows further (Smith, 2009, p. 559), it is important that these types of argument can be categorised accurately so that systems can make full use of this information source.

1.2 Aim & Research Contribution

The ultimate aim is to provide a means of *"furthering reflection"* of argument over the social web. The principle of this reflection is to emphasise *"...the possibilities of using argumentation to resolve differences of opinion and...to stimulate people to engage in a critical dialogue if they want to convince another person"* (Van Eemeren and Grootendorst, 2004). In other words, by showing a user of social media the overall structure of an argument and highlighting particular points of value (or downplaying those of little value) they can ideally be encouraged to contribute to the discussion in a more thoughtful manner, rather than resorting to insulting or demeaning their opponents.

To achieve this, it is first important to correctly model and represent the arguments that occur socially. In this way, the key features of informal arguments can be identified and systems adapted around them to produce more valuable argumentation. The first approach is to examine how formal models map arguments, and apply an appropriate model (or combination of models) to an argument (or arguments) on the social web to determine which features are well captured, and those that are not.

1.2.1 Objectives

1. To examine the methods by which people model (and represent) argumentation (and for what purpose)
2. To apply these models to the domain of the social web and, in doing so, analyse the strengths and weaknesses of these models
3. To use this analysis to construct a novel, or extend an existing, model to better represent argumentation on the social web
4. To use this model to reflect a users argumentation back to them, to encourage them to improve their critical reasoning and conduct

This paper proceeds by covering the investigation and completion of objectives 1 and 2 and concludes with addressing 3, leaving 4 for future work.

1.3 Report Structure

The background of the field, both in argumentation and online behaviour, is discussed in Chapter 2. Chapter 3 details the preliminary work carried out to investigate the capabilities of current models when modelling social argumentation and analyses the results. Finally, Chapter 4 sums up the findings and outlines a plan for future work.

Chapter 2

Background

2.1 Rhetoric & Argumentation

Before examining what attempts have been made to model argument, it is important to identify how arguments are analysed and to define some important terms.

Rhetoric is the art of discourse (both spoken and written) and is employed to persuade, inform or motivate. Although often used in modern parlance as a derogatory term to mean hiding the truth behind propaganda, in antiquity the highest form of rhetoric was considered the ability to maintain a *"balance between eloquence and wise silence"* (Hutto, 2002). Argumentation is a commonly used rhetorical mode (other examples include description, narration and exposition) in which the rhetor attempts to sway the audience's point of view on a subject or galvanise them into action (or even prevent them from action, if that better suits the rhetor's purpose).

2.1.1 Modes of Persuasion

Aristotle, in his treatise on rhetoric, described three "persuasive modes" that can be employed in an attempt to sway an audience, namely through the speech (*logos*), through the character (*ethos*) and through the emotions of the audience (*pathos*) (Kennedy, 1991). These modes can be applied individually but can also be applied in conjunction with one another.

At the core of many theoretical argument frameworks (see Section 2.1.3) is the examination of *logos*, an appeal to logic or reason. This is the method by which one might rationalise a position, often backing it up with evidence or statistics. It is important to note that, when enacting *logos*, it is not strictly necessary for the logic to be sound, or the evidence provided to be factual – it can be warped to fit a particular purpose, or even outright fabricated (however, this will usually invoke another of the modes described below). The key element is that it appears to be reasonable and thus, appeals to an audience's sense of reason (Kennedy, 1991; Braet, 1992).

Ethos is an appeal a person's character or sense of ethics and morals. This can be used in an attempt to strengthen the position of the rhetor's argument or to weaken their opponent's position. For example, if a rhetor can state that they are an expert in the field that they are debating then it is likely their audience will lend their argument more credence than if they were a novice. This specific case is known as an argument from authority, or *argumentum ab auctoritate* (Kennedy, 1991; Braet, 1992). Similarly, an argument can be made that attacks an opponents position indirectly, by attacking their credentials rather than refuting their claims (an *ad hominem* argument). Although such an argument is not logically sound (and constitutes a fallacy), it is still often used in practice and in certain circumstances is a viable (and often effective, if somewhat underhand) means of persuading an audience (Budzynska and Reed, 2012).

Finally, *pathos* is an appeal to emotion, whereby an attempt is made to evoke a particular feeling in an audience in the hope that this will influence their opinion on a position. This can be done in both positive and negative terms. For example, flattering an audience, or promising them a boon, can shift them towards accepting a particular course of action. On the other hand, threatening them with the potentially undesirable consequences of their actions can cause them to reconsider even if these consequences are unlikely or, indeed, impossible. A classic example is the appeal to fear (*argumentum ad metum*) (Kennedy, 1991; Braet, 1992).

2.1.2 Dialectic & Eristic Argumentation

A dialectic argument takes the form of two or more parties engaged in rational discourse with the aim of either discovering the particular truth behind a matter, or formulating a solution or resolution for a set of circumstances (Bloom, 1991, p. 454a). For example, an academic presenting their findings and rationalising that they are indeed valid, given the rigorous methodology they have used and the weight of evidence this has provided is an example of a dialectic argument. Likewise, a peer reviewer that disagrees with the findings by pointing out a specific flaw in the experimental methodology and explaining how this should be resolved, is another example.

By contrast, an eristic argument is an argument in which there is no clear resolution in the minds of the participants – they are quarrelling for its own sake, either for catharsis or with the aim of being seen to "win", either in the eyes of their opponent or, more usually, in the eyes of any spectators present (Jørgensen, 1998). This makes the role of audience an important feature to consider: when an individual responds to a post on the social web their post is often seen not just by the author of the post they reply to, but by many other users as well. In fact, many posts may be directed at this wider audience to seek approval, voice dissent, or provoke other emotions (Berland and Forte, 2010).

Many theoretical models of argumentation are based on the assumption of a dialectic argument. However, in social media there is a clear proliferation of eristic argumentation (Sood et al., 2012). Interestingly however, Schneider et al. downplay the value of modelling arguments that can be recognised as eristic. They argue that, as an eristic argument has no value other than

cathartic release, participants are free to *"sling propositions that they would not commit to under other circumstances"* and, as such, should not be retained (Schneider et al., 2014).

2.1.3 Models of Argumentation

There are many different models and frameworks used to capture particular aspects of argumentation. These aspects include notions of trust(Wigmore, 1913, p. 752), focus on argument topic or chronology (Klein, 2010) and the ability to demonstrate support for or refutation of other points in the structure (Dung, 1995). Some examples are discussed below, with respect to their technical structure, their influence in the field and their practical applications.

Toulmin developed his model from the school of philosophy in the 1950s as a means of demonstrating an approach to practical (rather than theoretical) argumentation, by attempting to show the internal structure (and thus, consistency) of an argument (Toulmin, 1958). The general form of Toulmin's argument, shown in figure 2.1, follows the structure of a *claim*, or conclusion, that is backed up with generally agreed upon facts (the *data*). The *claim* can be *qualified* ("definitely", "maybe", "probably", etc.) and any potential *rebuttals* accounted for. Then, key to the Toulmin model, the *claim* and *data* are connected using either an implicit or explicit *warrant*, or justification – this can then be supported by a particular *backing* (Verheij, 2005, p. 347-350). A specific example can be seen in figure 2.2, which shows an argument reasoning that Alice is a British citizen. Toulmin's model has been a particularly influential piece of work and has had an impact of decades of argumentation research in fields as far ranging as law, rhetoric and education (Newman and Marshall, 1992, p. 8-10; Schneider et al., 2013, p. 5, 12). However, there has been discussion as to the effectiveness of different aspects of the framework. In its favour, the means of explicitly stating the connecting warrant (and associated backing) can improve cross domain discourse. On the other hand, because models themselves are focused towards internal structure, there is no criteria for modelling overall structure (such as a group of arguments that refutes or support one another's claims). There is also no concept of resolving an argument (for example, on the grounds of logic or value); although this may have been by design, it negates the possibility of evaluating the strength of a given argument (Newman and Marshall, 1992, p. 349-350; Verheij, 2005, p. 5, 12). Among other applications, the Toulmin model has been incorporated into the Argument Model Ontology[1], an OWL ontology to allow classification of academic arguments. This is used in conjunction with CiTO, an ontology for factually and rhetorically categorising citations (Peroni and Shotton, 2012, p. 8).

[1] http://www.essepuntato.it/2011/02/argumentmodel

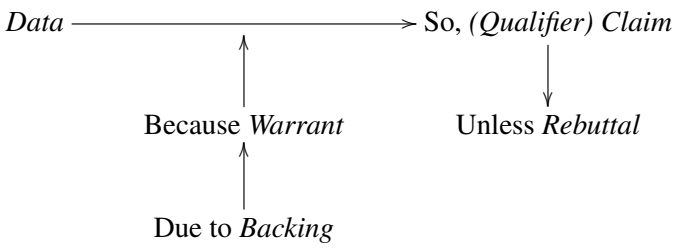


Figure 2.1: General form of Toulmin's diagram (Toulmin, 1958, p. 104)

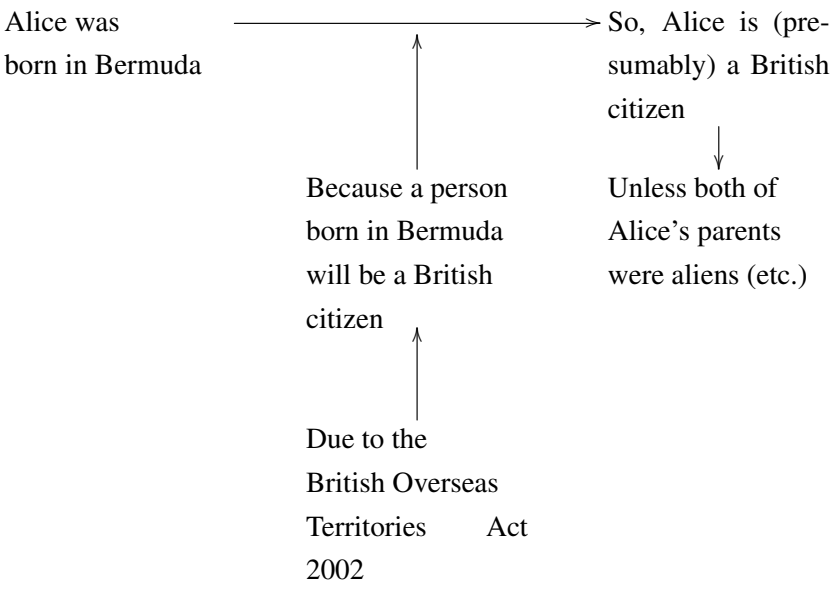


Figure 2.2: Example usage of the diagram, adapted from (Toulmin, 1958, p. 105)

