Lund University
Department of Political Science

STVK02
Tutor: Christian Fernández

The Securitization of Cyberspace

How the Web Was Won

Ola Hjalmarsson

Abstract

This essay explores how the American Government understands and characterizes cyberspace and its relation to security. Building on the theory of securitization, the essay seeks to understand and describe the speech-acts that the American Government under the Obama Administration engage in in order to securitize the cyber domain and related referent objects. To accomplish this, this essay takes both a conventional approach, using a qualitative-intensive method, and proposes a novel, quantitative-extensive method to analyse the prevalence of securitizing speech acts in a text corpus. The qualitative investigation demonstrates how securitizing actors engage in "hypersecuritization" by constructing an image of a threat capable of utilizing the networked nature of cyberspace to create destruction on a level that is comparable to previous disasters such as "Pearl Harbour" and "9/11". The results from the quantitative investigation support the notion that such speech-acts are representative of a broader tendency within the Department of Defense and the Department of State to engage in speech-acts aimed at presenting cyberspace as a domain filled with threats and in need of securitizing, but fails to provide the level of context that the qualitative investigation achieves.

Key words: Securitization, Cyberspace, Copenhagen School, Speech-act, Obama Administration
Words: 8260

Table of contents

1 Introduction

The advent of the internet has fundamentally changed the way the world communicates. Since its inception, the internet has grown into a vast network spanning the globe, allowing roughly a third of the worlds population to communicate within and between states and allowing access to an ever growing mountain of information (World Bank, 2012). The internet has facilitated communication, diplomacy, trade and the (mostly) free exchange of ideas between its exponentially growing user base of individuals, corporations, organizations and states, who increasingly incorporate this new communication structure into their respective infrastructures. But for all its advantages, the internet has also facilitated a rise in varying sorts of malicious activity. Viruses and other pieces of malicious code with a range of different functions are reguarly disseminated throughout the global network.

This has lead some actors to deem cyberspace a domain in need of military presence. In July 2012, President Barack Obama wrote an opinion piece in The Wall Street Journal, describing the threat of a cyber attack against the nation as "one of the most serious economic and national security challenges we face" (Wall Street Journal, 2012). His statements are symptomatic of a broader movement within the U.S. government to establish a military presence in the cyber domain that includes, but is not limited to, establishing "The United States Cyber Command" with the expressed goal of "planning, coordinating, integrating, synchronizing, and directing activities to operate and defend the Department of Defense information networks and when directed, conduct full-spectrum military cyberspace operations" (U.S. Strategic Command, 2012).

Because the cyber domain plays an increasingly important role in the lives of people around the globe, it is important to critically assess and understand the attempts that are made to define and confine aspects of this new domain of communication.

1.1 Purpose

This essay will attempt to describe and understand the process of securitization of cyberspace. As the internet will most likely continue to reach an even more substantial portion of humanity and thus become an even more integral part of our lives, it is important to examine the processes and actions that work towards defining and framing the discussion of security and its relation to internet. The increasingly frequent portrayal of internet-related threats in the media, the

creation by several states of military subdivisions tasked with establishing a military presence in cyberspace, and even the frequent inclusion of cyber attacks as a pop culture trope are all indications of a general move towards a conceptualization of cyberspace that to an increasing extent is characterized by the language of war.

In order to further the understanding of this process and its implications, it is the aim of this essay to explore the way the American Government under the Obama Administration has approached and characterized the cyber domain. This attempt will be informed by and conducted with the help of the theory of securitization that has been developed as part of the "Copenhagen School" of security. Using the traditional concepts and frameworks developed by its proponents, it is my intention to illuminate and analyze the role that the securitizing actors, and the speech-acts in which they engage, play in the securitization of the cyber domain. Because of its prominent role in the overall discourse on international security in general, and cyber security in particular, the focus of this essay will be limited to analyzing the role played by representatives of the American Government under the Obama Administration as securitizing actors. The guiding question to this investigation is *"How has the American Government under the Obama Administration understood and characterized the role of cyberspace and its relation to security?"*