Information-Based Issue System (IBIS) models are a particular type of dialectic process originally designed to aid in solving so-called "wicked problems" (Kunz and Rittel, 1970) – problems of social policy to which there is no clear definition, methodology or even end-goal (Rittel and Webber, 1973). IBIS models are represented as trees, made up of four different types of node. Firstly, *Issues* represent the problems that need to be solved, or questions that must be answered. Generally, there is one "root" *Issue* to be deliberated, but other sub-*Issues* can be created as necessary during the reasoning process. *Ideas* are proposed solutions or answers to these *Issues*, and each *Idea* can then be weighted positively or negatively using *Arguments For* and *Arguments Against*. IBIS models have seen wide usage in the field of design rationale and cognitive ergonomics where the assimilation of collective knowledge is required to solve problems (Conklin and Begeman, 1987; Aurisicchio and Bracewell, 2013). Because of its dialectic context, the application of IBIS models is ideal when two or more parties are trying to resolve a complex problem, especially if they have differing (or even opposing) stakes. As might be expected, there are many IBIS-like systems used in system-design and knowledge aggregation. Delibatorium[2]

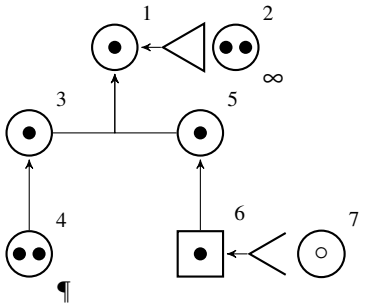
[2] http://deliberatorium.mit.edu

is a tool, developed by Klein (2010) of MIT, that uses an IBIS approach to solving challenging problems such as *"Is carbon offsetting a good idea?"*. The IBIS approach is invoked to aid the collaboration of large amounts of people separated across space and time by preserving a topic-centric (rather than time-centric) structure. IBIS structures have also been included in an extension to the SIOC ontology devoted to representing argumentation[3]. This ontology uses the IBIS notation of *Issues* and *Ideas* to formalise the process of solving a problem over social media (Lange et al., 2008).

"Wigmore's chart", conceived in 1913, is a means of recording argumentation originally devised for use in legal trials. The chart models the chain of interactions between competing arguments from both participants and can be used to evaluate the overall conclusion that should be drawn (Wigmore, 1913, p. 751). It takes the form of a directed graph where each node represents a particular fact. The shape of each node relates to the nature of the assertion; squares represent testimony given under oath; a triangle represent an explanation of or support for the node it "points" to; an open angle refutes the argument it points to and all other assertions (such as claims, physical evidence or related legal statutes) are represented by circles. These can additionally be marked to denote arguments by the defence or prosecution, but are not discussed here for clarity (Chalamish et al., 2011, 2013). Symbols relate further information about the nature of these assertions: an infinity symbol (∞) states that a node denotes sensory evidence that may be (re)produced in court; a pilcrow (¶) denotes an assertion that can be taken as fact with no further evidence (such as a precedence case); a lack of a symbol shows that the claim is implied from further reasoning in the graph. In addition, Wigmorean analysis can incorporate the notions of *strong belief* (••), *belief*(•) *doubt* (?) *disbelief* (○) and *strong disbelief* (○○) (Wigmore, 1913, p. 751-756; Goodwin and Fisher, 2000). Little is known about precisely how often this type of analysis is used manually, although it is thought that it is carried out in courthouses around the world (Chalamish et al., 2011). However, efforts are being made to automate the process by parsing the natural language propositions made in court and transforming these into a Wigmore diagram to aid judges, barristers and juries in their deliberations (Chalamish et al., 2013).

Similar to Wigmore's method, Dung's framework (which uses the format of set theory) focuses on the aspect of arguments attacking, (implicitly) supporting and, ultimately, defeating one another (Dung, 1995). Dung defines an *Argument Framework* as a pair such that $AF = \langle AR, attacks \rangle$ where AR is a set of arguments $\{a_1, a_2, ..., a_n\}$ and $attacks$ is a binary relation such that $attacks \subseteq AR \times AR$. $attacks$ describes which arguments are "defeated" by one another: for example, if a_1 is the argument "Alice is not a British citizen" and a_2 is the argument "Alice has a British passport" then $(a_2, a_1) \in attacks$. The set of *conflict free* arguments is a maximal set of arguments that do not attack each other. An argument a_1 is *acceptable* with regard to a set of arguments S if there is no argument a_2 that attacks a_1 that is not itself attacked by an argument in S. A set of arguments is *admissible* if each argument is considered *acceptable* with respect to the set. The maximal *admissible* set is known as a *preferred extension* (Schneider et al., 2013). Bench-Capon (2002, 2003) has extended this framework to incorporate the idea of "value" or

[3] http://rdfs.org/sioc/argument



1 Alice is a British citizen

2 Alice has a British passport

3 A person born in a British territory will be a British citizen

4 British Overseas Territories Act 2002

5 Alice was born in Bermuda

6 Alice's parents testify that she was born in Bermuda

7 Alice's parents' testimony could be biased in her favour

Figure 2.3: Example Wigmore graph

principle to arguments. When circumstances arise such that two possible resolutions to a dispute are equally (logically) valid, different audiences will have differing preferences based on the principles they feel are most important. For example, say that two solutions for combating crime are put forward: reading the general public's private correspondence or an expensive social program of education and rehabilitation. If each has been proven to be equally effective, audiences that value minimisation of cost may favour the former whereas audiences that value individual privacy might choose the latter.

The Argument Interchange Format (Chesñevar et al., 2006), developed as part of the Argument Web project (which is a concerted effort to map out a connected world wide web of argument), is an important modern model for the representation of web-based arguments. This model is described in further detail in Section 2.3.2.

2.2 Online Interactions

2.2.1 Social Media

Kaplan and Haenlein (2010) classify six distinct categories of social media: collaborative projects, blogs, content communities, social networking sites, virtual game worlds and virtual social worlds. These different classifications are detailed below.

Collaborative projects allow many different users to create, maintain and often discuss content. This category includes sites such as the online encyclopedia*Wikipedia*[4], which allow users to

[4]https://en.wikipedia.org/

write and edit articles and *Urban Dictionary*[5], a user generated dictionary of slang and internet culture.

Kaplan and Haenlein compare blogs (web-logs) to personal websites, in that they allow users to post information about the subject of their choice – these posts are often timestamped and presented reverse-chronologically. *Wordpress*[6] and *Blogger*[7] are two social media sites specialised for this purpose. "Micro"-blogging sites that pose limits on the amount of content that can be shared in a single post, such as *Twitter*[8], also fall into this category.

Content communities revolve around the concept of publishing (and ultimately sharing) different forms of media. These include sites for publishing video (such as *Vimeo*[9]), images (such as *Flickr*[10]), audio (such as *SoundCloud*[11]) and many other different types of media.

Social networking sites allow users to create a profile detailing information about themselves (such as home town, or music preferences) and then connect their profiles with the profiles of others on the site. Examples include *Facebook*[12] and *Google+*[13].

Virtual game worlds (such as *World of Warcraft*[14]) encompass online games in which a user controls a digital avatar to accomplish certain tasks (such as slaying a virtual dragon, or defeating another player's avatar). Similarly, virtual social worlds (such as *Second Life*[15]) encompass virtual spaces in which users have an avatar, but there is no specified aim or end-goal – the medium exists solely to facilitate social interaction. In this work, less focus is afforded to these latter two areas of the social web due to the tendency for participants to be playing a particular role, rather than their "real" self, which would affect their engagement in a discussion.

2.2.2 Anti-Social Behaviour on the Web

Anti-social behaviour is a growing problem on the social web, and often arises from debates or discussions that get out of hand (Suler and Phillips, 1998; Davis, 2002; Sood et al., 2012). This behaviour can arise from simple misunderstandings due to the difficulty in conveying tone through text, or as a deliberate act by individuals lashing out at other participants in a discussion. Incidents include flaming, in which a user simply hurls emotional abuse (Konijn et al., 2008, p. 13); spamming, in which a user floods the medium with content, often unrelated to the topic in hand, in the hope of drowning out other participants or as a means of advertising a commercial product (Krause et al., 2008); trolling, in which a user posts seemingly innocuous but

[5] http://urbandictionary.com/
[6] http://wordpress.com/
[7] http://blogger.com
[8] http://twitter.com/
[9] http://vimeo.com/
[10] http://flickr.com/
[11] http://soundcloud.com/
[12] http://facebook.com
[13] http://plus.google.com/
[14] http://battle.net/wow/
[15] http://secondlife.com

deliberately fallacious argument to provoke other members of the group into becoming outraged (although there is debate as to whether this term refers to the bridge-dwelling monster of myths, or the fishing term for dangling a baited line behind a boat) (Herring et al., 2002); and much more serious incidents of directed threats and stalking (Spitzberg and Hoobler, 2002).

As a result, there is a concerted research effort into the best way to tackle these issues before they cause serious harm to individuals, or the field as a whole. Suler and Phillips (1998) discuss a wide variety of approaches (specifically in regard to the virtual social world *The Palace*[16], but these could be applied to other online spaces as well). The simplest solution is to moderate users' interactions and dispense warnings, "mutes" (where a user may observe, but not contribute) or, in extreme cases, bans as and when the situation warrants. Needless to say however, this approach does not scale when considering the social web.

A different approach is to allow the community a degree of self-moderation. Reputation systems, for example, allow users within a community to assign "votes" to a particular account, or post, to show its trustworthiness. This allows new users to make judgements on whether to take a comment seriously, for example, or to purchase something from a particular seller in an online auction (Resnick et al., 2000; Anderson et al., 2012). However, this can also lead to a feedback loop in which communities become self-reinforcing; if users always vote for posts of similar sentiment (or against those that disagree), then gradually these sentiments will become dominant. Over time only users who hold these views will contribute to the site (further reinforcing the disparity) and the community as a whole will stagnate.