A secondary aim of this essay is to explore the possibility and utility of a quantitative approach to the concept of the speech-act – a concept which in many interpretations is central to the process of securitization. The theory of securitization takes a largely constructivist approach to security that often focuses on the intensive study of the different mechanisms that enable the securitization of a referent object. The multitude of studies that have added to the understanding of security by means of qualitative inquiry are a testament to the value of this approach (i.e. Balzacq, 2011). Nevertheless, I believe that a quantitative investigation of speech-acts might also add to the understanding of securitization by virtue of its ability to take large amounts of information into account when attempting to characterize the prevalence of a specific phenomenon. It is my ambition to develop a method of inquiry that uses a quantitative approach to analyze speech-acts, apply it to the research problem and evaluate its strengths and weaknesses. In order to further the understanding of the problem itself, as well as provide a point of comparison of the strengths and weaknesses of the qualitative approach, I will also engage in a qualitative analysis using a more limited number of texts relevant to the problem.

1.2 Theory

Traditional theories concerning security often consider the concept to be a given – a more or less objective, if abstract, condition wherein the threat of the annihilation of a nation state is absent. This narrow and state-centric notion of security that to a great extent constituted the prevailing understanding of security

during the 20th century has since been challenged by attempts to introduce an alternative, broader understanding of security. In this proposed new approach, security should not be a concept that is exclusively applied to the state. Instead, argue its proponents, security should be applied to all facets of life where the well-being of individuals is threatened. But, as Ole Wæver points out, this understanding of security creates the potential problem of making security synonymous with all the problems that plague humanity, thus losing its utility as of a tool for describing and understanding a specific phenomenon (Wæver, 2007, p. 67f). Furthermore, it relies on the assumption that security can exist independently of the social processes that, in the constructivist view, gives rise to the concept (Wæver, 2007, p. 66).

Securitization theory, therefore, has arisen as a comprehensive alternative that attempts to bridge the gap between the old and too static and the new and too malleable notions of security by drawing on useful insights from both traditions – it is constructivist in its emphasis on social processes as the origin of security and in its view of security as not just relevant to matters pertaining to the state, while attempting to retain elements of the classical realist notion of security as relating to a broader, existential threat to a limited collective (Williams, 2003). In the understanding of security offered by securitization theory, "threats are not separable from the intersubjective representations in which communities come to know them" (Balzacq, 2011, loc. 214). In this view, there is no distinction made between a "real threat" and a "perceived threat", there is only an intersubjective understanding of a threat.

Within the theory of securitization a few concepts are key: *the securitizing actor* is the designation given to an actor who prompts the securitization; the *referent object* is the object that is deemed by the securitizing actor as being in need of securitization; *the audience* is the population that needs to be convinced of the vulnerability of the referent object, and the necessity of extraordinary measures in order to protect it, in order for the securitization to be successful (Balzacq, 2011, loc. 321). Ole Wæver and Barry Buzan offers the following as a definition of securitization: "the discursive process through which an intersubjective understanding is constructed within a political community to treat something as an existential threat to a valued referent object, and to enable a call for urgent and exceptional measures to deal with the treat" (Buzan & Wæver, 2003, p. 491).

One of the most important contentions of securitization theory is that security is intrinsically connected with a speech-act. The speech-act is the main mechanism through which security is constructed. The idea behind the speech-act is that certain acts of speech are actively changing reality through their very occurrence. In securitization theory, thus, security can only arise through a performative utterance by a securitizing actor (Strizel 2007, 360f). By referring to the urgency of action posed by an existential threat, a securitizing actor can transform an issue in to one of security (Buzan & Wæver, 2003, p. 71). "The word "security" is the act", elaborates Wæver, "the utterance is the primary reality" - there can be no security without the act of saying security (Wæver, 2007, p. 73f).

The theory of securitization seems to provide an excellent framework for understanding security as it relates to the cyber domain. Cyber issues, like economic and environmental issues, do not exist in a vacuum – more often than not they exist and intermingle on a global scale, taking little regard to the boarders of nations. Where theories that place too much importance on the actions of a single set of rigid actors (i.e. states) might have difficulty describing and explaining an increasingly fluid and interconnected world, securitization theory provides an understanding that is not contingent upon the centrality of states, while at the same time acknowledging their continued importance. Securitization strikes a balance between rigidity and fluidity, between the state and the individual and between consistency and adaptability; qualities that I believe are necessary for any theory striving to understand security in the globalized era.

1.3 Method

The theory of securitization takes a clearly constructivist position in its contention that security arises from labeling an issue a security issue (Buzan & Wæver, 2003, p 71). It is perhaps not surprising therefore, that many of the attempts made to formalize and apply the theoretic framework of securitization take a qualitative, intensive approach (see Balzacq, 2011). There are clear advantages to this approach – qualitative methods of inquiry allows for more sensitivity to context, contributing to a generally higher validity in the study. It also allows for greater freedom when choosing what material to consider when conducting an investigation.

But as Williams points out, securitization has also been informed by ideas brought forth by classical realism; the Copenhagen School's view of security as "a phenomenon that is concretely indeterminate and yet formally specific: constituted by a particular kind of speech-act" is shared by the classical realist Carl Schmitt (Williams, 2003, p. 516). Securitization is, after all, an attempt to retain security as a specific problem, characterized by a portrayal of urgency and the need for extraordinary measures (Wæver, 2007, p. 70). The theory's ambition to exist within and contribute to a broader discourse of security, as well as its clear focus on speech-acts as the main mechanism through which security arises, lends hope to the idea that a quantitative, extensive approach also can contribute to the understanding of security within the context of securitization.

In my attempt to characterize and understand the securitization of cyberspace, I will reflect this dual heritage by approaching the problem on the one hand by conducting an intensive study using a few selected texts as a basis of an idea analysis, and by proposing a method that involves a more extensive text analysis of a wider selection of material. The main reason for this two pronged approach is to provide both the broad scope and reliability associated with extensive inquiries and the deep understanding and validity associated with intensive inquiries, thereby providing complementary value to the study as a whole (Teorell & Svensson, 2007, 264ff).

There are a few potential benefits to an extensive approach. The sheer quantity of texts produced by states and other potential securitizing actors is more often than not so enormous that it effectively prohibits any one person from giving them all proper consideration. To illustrate, The United States Department of State, a single department within the American Government, produced 885 press releases in 2011 (Department of State, 2012a). An extensive text analysis could potentially be effective at capturing the scope of securitization by giving an indication as to the prevalence of securitizing speech-acts. One of the criticisms brought forth against securitization is the lack of a framework that allows for "systematic and comparative empirical analysis" (Strizel, 2007, p 358). Holger Strizel identifies two distinct "centers of gravity" within the theory of securitization and forwards his own proposal for a systematic approach centered on the process-approach to securitization. A quantitative-extensive approach that provides a broad indication of how a large quantity of texts relate to referent objects and to security, could be useful in the development of the opposite (speech-act) center of gravity by expanding the potential scope of any inquiry into the prevalence of speech-acts.

When attempting to analyze a large quantity of text, there is an obvious need to maximize the sophistication of the technique, while at the same time acknowledging the limitations that an automated process impose. In finding a balance between the automation and sophistication, there is bound to be a trade-off between reliability and validity (Teorell & Svensson, 2007, p. 269). A method of text analysis that relies solely on simple word frequency count might therefore be too blunt of an instrument to be of any greater value when attempting to investigate the securitization of referent objects. Conversely, a method of text analysis that utilizes too complex a coding scheme, requiring manual classification of sentences and sentiments, puts too strenuous a limit on the amount of text that can realistically be analyzed.

With this and with the basic premise of security arising from a securitizing actor "speaking security" in reference to a referent object in mind, the quantitative-extensive text analysis that I will test the merits of uses an approach that I hope will escape some of the limitations discussed above, and capture the broad features of a securitization process. This method centers around analyzing what words are most commonly used in relation to the referent object of interest.

By identifying instances where words that describe the referent object are mentioned by a securitizing actor (primary words), and by identifying what terms that are used in conjunction with the words describing the referent object (secondary words), the overall prevalence of instances where the referent object is expressed along with terms that would be indicative of securitizing speech-acts (such as "security", "threat", "defend", "protect") should emerge. This method differs from methods that simply count word frequency in that it provides some context by only including words that are semantically linked to the referent object by virtue of occurring in the same sentence By narrowing the inclusion of words to only those that are of interest as indicators of referent objects, and by letting those words function as a keys to the inclusion of surrounding words, this method can hopefully capture some of the complexity of a manual coding scheme while avoiding some of the pitfalls of more simplistic approaches.