In another example of direct self-moderation, the popular online game *League of Legends*[17] implements a "tribunal" system in which players that are reported for poor behaviour in matches (such as verbally abusing team-mates), are judged by their peers. These peers can examine evidence such as chat logs and scores, then decided whether to "pardon" or "punish" the offending player (Hodson, 2013; Kou and Nardi, 2013).

A more covert attempt to manipulate users' behaviour can be found in certain implementations of human-computer interaction design. HCI can be leveraged to "trick" users into performing (or not performing) an action desirable to the designer. These so-called "malicious interfaces" (Conti and Sobiesk, 2010) are often used to trick users into spending time or money that they otherwise would not (for example, advertising banners that suddenly cover page content). In 2008, YouTube temporarily added an "Audio Preview" button to its comment system that would read aloud what the user intended to post. This was placed in the previous place of the "post" button (which had been moved further to the right), such that a user was likely to unintentionally preview their comment before posting it (Munroe, 2008).

[16] http://thepalace.com
[17] http://leagueoflegends.com

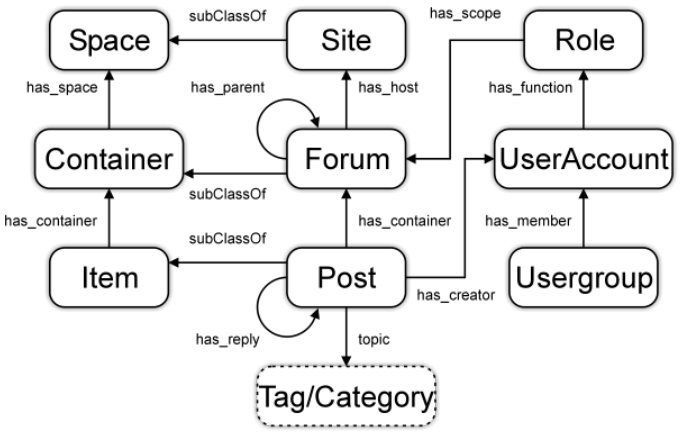


Figure 2.4: An overview of the SIOC Core Ontology[18]

2.3 Machine-Readable Argumentation

2.3.1 Social Interlinked Online Communities

The Social Interlinked Online Communities project (SIOC) (Breslin et al., 2006) aims to enable the cross-platform, cross-service representation of data from the social web. SIOC allows for semantic representations of Sites, which hold Forums, which contain Posts, authored by the owner of a UserAcount. This structure is shown in Figure 2.4. SIOC is often used in conjunction with the Friend of a Friend (FOAF) project, to show how individuals map to their online personas.

While an extension to SIOC, for the purposes of capturing and representing argumentation, does exist (Lange et al., 2008), it is based on the Issue Based Information System (IBIS) principals of modelling an argument as an issue that needs to be solved, with users suggesting ideas, then providing arguments for or arguments against these ideas. While this approach is highly useful when dealing with arguments centred around deliberation, and to a lesser extend criticism or inquiry, they are not as suitable when modelling negotiations or eristic arguments.

2.3.2 The Argument Interchange Format

Created as part of the Argument Web project (Rahwan et al., 2007a), which aims to link the concepts of natural language argumentation with abstract mathematical modelling (including capturing *"linguistically sophisticated maneuvers"* such as personal attack or faulty reasoning in the name of rhetoric (Bex et al., 2013)), the Argument Interchange Format (AIF) is a framework for representing argumentation as a directed graph (Chesñevar et al., 2006). While the ontology itself is primarily a natural language description of the format, there are several specifications for specific implementations in formal languages such as RDF or OWL (Rahwan et al.,

[18] http://sioc-project.org/ontology

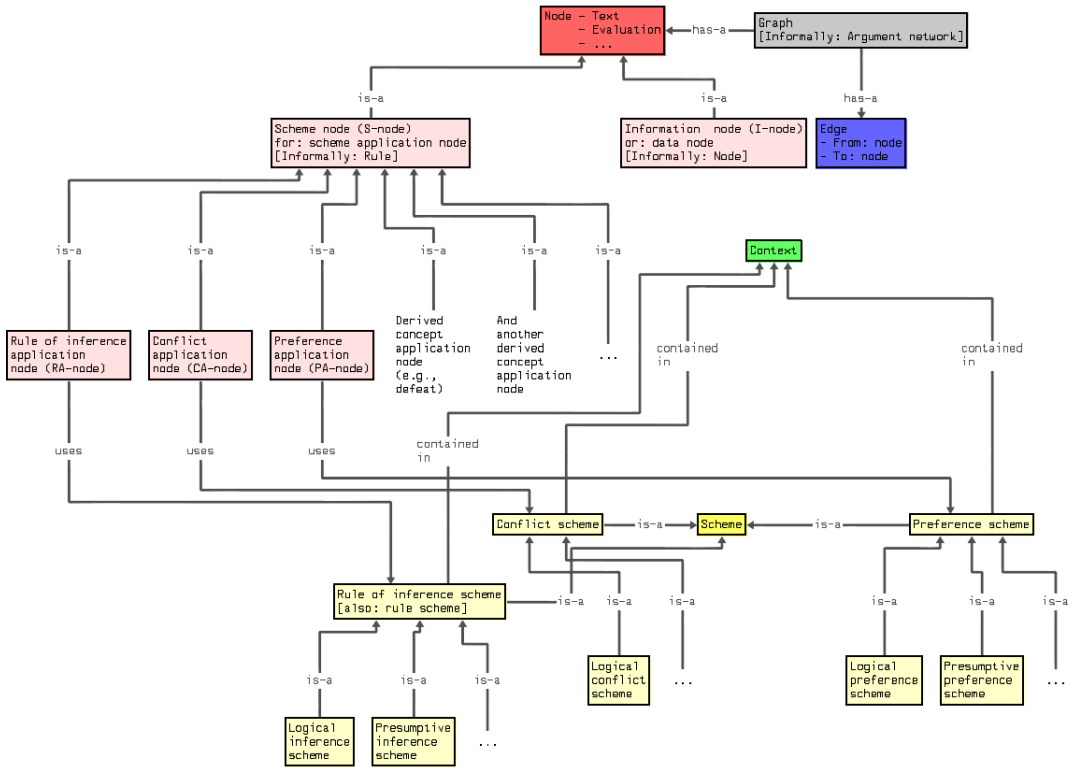


Figure 2.5: An overview of the AIF Ontology[21]

2007b; Rahwan and Banihashemi, 2008). A representation of this structure and some of its instantiations has been implemented in the Argument Interchange Framework database (AIFdb)[19]. Services which use argumentation data can draw on this repository to insert data to or query the Argument Web (Lawrence et al., 2012). One such service is ArguBlogging[20], a tool developed by Snaith et al. (2012) that interfaces with the Argument Web (and AIFdb). Users can highlight portions of text in their web-browser and explicitly state their agreement or disagreement and their reasoning behind it. This is then added to AIFdb and can be optionally posted to the user's blog. Other ArguBlogging users can respond to these arguments, further expanding the Argument Web.

At its highest level, the AIF can be conceptually divided into an "upper" ontology and a "forms" ontology. The upper ontology consists of the building blocks or "nodes" of an argument. The data, claims and conclusions are modelled by I-Nodes. Relationships between I-Nodes are represented by S-Nodes: these are subdivided into nodes that denote an inference (RA-Nodes), a conflict (CA-Nodes) or a preference (PA-Nodes). An overview of this ontology is shown in Figure 2.5. The forms ontology consists of a set of scheme descriptions which can be applied to nodes to give further context (for example, denoting an RA-Node as an argument from expert opinion).

[19] http://aifdb.org/
[20] http://argublogging.com/

In their work on an extension to the AIF, dubbed AIF+, Reed et al. built on the work of O'Keefe (1992, p. 79-82) to differentiate between two separate notions of argumentation: the first, which they term argument$_1$, is a logically constructed set of claims and evidence used to back these claims (or attack other claims); for example, *"Alice put forward the argument that..."*. The second, termed argument$_2$, refers to a dialogue – the exchange of ideas and opinions between two or more people; for example: *"Alice and Bob were having an argument"* (Reed et al., 2008). A result of this work was to introduce a new set of nodes. Locutions (L-Nodes), a subset of I-Nodes, model a locutionary act (or utterance) in an argument$_2$. Transitions (TA-Nodes), a subset of S-Nodes, represent the flow of conversation between L-Nodes. Finally, YA-Nodes, also a subset of S-Nodes, represent the "illocutionary force" and serve to link the L-Nodes of the argument$_2$ to the I-Nodes of the argument$_1$ – in other words, linking what was *said* to what was *meant*.

2.3.3 Automatic Identification of Argumentation Features

Anand et al. (2011) have worked on a means of automatically detecting persuasive language in blog posts. Building on the work of Marwell and Schmitt (1967); Cialdini (2001) and Walton et al. (2008) they defined five major categories of persuasive tactics: "Outcomes", such as threats or bribery; "Generalisations", such as appealing to duty or morality; "External" factors, such as the actions of authority figures, celebrities, etc.; "Interpersonal", such as asking for "a favour for a friend" and "Other", such as logical argument or analogy. In an "oracle" study they compare a hypothetical system capable of perfectly detecting "tactics" with previous techniques for classifying persuasion (such as word- or topic-based systems) and find that systems incorporating tactics perform better.

Sood et al. (2012) examine the possibilities of aiding community management of social sites and encouraging new participants to contribute by automatically detecting posts of an abusive or insulting nature (so that they can be removed by a moderator, depending on the culture of the site in question). Using a multi-stage process that takes into account valence, relevance and bigrams, Sood et al. found they could detect insulting posts with excellent accuracy and could even distinguish insults directed at other authors from those directed at the topic at hand.

2.4 Field Review and Outstanding Issues

Schneider et al. (2013), after reviewing a collection of tools that implement some of the argument frameworks discussed in Section 2.1.3, have presented a selection of outstanding issues in the field. Firstly, by their very nature, the informal arguments that take place on social media are particularly difficult to represent with formal modelling systems. While this can be somewhat mitigated by the investment of human input to resolve, for example, by manually annotating

[21] http://arg.dundee.ac.uk/aif

implicit premises and unprecedented or informal argument structures (such as heckling or outright abuse), this leads to the second major area of work. Currently, substantial human input is required to transform a raw argument into a formal representation. Minimising the amount of human input required and maximising the capability to automate the modelling of argumentation is a key goal for future work in this field. Finally, there is the issue of maintaining rhetorical meaning after translation and/or annotation. Somehow, the context surrounding the argument (such as the medium in which it took place, premises that may be taken as read or the number and identity of the participants) must be preserved along with the argument, so that the meaning behind it is not lost when it is stored.