1.4 Material

One of the strengths of securitization theory is that it is able to disjoin the concept of security from the exclusive domain of the state, while at the same time fully retaining the applicability of the concept on state conduct. It follows naturally from the constructivist position of securitization that any non-state entity that is in a position to effectively communicate a securitizing speech-act is of relevance to the study of international security – which efforts by Kurdish and Palestinian organizations aptly demonstrate (Buzan & Wæver, 2003, p. 195). But in order for a securitization to be successful, the securitizing actor must be in a powerful enough position for the speech-acts to be effective (Williams, 2003, p. 514). All states except perhaps a few "failed states" retain such a powerful position to varying degrees and few doubt the continued importance of states in the realm of security. The especially privileged role that states enjoy is the reason why this study will focus on state conduct as it relates to the securitization of cyberspace.

If states enjoy a privileged position in this regard, few states do to the extent that the United States of America does. Leading the world in economic development and military spending, the United States occupy a role that is unique in the world arena. For this reason, the material that will be examined in this study will be texts produced by American officials with the intent of communicating ideas, opinions and claimed facts to a broader public. The texts that I will subject to a qualitative-intensive study have all been selected because I believe they are representative of the American Governments approach to the discourses of security and the cyber domain. The first such text is a speech given by Secretary of Defense Leon Panetta to members of "Business Executives for National Security" (Panetta, 2012). The second text is an opinion piece published in the Wall Street Journal, written by President Barack Obama, titled "Taking the Cyberattack Threat Seriously" (Obama, 2012). These texts have both been produced with the authors speaking from the context of their respective posts, and with the explicit purpose of public consumption, albeit with different audiences in mind.

Texts that are subject to a quantitative-extensive method of inquiry should preferably not be subject to the same kind of unsystematic sampling that would likely be of benefit to an intensive study (Teorell & Svensson, 2007, p. 84). In choosing the material for the extensive part of the investigation, I therefore set out to utilize the automated process to its fullest extent, hoping to capture a substantial part of the available corpus of text produced by the relevant actors. To this end, I strategically selected much of the combined output of text produced for public consumption by the U.S. Department of Defense (DoD) and the U.S. Department of State (DoS) during the Obama Administration. I chose to include material from January 2009, the month that Barack Obama came into office, to the most currently available material in December of 2012.

The final corpus to be analyzed includes all 2522 DoS press releases for this period (U.S. Department of State, 2012a), all 6506 DoS "Remarks, Testimony, Speeches and Briefings by Department of State Officials" from this period (U.S

Department of State, 2012b) as well as all 324 speeches during the period made by the Secretaries of Defense (U.S. Department of Defense, 2012a), all 2522 DoD news releases from the period (U.S. Department of Defense, 2012b) and all 613 DoD transcripts of news briefings and "significant speeches" during the period (U.S. Department of Defense, 2012c).

My hope is that by using a substantial part of the complete output of text for public consumption produced by these departments during the Obama Administration, the reliability of the inquiry will be increased.

2 Qualitative Investigation

Securitization theory posits that in order for a referent object to be successfully securitized the securitizing actor must be in a position powerful enough within the specific social context that the speech-act uttered has an effect on the audience (Williams, 2003, p. 514). U.S. President Barack Obama and U.S. Secretary of Defense Leon Panetta, hold two of the arguably most powerful positions in the world. Their words carry heavy weight both within the U.S. and in the international community. Due to the positions they hold they have the capacity to carry out securitizing speech-acts to an extent that is almost unparalleled.

One of the premises behind securitization is the need for the securitizing actor to point to the critical vulnerability of the referent object that is to be securitized (Balzacq, 2011, loc. 321). In his opinion piece in the Wall Street Journal, Barack Obama begins by describing a meeting where a catastrophe had just occurred, vividly describing a hypothetical scenario wherein "country trains […] carrying industrial chemicals [had] exploded into a toxic cloud" and where "Water treatment plants in several states had shut down, contaminating drinking water and causing Americans to fall ill", all as a direct result of a cyberattack (Obama, 2012). Reiterating this danger in his speech to the Business Executives for National Security, Panetta elaborates: "The most destructive scenarios involve cyber actors launching several attacks on our critical infrastructure at one time […] The collective result of these kinds of attacks could be a "cyber Pearl Harbor:" […] it would paralyze and shock the nation and create a new, profound sense of vulnerability" (Panetta, 2012).

References to disaster scenarios where a hypothetical attack could potentially cascade and spread throughout a network causing devastation in its wake is a distinct feature of the cyber discourse. Hansen & Nissenbaum uses the term "hypersecuritization" to describe this phenomenon and points out that even though actors often conjure up images of catastrophes, there are no clear precedents for such events in the cyber domain (Hansen & Nissenbaum, 2009, p. 1164). The cascading almost domino-like effects that are posited during hypersecuritization allows the securitizing agent to link the rather abstract referent object that is "the network" to more defined referent objects such as "businesses" and "infrastructure", and in extension "society".

In place of actual precedents, securitizing actors who seek to illustrate the urgent need to take extraordinary measures in order to protect the referent object are left to invoking images of previous catastrophes. Leon Panetta thus likens the potential devastation of a serious cyber attack both to Pearl Harbor and to 9/11, claiming that "A cyber attack perpetrated by nation states are [sic] violent extremists groups could be as destructive as the terrorist attack on 9/11" (Panetta, 2012). By invoking the images of previous disasters, the vulnerability of the

referent object is effectively established and the case for the existence of an existential threat capable of perpetrating such an attack can be presented (Buzan & Hansen, 2007, p. xxxv).

Obama describes the threat of a cyberattack as almost omnipresent and originating from a range of different actors such as "foreign governments, criminal syndicates and lone individuals" who are attempting to gain access to "financial, energy and public safety systems every day" (Obama, 2012). By further claiming that there has been an increase in attacks against "nuclear and chemical industries", the presented image of the threat turns existential (Obama, 2012). Adding to this understanding the dangers of the cyber domain, Leon Panetta describes the threat posed by cyber attacks as "every bit as real as the more well-known threats like terrorism [and] nuclear weapons proliferation" (Panetta, 2012). The images of nuclear and chemical plants, along with terrorism and nuclear weapons imply the existence of a threat to the sovereignty of the state on a scale that requires great urgency of action to prevent (Wæver, 2007, p. 70).

In order to alleviate the threat against the referent objects, Obama asks his audience to support efforts that would among other things "make it easier for government, if asked, to help […] companies prevent and recover from attacks". To this end, he also urges "the Senate to pass the Cybersecurity Act of 2012 and Congress to send me comprehensive legislation so I can sign it into law"(Obama, 2012). Leon Panetta notes that the Cybersecurity Act is "is victim to legislative and political gridlock" and calls this "unacceptable not just to me, but to you and to anyone concerned with safeguarding our national security" (Panetta, 2012). An equivalence is drawn between preventing the implementation of measures proposed by the securitizing actor and being unconcerned with the security of the referent object. Panetta argues that the Department of Defense must have "capabilities" to act in the cyber space – they must be able to employ extraordinary means to defend the people: "If a crippling cyber attack were launched against our nation, the American people must be protected [and] the Defense Department must be ready [...] to act" (Panetta, 2012).

Michael Williams reiterates the point made by Buzan et. al, that "in the security discourse, an issue is dramatized and presented as an issue of supreme priority" (Williams, 2003, p. 514). Panetta illustrates how the threat of a cyberattack is already treated as having supreme priority by his department: "the department is continuing to increase key investments in cybersecurity even in an era of fiscal restraint". He emphasizes the need to invest in "skilled cyber warriors", making the comparison to the development of "the world's finest counterterrorism" in the previous ten years (Panetta, 2012). In the same way that the U.S. used extraordinary means to respond to the threat posed by terrorism after the attacks on 9/11, so too should also a cyber force be developed in anticipation of a coming cyberattack, is the argument put forth by Panetta.