By conducting a short survey into argumentation on the social web by bringing together some of the tools currently available, these issues are identified and highlighted for the purpose of learning how to adapt and extend these tools when moving towards a resolution.

Chapter 3

Preliminary Work

To explore the different structures of argumentation on the social web and examine how effectively these are modelled using current methods, a preliminary experiment was carried out. The approach described here brings together the AIF and SIOC by implicitly linking the notion of a Post with that of a Locution.

3.1 Methodology

3.1.1 Data Collection

To examine how arguments evolve across different communities on the web, and how these can be modelled, a single topic was chosen to be examined across different social web services. To ensure the stimulation of debate, the selected post needed to be publicly accessible, contain a controversial topic and have a large number of respondents. The Oct. 2013 United States government shutdown caused by Congress's failure to agree on a budget, and the following condemnation this received from the presidency, was a suitable match for these requirements.

This topic was then tracked across three of Kaplan's social media categories: YouTube, a content creation site where users can create and upload videos, or playlists of videos; Twitter, a microblogging service that allows users to publish messages of up to one-hundred and forty characters; and Facebook, a social network, that allows users to create a network of "friends" and share text or images. The former account is managed by the White House press office; the latter two are Barack Obama's "official" profiles (though managed by a third party).

The first post, initially posted on 8 Oct. 2013 from the White House's YouTube channel[1], is a 14m 40s video recording of Obama delivering a statement to press from the West Wing of the White House, condemning the shutdown. The second post, initially posted on 15 Oct. 2013 from

[1] https://www.youtube.com/watch?v=7LwoudGfug0

Table 3.1: Aspects of raw data from social media APIs capable of being modelled using the AIF or SIOC ontologies

Features present in social media APIs	**Represented in:**	
	AIF	**SIOC**
Locution (content)	✓	✓
Illocution (premises/conclusions)	✓	
Argumentation structure (attacks/support)	✓	
Author	✓	✓
Avatar		✓
Replies	✓	✓
Creation Date	✓	✓
Reputation (e.g. "Likes")		
Location		
User "Type" (i.e. individual/business/etc.)		

Obama's official Twitter account[2], reads: *"This is unacceptable. Tell Tea Party Republicans to stop holding our economy hostage: http://OFA.BO/qNmA3Y"*. The third post, also posted on 15 Oct. and posted from Obama's official Facebook account[3], reads: *"Tea Party Republicans in the House of Representatives forced a government shutdown, and now they're threatening an economic shutdown. This has gone on for too long. Tell them to #EndThisNow: http://OFA.BO/ACC7qB"* (however, this post captions an image displaying the message "End this now", rather than plain text).

The discussions surrounding these posts was acquired by collecting comments replying to each initial post, and those replying to subsequent posts in the discussion, with the use of the public YouTube, Twitter and Facebook APIs respectively. An extract of the raw data is shown in Appendix A.1. This data was translated to an RDF triple-store. The AIF was used to represent the text of each post or comment as an L-Node, within the argument$_2$. Transitions between posts, such as when a post is a direct reply of another post, or specifically mentions another post, are modelled using TA-Nodes. SIOC was used to model the data specific to the social media platform, such as which User created which Post and which Thread stores which Posts. This was used in conjunction with the DCTerms ontology, which held supplementary data such as timestamps.

In addition, when collecting this data it became apparent there was data that had no appropriate representation in either ontology (such as reputation systems, e.g. Facebook "Likes"); these omissions are shown in Table 3.1. Where available, this data was collected and represented

[2]https://twitter.com/BarackObama/status/390288744235823104
[3]https://www.facebook.com/photo.php?fbid=10151874920756749

using a novel ontology, named the Argumentation on the Social Web Ontology (ASWO). This is discussed in more detail in Section 3.3.2.

3.1.2 Data Sampling and Annotation

Because of the volume of the data produced over the course of the tracked event and the time-intensive nature of manually annotating the data, it was necessary to sample the data to a more manageable size before annotation could take place. As noted in Schneider et al. (2013), the reliable automation of this (and similar tasks) is another important area for future work in the field of argumentation research – this is discussed further in Section 3.1.4. To prevent information being lost when the dataset was scaled down, it was important to ensure that the sampled graph maintained properties (such as diameter and average path length) similar to those of the raw data. To maintain these characteristics, "forest fire" sampling (Leskovec et al., 2005; Leskovec and Faloutsos, 2006) was used to create a sub-graph that preserved the overall structure of the parent. The algorithm for forest fire sampling is as follows:

1. Choose a "forward burning probability" pf – in this instance a value of 0.7 was chosen based on the recommendation by Leskovec and Faloutsos (2006) for scaling down a larger graph
2. Choose a random starting node
3. Add this node to the sample graph. Select x nodes at random from all nodes linked to the chosen node, where x is a random number geometrically distributed with mean $pf/(1pf)$. If the selected node has fewer than x nodes, select all linked nodes, and return to step 2.
4. With each selected linked node, recursively repeat step 3 until the desired sample size has been reached.

Thirty posts from within the discussion (i.e. not including the original post by Obama) were selected using this method. This data was then manually annotated with the more abstracted argument_1 information. Specifically, from each L-Node, the implied I-Nodes were extracted and related together using the most appropriate S-Nodes. For example, Obama's original Twitter post (an L-Node) states: *"This is unacceptable. Tell Tea Party Republicans to stop holding our economy hostage: http://t.co/y8fPF8s3bG"* (the hyperlink leads to an Organising for Action page allowing users to automatically generate and send tweets to voice their displeasure at the shutdown). From this the following I-Nodes can be extracted: *"The Tea Party Republicans are holding the economy hostage"*, *"Holding the economy hostage is an unacceptable tactic"* and *"The Tea Party Republicans should stop holding the economy hostage"*. From this, it is easy to see that the single locution contains two premises and a conclusion (which therefore need to be joined using an RA-Node). This argument_1 can then be mapped to the specific locution by means of a YA-Node.

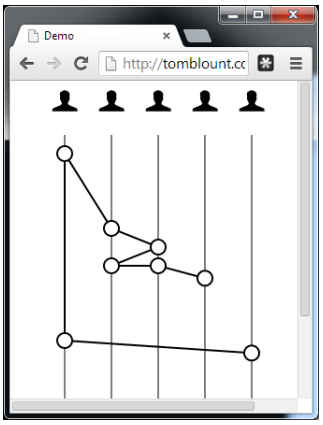


Figure 3.1: A simple example of the "Lifeline" visualisation

3.1.3 Data Visualisation

To aid the analysis of the constructed dataset, two visualisations were developed to display the output in a more human-readable form. Both are generated (based on the annotated RDF), interactive, javascript visualisations allowing users to elicit further information such as content of posts, URI, timestamp, etc. (by mouse-over) as required.

The first is a means of displaying the data represented with SIOC: each user has a "lifeline" associated with them (drawn in the order in which they contribute to the discussion). Each node represents a single post, and the height of each node is relative to the time it was posted. Lines are drawn between nodes when one post replies to, or makes mention of, another.

Figure 3.1 shows a simple argument structure visualised in this fashion. There are five users, contributing eight posts between them. The first post provokes a response, which then stimulates a rapid sub-argument between some of the users. A few minutes later, the first user posts again, and another user joins the debate.

This visualisation gives an overall impression of the speed with which people are contributing to the discussion. It also highlights any "pockets" of argumentation that form between groups of participants within the main dialogue.

The second visualisation focuses on the information contained within the AIF, and highlights the elicitation of information from locutions, showing how the dialogue of the argument_2 builds into the abstract argument_1 structure. The rightmost side shows the argument_2 – the L- and TA-Nodes – in blue. The leftmost side shows argument_1 – the information extracted from each locution, and any scheme applications – in red. YA-Nodes, shown in green, connect the two sides together.

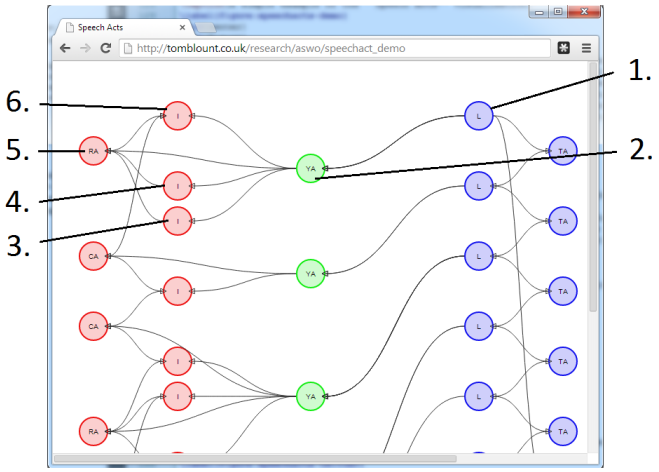


Figure 3.2: A simple example of the "Speech Acts" visualisation

Figure 3.2 shows the beginning of the same argument structure as visualised in Figure 3.1. Focusing on the first locution (1.), this can be seen to linked, via a YA-Node (2.), to a well-formed argument$_1$, consisting of two premises (3., 4.) which infer, through an RA-Node (5.), a conclusion (6.).

This visualisation highlights how an individual post can contain one or more abstract arguments, and how these then interact with one another. It also clearly shows how much information is contained within the dialogue that takes place; in other words, a complex debate will have a large number of I-Nodes per locution whereas an argument$_2$ of the form "Yes it is/No it isn't", *ad infinitum*, will have a much lower information content, and thus a much lower number of I-Nodes per locution.

An additional visualisation, aiming to combine the major features of the two visualisations already discussed to better display how the SIOC and AIF ontologies interact, is currently under development.

3.1.4 Limitations and Assumptions

The primary limitation is the necessity to manually annotate all of the data. This is a time consuming and subjective process, but as yet there is no way to circumvent this process and automatically extract premises and conclusions.