Thierry Balzacq notes that in order for a speech-act to achieve the desired effect, a securitizing actor needs "tune his/her language to the audience's experience" (Balzacq, 2011, loc. 472). This attempt to conjure an emotional response by appealing to the common experience of the audiences is a reoccurring theme throughout both texts. But it is perhaps best illustrated in the final part of

Panetta's speech, which also aptly provides a summary of the way a securitizing speech-act is constructed; from describing the vulnerability of the referent object, and characterizing the nature of the existential threat, to invoking a sense of urgency and portraying the need for extraordinary measures in order to protect the referent object (construed here to especially encompass "the children"):

> "Before September 11, 2001, the warning signs were there. We weren't organized. We weren't ready and we suffered terribly for that lack of attention. We cannot let that happen again. This is a pre-9/11 moment. The attackers are plotting. Our systems will never be impenetrable just like our physical defenses are not perfect, but more can be done to improve them. We need Congress and we need all of you to help in that effort […] [W]e always have been able to defend our interests and our values. That must remain our most important mission on land, at sea, in the air, in space and yes, in cyberspace. This is not just a responsibility, it is a duty that we owe to our children and their children in the future." (Panetta, 2012)

3 Quantitative Investigation

3.1 Design

The Department of Defense and the Department of State together make up the two of the most integral parts in the U.S. government's capabilities to act in the international community, if by quite different means. While the Department of State uses diplomacy as its main tool in the strive for "a more democratic, secure and prosperous world" (U.S. Department of State, 2012c, p. 4), the Department of Defense provides the military means to back the diplomatic efforts. The two departments and their representatives are also by virtue of their roles potential instigators of security, or more specifically, securitizing speech-acts (Buzan & Hansen, 2007, p. xxxvi). Conforming to the zeitgeist, the departments have steadily increased the frequency and volume of their communication with the outside world – from 1995 to 2012, the number of news items produced by the Department of Defense has more than doubled from 366 to 761 (U.S. Department of Defense, 2012b).

More information is ostensibly a good thing for those who attempt to make sense of the world, but when it becomes ubiquitous, information can easily have an overwhelming effect. And as Nate Silver points out, the promise of a world where the interpretation of this information can be left solely to computers has not come to fruition (Silver, 2012, p. 9). A computerized process presents its value when used with its limitations in mind and in combination with a human interpretation of the results it produces. Never the less, it is important to continue to explore the potential and subsequent limits of automated processes. And given the prevailing understanding of speech-acts within the theory of securitization, not as a metaphorical notion but as a fairly concrete one, wherein the utterance of the word security in reference to a referent object gives rise to security (Hansen, & Nissenbaum, 2009, p. 1159), the theoretic framework seems to lend itself to the possibility of the kind of concrete approach that a quantitative-extensive study entail.

In order to fulfill the quite specific technical and methodological needs of the method to be employed, I wrote a relatively simple computer program that can analyze large amounts of text and produce an output that is consistent with the intent of the method. The program works in a few steps, the general procedure being detailed here for the sake of transparency. First, it takes a file consisting of a corpus of preprocessed html files and removes the residual html code to reduce the chance of interference. Next, it prompts the user to input a keyword – the

keyword being the primary word of interest (in the case of this study, words associated with the referent object). The program then finds all the instances where the keyword (and any variations of it, i.e. if the keyword is "example", it also includes "examples") is used and extracts the sentences in which the keyword occurs. The program then counts the frequency of all the words in the sentences that match the keyword, resulting in a list of frequencies. However, some words, like "the" and "of" are very frequently occurring in the English language but does not bear much relevance to the inquiry into the words associated with our keyword. The program deals with this by ignoring the 100 most common words in the English language (Perc, 2012). The end result is a list of all the words that are used in the same sentence as the keyword, in ascending order of frequency.

The keywords that represent the referent object are central to the inquiry and in order for the investigation to be conducted, we must establish what the proper keywords should be. If the securitizing actor in the case of this investigation is represented by the American Government (more specifically representatives of Department of Defense and the Department of State) the referent object is a bit more abstract. The word that offers the perhaps most intuitive description of the referent object, and the word that I have been using to connote the referent object of interest is "cyberspace". This therefore seems like a natural first candidate for a keyword as it captures the general aspects of the discourse quite succinctly.

But as noted earlier, the cyber domain does not exist as a totally insulated plane. It is occupied by states, individuals, private companies and many other organizations. The multitude of security discourses that relate to these groups and individuals in the physical world are often mirrored in discourses of cyber security. Lene Hansen and Helen Nissenbaum therefore view the discourse of cyber security as "arising from competing articulations of constellations of referent objects rather than separate referent objects", exemplified by the "linkage between 'networks' and 'individuals' and human collective referent objects" present in this discourse (Hansen & Nissenbaum, 2009, p. 1163).