A further constraint is that only English-language sites are examined. There are, of course, many other social media services that cater to audiences of different languages, such as *Renren*[4] for China or *VKontakte*[5] for eastern Europe. However, this separation is mitigated by the fact that different languages (and different cultures) have their own rhetorical structures and argumentation schemes (Van Eemeren and Grootendorst, 2004, p. 21). As a result, attempting to

[4] http://renren.com/
[5] http://vk.com/

Table 3.2: Metrics (e.g. total number of posts, total number of users, posts per user, etc.) of discussions collected from YouTube, Twitter and Facebook

Metric	**YouTube**	**Twitter**	**Facebook**
Total number of posts	2719	137	9494
Total number of users	1255	33	6224
Average posts per user	2.17	4.15	1.53
Average words per post	26.74	15.91	40.12
Average characters per post	150.13	97.63	241.14
Time between first and last posts	101.68 days	0.57 days	90.83 days
Average time between posts	53m 52s	3m 2s	13m 47s

analyse multiple sites with different primary languages concurrently would distort any patterns that might emerge in the argument structure of the users.

In addition, there are limitations on the conclusions that can be drawn from such a small dataset when working with proverbial "big data". As such, the findings presented in Section 3.2 cannot be used to justify broad claims that this is how *all* arguments on a particular example of social media must be structured. However, these examples instead serve to demonstrate the important fact that different types of structures *can* evolve, and provide some examples of the argumentative and rhetorical tactics people use when arguing over social media and how the conjunction of the AIF and SIOC projects (as well as any future extensions) can be used in attempts to map them.

3.2 Results

An overview of the raw data collected from each platform is shown in Table 3.2. In total, the discussion generated by the Twitter post has slightly over one-hundred and thirty replies – in contrast, the YouTube comments total nearly three thousand posts, and the Facebook discussion has well over nine-thousand. Each platform sees the vast majority of posts contributed soon after the initial post. However, each has a "long tail" of responses that gradually decrease in frequency as time goes on. The discussion on Twitter in particular seems particularly ephemeral, with participants only contributing for a short time before moving onto other topics; while the Facebook and YouTube posts appear more "permanent", with users finding and contributing to them months later.

Table 3.3 shows the statistics collected after annotating the data with premises and conclusions, represented as AIF nodes. Given this data it can be seen that Twitter is the only sample that contains intra-thread links; that is, replies to other posts within the thread. While this may appear to suggest that the platform is used more fore debate than the others, it is possible this is

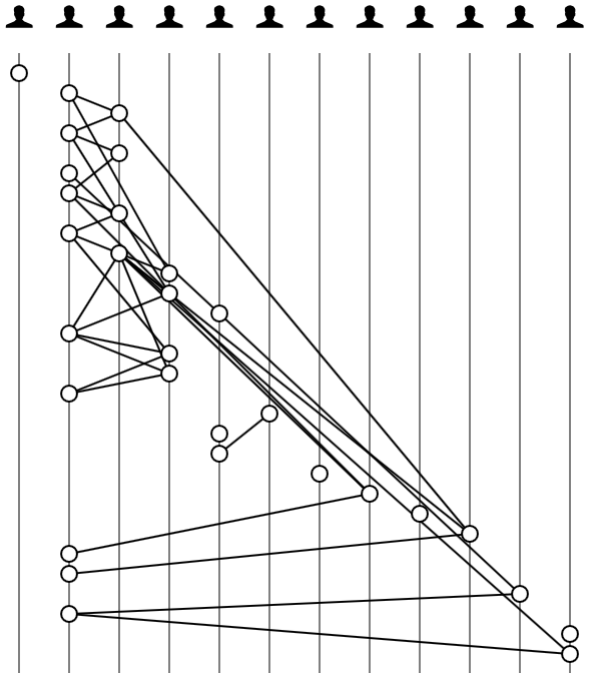


Figure 3.3: A subset of thirty Twitter replies, sampled using the forest-fire technique, visualised using Lifelines

Table 3.3: Count of different AIF nodes found in discussions collected from YouTube, Twitter and Facebook

Metric	**YouTube**	**Twitter**	**Facebook**
L-Nodes	30	30	30
TA-Nodes	0	20	0
YA-Nodes	31	30	41
I-Nodes	88	116	110
S-Nodes	13	30	26
L- to I-Node ratio	15:44	8:29	3:11

down to deficiencies in the APIs of the other platforms, which often do not accurately highlight replies. It can also be observed that the debates on Twitter and Facebook have a much higher information content than that of YouTube. The resulting structures are visualised in Figure 3.4, which shows a side-by-side comparison of the three different samples.

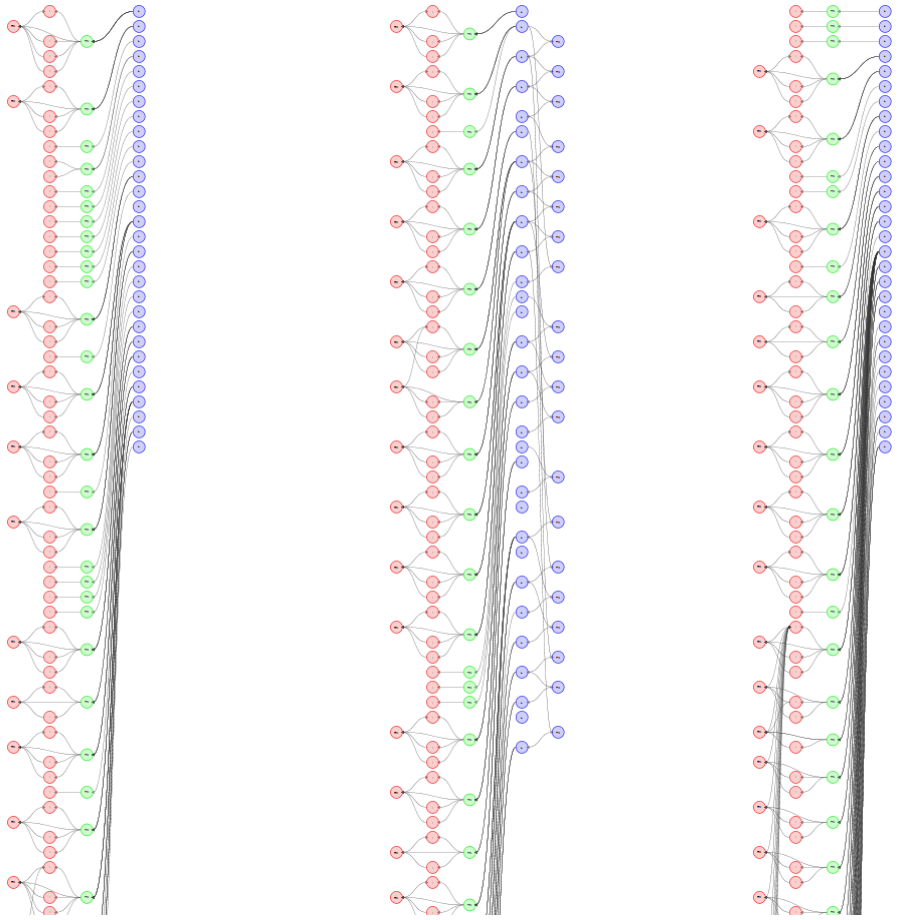


Figure 3.4: A side-by-side comparison of the emergent structures of discussions taken from YouTube (left), Twitter (centre) and Facebook (right), visualised as Speech Acts

3.3 Discussion

3.3.1 Analysis

On the surface, the sample of posts taken from Twitter and Facebook appear to have similar information content. However, upon manual inspection, it can be seen that this average is actually heavily skewed by one particular Facebook post that is thirteen paragraphs long (likely reused in full by multiple users for any argument on that topic) and contains a total of twenty six information nodes.

To highlight the overall information disparity take, for example, the tweet *"@BarackObama Stop expanding government, spying on Americans and driving up the deficit."*. This is an enthymeme – the literally derived I-Node acts as a conclusion, while the premises (that Obama is expanding government, spying on Americans and driving up the deficit and that to do so is a bad thing) are left implicit. In turn, contrast with the posts *"first"*, *"wow obama"* and *"lolollll i love this"* which contain very little information, either explicit or implicit. In addition, not all posts with a large amount of literal content have a comparatively large amount of information.

For example, posts such as *"Give DIRETIDE Give DIRETIDE Give DIRETIDE"* (repeated upwards of fifty times in a single post) show a desire to derail the discussion by flooding it with completely irrelevant information ("Diretide" was an event in the popular online game *Defence of the Ancients 2* that was cancelled, sparking uproar from the fanbase – this led to a number of social media platforms being flooded with this message).

3.3.2 Argumentation on the Social Web Ontology

In the process of capturing these arguments, there were many features both explicitly provided by social media APIs (see Table 3.1) and implicitly within the posts (see Table 3.4) that could not be adequately represented by the AIF or SIOC ontologies. As a result a new ontology named the Argumentation on the Social Web Ontology (ASWO) has been developed to be used in conjunction with the AIF and SIOC project to allow for the richer representation of social semantic data, particularly focused on how this affects arguments that take place on the social web.

Table 3.4: Proposed additions to the ASWO, for richer data representation

Name	Metric	Description	Example
Reputation	Positive or negative value indicating how many "votes" a user has received (of course, this does not normalise over different biomes – i.e., while some sites consider several hundred votes to be the norm, another might consider five to be particularly high)	User-driven reputation systems that allow users to vote positively (or negatively) on UserAccounts or Posts (such as Facebook "likes") can provide an insight into how well-liked, or trusted a particular individual or opinion is in a given community. A trusted user's post may carry more weight, even if it is fallacious, than a less trusted user's (Jøsang et al., 2007).	The most liked post in the Facebook thread (with 21 likes) was: *"I, for one, will remember this coming voting time. Enough is enough. This is about destroying Obama presidency so a republican will look good come next election. It is going to backfire you republicans and tea party people. How dumb can you be!! President Obama, please stick to your guns, in the end the republicans are going to loose as they deserve to!!"*. Despite the poor spelling and grammar, people were willing to show their support for the opinion behind it.
Credibility	Scale over the range $\pm 100\%$	As opposed to user-driven reputation systems, credibility is a measure of how "true" the content of a post is, and can be ascertained through analysis of the number of unique characters, presence of "emoticons" and the length of username (Gupta and Kumaraguru, 2012).	Consider the tweet *"@BarackObama they are not the ones "holding this country hostage". They are the ones truly representing the people of the United States!"*. Suggesting that the Tea Party Republicans represent the majority of the people of the US is a dubious claim.
Relevance	Scale over the range $\pm 100\%$	A measure of how closely related replies are to the initial post in terms of content. This could, among other things, aid the detection and flagging of "spam" posts aiming to advertise a person or business while contributing nothing to the argument.	The comment *"plz guys like my page plzz nobody likes my page i have challenged my friend that i will get more likes then him in one week and he has more likes then me"* has no relevance to the government shutdown, and instead is trying to "piggy-back" attention
Sentiment	Tagging system; including tags such as *"sarcasm"*, *"humour"*, *"abuse"*, etc.	Analysing the sentiment of a Post could provide additional context to the argument – for example, if a particular thread is mostly hostile, or if a particular UserAccount has a habit of being sarcastic.	Consider the comment *"gotta love tea baggers"* – it is unlikely that this was meant literally, and is instead sarcastically mocking the Tea Party Republicans
Mentions	SIOC UserAcount or Post	Some services (such as Twitter) distinguish between a post that directly replies to another from those that simply mention another post or user. In addition, they can "reply" to a particular UserAccount, rather than a post.	The tweet *"@thetropico @BarackObama YOU ARE A HORRIBLE PERSON!! Racist pig!"* replies directly to "thetropico" but also indirectly mentions Barack Obama

Chapter 4

Conclusions and Future Work

4.1 Conclusions

Through a combination of the background literature and the preliminary work carried out, the conclusion can be drawn that social argumentation on the web is currently not adequately modelled. As the survey in Chapter 3 shows, there is often a wide information disparity between arguments$_1$. However, there are many features of social and web-based argumentation that are not captured, such as spamming, abuse or use of humour.

Argumentation on the social web is made up of many complex parts; as a single thread can have thousands of individual participants, there will be many different goals and objectives present, ranging from information seeking to criticism to quarrelling. Simply eliminating one of these types (regardless of how little value it might contain) would remove valuable context from the argument$_1$ as a whole.

On top of these shortcomings, even the structure of relatively formal argumentation takes considerable human input to annotate (such as the extractions of explicit and implicit premises and conclusions). While this does provide an estimate of the information content of a post (and hence, an estimate of its value) and provides a record for further study in the Argument Web, other features such as relevance or sentiment (and others described in Section 3.3.2) may provide an equally good estimation of value automatically. This evaluation can then be used to determine the arguments$_1$ that time should be devoted to when annotating posts for the Argument Web.

4.2 Future Work

There are two obvious avenues of approach when considering the idea of reflecting a user's argument: namely whether it is the argument$_1$ that is presented to the author in a new light, or

the argument$_2$ as a whole. In the first case, this might be achieved by analysing the content, structure and rhetoric of a user's argument (such as the use of inflammatory language or insults) *as they write it* in the hope that they would restructure their argument to. The downside to this approach is that by "keeping score" of undesirable features could cause antisocial users to try to maximise their negative rating simply for fun. As a result, this approach is not considered further. A more practical approach then would involve analysing the features of each *previous* post in an argument$_2$ thread. In this way, a user could quickly scan a large debate on a particular topic and pick out key information, disregarding irrelevant or deliberately inflammatory posts. This is the approach that will be carried out in the following work packages.

4.2.1 Work Package 1: Experimentally Determining Value Weighting

Description: To design a system that can accurately describe which posts in an argument$_2$ are valuable, it must be determined what features of an argument$_1$ users of social media feel are valuable and, most importantly, in what proportions. For example, the relevance of a given post to the topic at hand might be considered more important than the reputation of the user that posted it. To determine this, an experiment will be carried out in which social media users are given a sample of argumentative posts (such as those sampled in the preliminary work in Section 3.1.1), asked to score a number of features of each presented argument (such as credibility, level of abuse, etc.) and then asked to rate the overall "value" of the argument, i.e. whether it advances the debate, and is worth reading or replying to.

Tasks:

1. Construct experimental framework
2. Seek ethical approval
3. Conduct experiment
4. Evaluate results
5. Produce conference paper

Outcome: An evaluation of the viability of the proposed system and a conference paper detailing the evaluation process and results.

Estimated Time: 2 months

4.2.2 Work Package 2: Creation of Value Annotation System

Description: Based on the outcome of work package 1 and the work of Jøsang et al. (2007), Gupta and Kumaraguru (2012), Sood et al. (2012) and Duan et al. (2010) a prototype system will be developed. This system will be capable of automatically weighting the "value" of a given argument$_1$ and modelling the result in an ontology that builds from and extends the AIF and SIOC.

Tasks:

1. Extend AIF/SIOC ontologies to account for desirable features
2. Create system to detect, annotate and weight these features
3. Test and debug system
4. Produce conference paper

Outcome: A system capable of annotating arguments taken from the social web with different levels of "value" based on different characteristics, and a conference paper detailing the construction process and use-cases of the system.

Estimated Time: 5 months

4.2.3 Gantt Chart

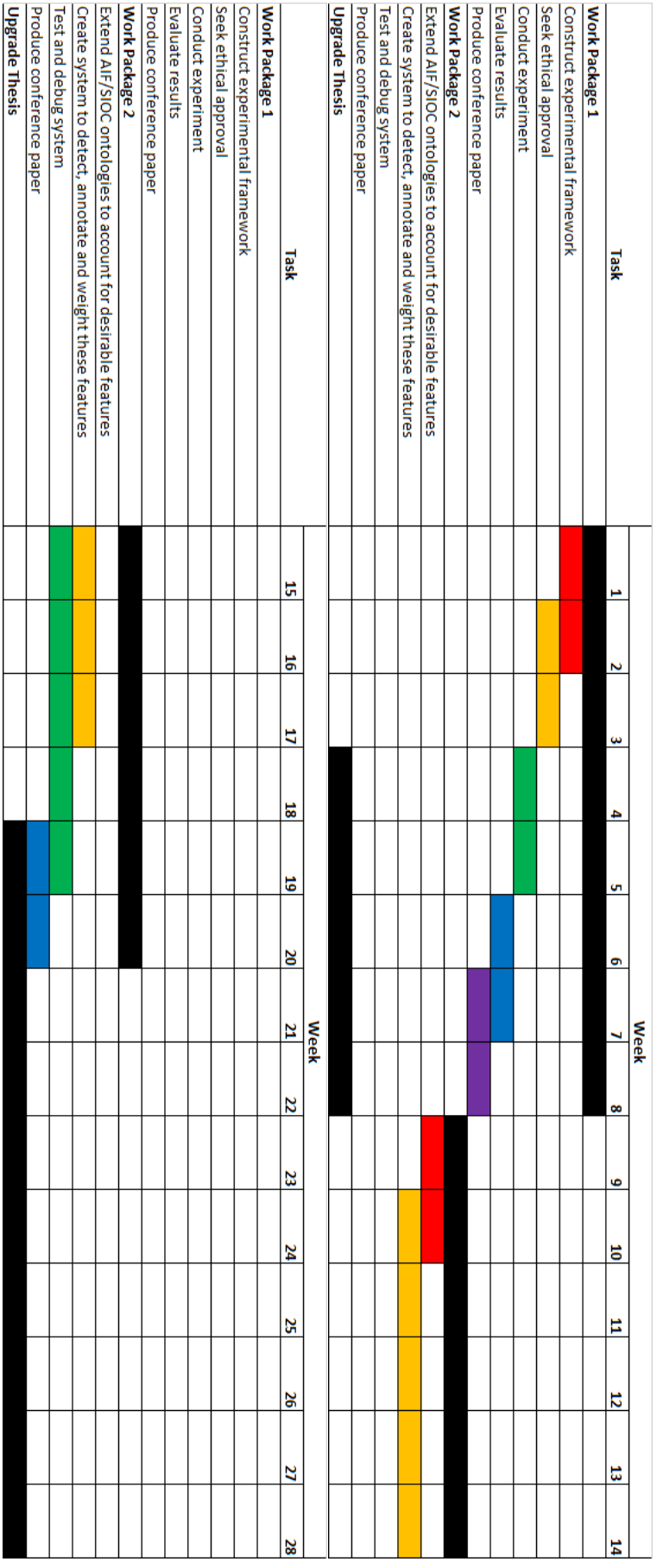


Figure 4.1: Gantt chart displaying the projected workflow across the next 9 months

References

Anand, P., King, J., Boyd-Graber, J., Wagner, E., Martell, C., Oard, D. W., and Resnik, P. (2011). Believe Me: We Can Do This! San Francisco, CA.

Anderson, A., Huttenlocher, D., Kleinberg, J., and Leskovec, J. (2012). Discovering value from community activity on focused question answering sites: a case study of stack overflow. In *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 850–858. ACM.

Aurisicchio, M. and Bracewell, R. (2013). Capturing an integrated design information space with a diagram-based approach. *Journal of Engineering Design*, 24(6):397–428.

Bench-Capon, T. J. (2002). Value based argumentation frameworks. pages 444–453.

Bench-Capon, T. J. (2003). Try to see it my way: Modelling persuasion in legal discourse. *Artificial Intelligence and Law*, 11(4):271—-287.

Bench-Capon, T. J. M. and Dunne, P. E. (2007). Argumentation in artificial intelligence. *Artificial intelligence*, 171(10):619–641.

Berland, L. K. and Forte, A. (2010). When students speak, who listens?: constructing audience in classroom argumentation. In *Proceedings of the 9th International Conference of the Learning Sciences-Volume 1*, pages 428–435. International Society of the Learning Sciences.

Bertea, S. (2004). Certainty, reasonableness and argumentation in law. *Argumentation*, 18(4):465–478.

Bex, F., Lawrence, J., Snaith, M., and Reed, C. (2013). Implementing the Argument Web. *Communications of the ACM*, 56(10):66–73.

Bloom, A. D. (1991). *Book V*. The Republic of Plato. Basic Books.

Braet, A. C. (1992). Ethos, pathos and logos in Aristotle's Rhetoric: A re-examination. *Argumentation*, 6(3):309–315.

Breslin, J. G., Decker, S., Harth, A., and Bojars, U. (2006). SIOC: an approach to connect web-based communities. *International Journal of Web Based Communities*, 2(2):133–142.

Budzynska, K. and Reed, C. (2012). The Structure of Ad Hominem Dialogues. pages 410–421.

Chalamish, M., Gabbay, D., and Schild, U. (2011). Intelligent Evaluation of Evidence Using Wigmore Diagrams. ICAIL '11, pages 61–65, New York, NY, USA. ACM.

Chalamish, M., Hazoom, M., and Schild, U. (2013). Semi-automatic Creation of Wigmore Diagrams. ICAIL '13, pages 181–185, New York, NY, USA. ACM.

Chesñevar, C., McGinnis, J., Modgil, S., Rahwan, I., Reed, C., Simari, G., South, M., Vreeswijk, G., and Willmott, S. (2006). Towards an argument interchange format. *Knowledge Engineering Review*, 21(4):293–316.

Cialdini, R. B. (2001). Influence: Science and practice. *Boston: Allyn & Bacon*.

Conklin, J. and Begeman, M. L. (1987). gIBIS: A hypertext tool for team design deliberation. pages 247–251.

Conti, G. and Sobiesk, E. (2010). Malicious interface design: exploiting the user. pages 271–280.

Davis, J. P. (2002). The experience of 'bad' behavior in online social spaces: A survey of online users. *Social Computing Group, Microsoft Research*.

Duan, Y., Jiang, L., Qin, T., Zhou, M., and Shum, H.-Y. (2010). An empirical study on learning to rank of tweets. In *Proceedings of the 23rd International Conference on Computational Linguistics*, pages 295–303. Association for Computational Linguistics.

Dung, P. M. (1995). On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games. *Artificial intelligence*, 77(2):321–357.

Goodwin, J. and Fisher, A. (2000). Wigmore's Chart Method. *Informal Logic*, 20(3).

Gupta, A. and Kumaraguru, P. (2012). Credibility ranking of tweets during high impact events. In *Proceedings of the 1st Workshop on Privacy and Security in Online Social Media*, page 2. ACM.

Hahn, U., Oaksford, M., and Corner, A. (2005). Circular arguments, begging the question and the formalization of argument strength. In *Proceedings of AMKLC'05, International Symposium on Adaptive Models of Knowledge, Language and Cognition*, pages 34–40.

Herring, S., Job-Sluder, K., Scheckler, R., and Barab, S. (2002). Searching for Safety Online: Managing "Trolling" in a Feminist Forum. *The Information Society*, 18(5):371–384.

Hodson, H. (2013). Moderate your language. *New Scientist*, 218(2912):18.

Hutto, D. (2002). Ancient Egyptian Rhetoric in the Old and Middle Kingdoms. *Rhetorica*, 20(3):213–233.

Johnstone, H. W. (1952). Philosophy and argumentum ad hominem. *The Journal of Philosophy*, pages 489–498.

Jørgensen, C. (1998). Public Debate – An Act of Hostility? *Argumentation*, 12(4):431–443.

Jøsang, A., Ismail, R., and Boyd, C. (2007). A survey of trust and reputation systems for online service provision. *Decision support systems*, 43(2):618–644.

Kaplan, A. M. and Haenlein, M. (2010). Users of the world, unite! The challenges and opportunities of Social Media. *Business horizons*, 53(1):59–68.

Kennedy, G. A. (1991). *Aristotle on rhetoric: a theory of civic discourses*. Oxford University Press.

Klein, M. (2010). Using metrics to enable large-scale deliberation. pages 103–233.

Konijn, E. A., Utz, S., Tanis, M., and Barnes, S. B. (2008). How technology affects human interaction. In Barnes, S. B., editor, *Mediated Interpersonal Communication*, chapter 1, pages 3—-13. Routledge New York.

Kou, Y. and Nardi, B. (2013). Regulating anti-social behavior on the Internet: The example of League of Legends.

Krause, B., Schmitz, C., Hotho, A., and Stumme, G. (2008). The anti-social tagger: detecting spam in social bookmarking systems. In *Proceedings of the 4th international workshop on Adversarial information retrieval on the web*, pages 61–68. ACM.

Kunz, W. and Rittel, H. W. (1970). *Issues as elements of information systems*, volume 131. Institute of Urban and Regional Development, University of California Berkeley, California.

Lange, C., Bojars, U., Groza, T., Breslin, J. G., and Handschuh, S. (2008). Expressing Argumentative Discussions in Social Media Sites. Karlsruhe, Germany.

Lawrence, J., Bex, F., Reed, C., and Snaith, M. (2012). AIFdb: Infrastructure for the Argument Web. pages 515–516.

Leskovec, J. and Faloutsos, C. (2006). Sampling from large graphs. In *Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 631–636. ACM.

Leskovec, J., Kleinberg, J., and Faloutsos, C. (2005). Graphs over time: densification laws, shrinking diameters and possible explanations. In *Proceedings of the eleventh ACM SIGKDD international conference on Knowledge discovery in data mining*, pages 177–187. ACM.

Marwell, G. and Schmitt, D. R. (1967). Dimensions of compliance-gaining behavior: An empirical analysis. *Sociometry*, pages 350–364.

Munroe, R. (2008). Youtube Audio Preview.

Newman, S. and Marshall, C. (1992). Pushing Toulmin too far: Learning from an argument representation scheme. *Xerox PARC Tech Rpt SSL-92-45*.

O'Keefe, D. J. (1992). *Readings in argumentation*, volume 11. Walter de Gruyter.

Peroni, S. and Shotton, D. (2012). FaBiO and CiTO: ontologies for describing bibliographic resources and citations. *Web Semantics: Science, Services and Agents on the World Wide Web*.

Pólos, L. and Hannan, M. T. (2002). Reasoning with partial knowledge. *Sociological methodology*, 32(1):133–181.

Rahwan, I. and Banihashemi, B. (2008). Arguments in OWL: A progress report. *COMMA*, 172:297–310.

Rahwan, I., Zablith, F., and Reed, C. (2007a). Laying the foundations for a world wide argument web. *Artificial intelligence*, 171(10):897–921.

Rahwan, I., Zablith, F., and Reed, C. (2007b). Towards large scale argumentation support on the semantic web. volume 7, pages 1446–1451.

Reed, C., Wells, S., Devereux, J., and Rowe, G. (2008). AIF+: Dialogue in the Argument Interchange Format. *FRONTIERS IN ARTIFICIAL INTELLIGENCE AND APPLICATIONS*, 172:311.

Resnick, P., Kuwabara, K., Zeckhauser, R., and Friedman, E. (2000). Reputation systems. *Communications of the ACM*, 43(12):45–48.

Rittel, H. W. and Webber, M. M. (1973). Dilemmas in a general theory of planning. *Policy sciences*, 4(2):155–169.

Rowe, M., Angeletou, S., and Alani, H. (2011). Predicting discussions on the social semantic web. In *The Semanic Web: Research and Applications*, pages 405–420. Springer.

Schneider, J., Davis, B., and Wyner, A. (2012). Dimensions of argumentation in social media. pages 21–25. Springer.

Schneider, J., Groza, T., and Passant, A. (2013). A review of argumentation for the Social Semantic Web. *Semantic Web*, 4(2):159–218.

Schneider, J., Villata, S., and Cabrio, E. (2014). Why did they post that argument? Communicative Intentions of Web 2.0 Arguments. In *Arguing on the Web 2.0*, Amsterdam. SINTELNET, European Network for Social Intelligence.

Schopenhauer, A. (2009). *The Art of Always Being Right*. Gibson Square.

Smith, T. (2009). The social media revolution. *International journal of market research*, 51(4):559–561.

Snaith, M., Bex, F., Lawrence, J., and Reed, C. (2012). Implementing ArguBlogging. pages 511–512.

Sood, S. O., Churchill, E. F., and Antin, J. (2012). Automatic identification of personal insults on social news sites. *Journal of the American Society for Information Science and Technology*, 63(2):270–285.

Spitzberg, B. H. and Hoobler, G. (2002). Cyberstalking and the technologies of interpersonal terrorism. *New Media & Society*, 4(1):71–92.

Suler, J. R. and Phillips, W. L. (1998). The bad boys of cyberspace: Deviant behavior in a multimedia chat community. *CyberPsychology & Behavior*, 1(3):275–294.

Toulmin, S. E. (1958). *The Uses of Argument*. University Press, Cambridge.

Van Eemeren, F. H. and Grootendorst, R. (2004). *A systematic theory of argumentation: The pragma-dialectical approach*, volume 14. Cambridge University Press.

Verheij, B. (2005). Evaluating arguments based on Toulmins scheme. *Argumentation*, 19(3):347–371.

Walton, D., Reed, C., and Macagno, F. (2008). *Argumentation schemes*. Cambridge University Press.

Wigmore, J. H. (1913). *The Principles of Judicial Proof As Given by Logic, Psychology, and General Experience and Illustrated in Judicial Trials*. Little, Brown and Co., Boston, MA, 1st edition.

Appendix A

Raw Data

A.1 API Results

This section shows the raw data received from the public APIs of YouTube, Twitter and Facebook. While broadly similar, each has its own structure and examples of unique content.

A.1.1 YouTube

Of particular note on the data received from the YouTube API is the links to related content, as well as the typing of some elements.


```
{
    "id": {
        "$t": "http://gdata.youtube.com/feeds/api/videos/7LwoudGfug0/comments/
    028p3HYaOlOwc3mEsWWbkA9R72LM4NOMtO7gsdpIMuI"
    },
    "published": {
        "$t": "2013-10-26T00:57:42.000Z"
    },
    "updated": {
        "$t": "2013-10-26T00:57:42.000Z"
    },
    "category": [
    {
        "scheme": "http://schemas.google.com/g/2005#kind",
        "term": "http://gdata.youtube.com/schemas/2007#comment"
    }
    ],
    "title": {
        "$t": "So by that logic ...",
        "type": "text"
    },
    "content": {
```

```
        "$t": "So by that logic people who have never been in accidents don't
    need car \ninsurance? Or isn't that what insurance is supposed to pay for,
    the \nunexpected? I'll admit Obamacare isn't a great system because it still
    \nrelies on private insurance, single payer systems in Europe consistently \
    nperform better than ours by almost every measure, cost effectiveness, \
    ninfant mortality, life expectancy. I don't agree with us maintaining a \
    ngigantic military, but I don't get to opt out from military spending.",
        "type": "text"
    },
    "link": [
     {
        "rel": "related",
        "type": "application/atom+xml",
        "href": "https://gdata.youtube.com/feeds/api/videos/7LwoudGfug0"
     },
     {
        "rel": "alternate",
        "type": "text/html",
        "href": "https://www.youtube.com/watch?v=7LwoudGfug0"
     },
     {
        "rel": "self",
        "type": "application/atom+xml",
        "href": "https://gdata.youtube.com/feeds/api/videos/7LwoudGfug0/comments/
    O28p3HYaOlOwc3mEsWWbkA9R72LM4NOMtO7gsdpIMuI"
     }
    ],
    "author": [
     {
        "name": {
                "$t": "jager290"
        },
        "uri": {
                "$t": "https://gdata.youtube.com/feeds/api/users/
    _3Tej5a09R2kR5knFWCBkQ"
        }
     }
    ],
    "yt$channelId": {
        "$t": "UC_3Tej5a09R2kR5knFWCBkQ"
    },
    "yt$replyCount": {
        "$t": 0
    },
    "yt$videoid": {
        "$t": "7LwoudGfug0"
    }
}
```


A.1.2 Twitter

The data received from Twitter is notable in that it includes a great deal of information about the user posting the tweet (such as number of followers, and even the colour of their profile) as well as the tweet itself.

```
{
    "contributors": null,
    "truncated": false,
    "text": "@BarackObama PLEASE FOLLOW ME PLEASE",
    "in_reply_to_status_id": 390288744235823104,
    "id": 390288961450835968,
    "favorite_count": 0,
    "source": "<a href=\"http://twitter.com/download/iphone\" rel=\"nofollow\">
    Twitter for iPhone</a>",
    "retweeted": false,
    "coordinates": null,
    "entities": {
        "symbols": [],
        "user_mentions": [
        {
            "id": 813286,
            "indices": [
             0,
             12
            ],
            "id_str": "813286",
            "screen_name": "BarackObama",
            "name": "Barack Obama"
        }
        ],
        "hashtags": [],
        "urls": []
    },
    "in_reply_to_screen_name": "BarackObama",
    "id_str": "390288961450835968",
    "retweet_count": 0,
    "in_reply_to_user_id": 813286,
    "favorited": false,
    "user": {
        "follow_request_sent": null,
        "profile_use_background_image": true,
        "default_profile_image": false,
        "id": 1650719012,
        "profile_background_image_url_https": "https://pbs.twimg.com/
    profile_background_images/431953559425282049/04og3Ngm.jpeg",
        "verified": false,
        "profile_text_color": "333333",
        "profile_image_url_https": "https://pbs.twimg.com/profile_images
    /470405045738274816/o_7QzOEy_normal.jpeg",
        "profile_sidebar_fill_color": "DDEEF6",
        "entities": {
            "description": {
                "urls": []
            }
        },
        "followers_count": 7928,
        "profile_sidebar_border_color": "FFFFFF",
        "id_str": "1650719012",
        "profile_background_color": "ACDED6",
        "listed_count": 34,
        "is_translation_enabled": false,
        "utc_offset": -14400,
        "statuses_count": 18175,
```

```
        "description": "I once saw michael clifford walk out of his tour bus and
    waVE TO US \u203a\u25e1\u0941\u2039",
        "friends_count": 5624,
        "location": "\u00bc \u261a 0/5",
        "profile_link_color": "000000",
        "profile_image_url": "http://pbs.twimg.com/profile_images
    /470405045738274816/o_7QzOEy_normal.jpeg",
        "following": null,
        "geo_enabled": false,
        "profile_banner_url": "https://pbs.twimg.com/profile_banners
    /1650719012/1400988277",
        "profile_background_image_url": "http://pbs.twimg.com/
    profile_background_images/431953559425282049/04og3Ngm.jpeg",
        "screen_name": "CINNAB0NOUIS",
        "lang": "en",
        "profile_background_tile": true,
        "favourites_count": 15104,
        "name": "PLEASE MICHAEL",
        "notifications": null,
        "url": null,
        "created_at": "Tue Aug 06 16:34:27 +0000 2013",
        "contributors_enabled": false,
        "time_zone": "Eastern Time (US & Canada)",
        "protected": false,
        "default_profile": false,
        "is_translator": false
    },
    "geo": null,
    "in_reply_to_user_id_str": "813286",
    "lang": "en",
    "created_at": "Wed Oct 16 01:31:54 +0000 2013",
    "in_reply_to_status_id_str": "390288744235823104",
    "place": null
}
```


A.1.3 Facebook

In contrast to the former two examples, Facebook returns fairly minimal data. However, like Twitter, this also includes some information about the user (in this case a “category”.


```
{
    "id": "10151874920756749_20645335",
    "can_remove": false,
    "created_time": "2013-10-15T15:03:22+0000",
    "from": {
       "category": "Local business",
       "category_list": [
          {
             "id": "175898962454294",
             "name": "Dog Training"
          },
          {
             "id": "188573091164316",
             "name": "Pet Store"
          }
       ],
```

```
        "name": "Dogs lovers \u10e6\u10e6",
        "id": "161928764002567"
    },
    "like_count": 0,
    "message": "plz guys like my page plzz nobody likes \ud83d\ude29 \ud83d\ude32
     \ud83d\ude1e \ud83d\ude35 \ud83d\ude30 \ud83d\ude12 \ud83d\ude0d \ud83d\
    ude24 \ud83d\ude1c \ud83d\ude1d \ud83d\ude0b \ud83d\ude18 \ud83d\ude1a \ud83d
    \ude37 \ud83d\ude33 \ud83d\ude03 \ud83d\ude06 my page i have challenged my
    friend that i will get more likes then him in one week and he has more likes
    then me \ud83d\ude29 \ud83d\ude32 \ud83d\ude1e \ud83d\ude35 \ud83d\ude30 \
    ud83d\ude12 \ud83d\ude0d \ud83d\ude24 \ud83d\ude1c \ud83d\ude1d \ud83d\ude0b
    \ud83d\ude18 \ud83d\ude1a \ud83d\ude37 \ud83d\ude33 \ud83d\ude03 \ud83d\ude06
    ",
    "user_likes": false
}
```

A.2 RDF Representation

This section shows how a particular post taken from Twitter can be broken down into premises and conclusions, and represented as RDF.

The extract below shows the structure of the post itself: its content (`aif:claimText`), its author (`sioc:has_creator`), its creation date (`dcterms:created`), etc.

```
<rdf:Description rdf:about="http://tomblount.co.uk/aswo#Post_390293189556715520">
    <rdf:type rdf:resource="http://www.arg.dundee.ac.uk/aif#L-node"/>
    <rdf:type rdf:resource="http://rdfs.org/sioc/ns#Post"/>
    <sioc:reply_of rdf:resource="http://tomblount.co.uk/aswo#
    Post_390292836153049088"/>
    <aswo:mention_of rdf:resource="http://tomblount.co.uk/aswo#
    Post_390288744235823104"/>
    <aswo:has_reputation rdf:datatype="http://www.w3.org/2001/XMLSchema#integer
    ">0</aswo:has_reputation>
    <dcterms:created rdf:datatype="http://www.w3.org/2001/XMLSchema#dateTime
    ">2013-10-16T01:48:42</dcterms:created>
    <aif:claimText>@kade6767 @BarackObama all presidents must work with Congress.
     Every budget they've submitted has been turned down.</aif:claimText>
    <sioc:has_creator rdf:resource="http://tomblount.co.uk/aswo#User_14772441"/>
</rdf:Description>
```

This extract shows the premises and conclusions taken from the post, each containing a claim (`aif:claimText`).

```
<rdf:Description rdf:about="http://tomblount.co.uk/aswo#
    I_Node_390293189556715520_P1">
    <aif:claimText>All presidents should work with congress</aif:claimText>
    <rdf:type rdf:resource="http://www.arg.dundee.ac.uk/aif#I-node"/>
</rdf:Description>

<rdf:Description rdf:about="http://tomblount.co.uk/aswo#
    I_Node_390293189556715520_P2">
    <aif:claimText>Turning down every budget constitutes not working with
    congress</aif:claimText>
    <rdf:type rdf:resource="http://www.arg.dundee.ac.uk/aif#I-node"/>
</rdf:Description>

<rdf:Description rdf:about="http://tomblount.co.uk/aswo#
    I_Node_390293189556715520_C1">
    <rdf:type rdf:resource="http://www.arg.dundee.ac.uk/aif#I-node"/>
    <aif:claimText>The president should accept a budget</aif:claimText>
</rdf:Description>
```

This extract shows the RA-Node, linking the premises (`aif:Premise`) and conclusion (`aif:Conclusion`).

```
<rdf:Description rdf:about="http://tomblount.co.uk/aswo#
    RA_Node_390293189556715520_1">
    <rdf:type rdf:resource="http://www.arg.dundee.ac.uk/aif#RA-node"/>
    <aif:Conclusion rdf:resource="http://tomblount.co.uk/aswo#
    I_Node_390293189556715520_C1"/>
    <aif:Premise rdf:resource="http://tomblount.co.uk/aswo#
    I_Node_390293189556715520_P2"/>
```

```
    <aif:Premise rdf:resource="http://tomblount.co.uk/aswo#
    I_Node_390293189556715520_P1"/>
</rdf:Description>
```

The final extract shows the YA-Node that links the components of the post. This links the premises, conclusion and RA-Node (`aif:IllocutionaryContent`) with the post itself (`aif:Locution`).

```
<rdf:Description rdf:about="http://tomblount.co.uk/aswo#
    YA_Node_390293189556715520_1">
    <aif:Locution rdf:resource="http://tomblount.co.uk/aswo#
    Post_390293189556715520"/>
    <rdf:type rdf:resource="http://www.arg.dundee.ac.uk/aif#YA-node"/>
    <aif:IllocutionaryContent rdf:resource="http://tomblount.co.uk/aswo#
    RA_Node_390293189556715520_1"/>
    <aif:IllocutionaryContent rdf:resource="http://tomblount.co.uk/aswo#
    I_Node_390293189556715520_P1"/>
    <rdf:type rdf:resource="http://www.arg.dundee.ac.uk/aif#Normal_Illocution"/>
    <aif:IllocutionaryContent rdf:resource="http://tomblount.co.uk/aswo#
    I_Node_390293189556715520_P2"/>
    <aif:IllocutionaryContent rdf:resource="http://tomblount.co.uk/aswo#
    I_Node_390293189556715520_C1"/>
</rdf:Description>
```