The view of referent objects as linked constellations potentially presents a challenge to the specific quantitative model to be used. The ability of the model to capture a broader discourse where the makeup of the referent object is contended might be limited. On the other hand, the model still might be able to provide a characterization of a referent object as it is understood, and subsequently securitized, by a specific securitizing actor. The Department of Defense, for an example, likely has a specific understanding of what the most important referent object in the cyber sector is. The military might be more inclined to see the territory of the nation as the collective referent object (Williams, 2003, p. 513), and so its proposed view of the relevant constellation is made by linking "networks" with the territory of the state. It would be more likely for the department with the mission to "deter war and to protect the security of our country" (U.S Department of Defense, 2012d) to be concerned with potential cyberattacks on the nations networks and infrastructure than for example with the corrosion of privacy as a result of increased governmental presence on the internet.

Given that the scope of this study is limited to the study of representatives of the American Government as securitizing actors, and given the likelihood that there is some degree of cohesion within the government, the model still might prove useful, even after accounting for the view espoused by Hansen and Nissenbaum. And even if there is disparity and competition between the way the DoD and the DoS view the referent object, this could potentially manifest in the results of the study. As a litmus test for the limits of the method, I will therefore use the word "network(s)" as a second keyword, indicating reference to a second understanding of the referent object – one that is contingent upon a invocations made by the securitizing actor between the keyword "network(s)" and a "bounded human collective" (Hansen & Nissenbaum, 2012, p. 1163), a "bounded human collective" in this understanding being "located at the 'middle scale of limited collectives', larger than the individual and smaller than humanity" (Buzan & Hansen, 2007, p. xxxvi). This will hopefully both illustrate invocations made to other terms included in a specific constellation of referent objects championed by the securitizing agent and any connections made between the constellation and words that would indicate securitizing speech-acts.

3.2 Results

Table 1.1.1 - *25 most frequent primary and secondary words from DoD speeches Jan 2009 – Dec 2012, * = keyword*

#	Word	Frequency	#	Word	Frequency
1	cyberspace*	143	14	security	16
2	space	49	15	threats	16
3	air	33	16	how	15
4	domain	32	17	forces	14
5	sea	28	18	protect	14
6	cyber	22	19	ensure	13
7	department	21	20	networks	13
8	capabilities	20	21	open	13
9	defend	17	22	strategy	13
10	new	17	23	use	13
11	defense	16	24	where	13
12	land	16	25	adversaries	11
13	military	16			

Table 1.1.2 - *25 most frequent primary and secondary words from DoD speeches Jan 2009 – Dec 2012, * = keyword*

#	Word	Frequency	#	Word	Frequency
1	networks*	329	14	systems	34
2	network*	91	15	protect	31
3	defense	82	16	how	30
4	military	82	17	attack	27

5	cyber	68	18	critical	27
6	security	61	19	threat	26
7	government	47	20	support	24
8	defend	43	21	command	23
9	attack	42	22	capabilities	22
10	department	40	23	against	21
11	computer	48	24	air	21
12	information	38	25	private	21
13	defenses	34			

Table 1.2.1 - *25 most frequent primary and secondary words from DoD news reports Jan 2009 – Dec 2012, * = keyword*

#	**Word**	**Frequency**	#	**Word**	**Frequency**
1	cyberspace*	29	14	headquarters	6
2	space	21	15	information	6
3	air	19	16	joint	6
4	force	14	17	washington	6
5	operations	12	18	bas	5
6	security	12	19	colo	5
7	director	11	20	development	5
8	command	8	21	international	5
9	new	8	22	peterson	5
10	continue	7	23	systems	5
11	strategy	7	24	capabilities	4
12	dc	6	25	department	4
13	defense	6			

Table 1.2.2 - *25 most frequent primary and secondary words from DoD news reports Jan 2009 – Dec 2012, * = keyword*

#	**Word**	**Frequency**	#	**Word**	**Frequency**
1	networks*	42	14	service	15
2	network*	32	15	washington	15
3	information	30	16	dc	14
4	defense	28	17	including	14
5	office	20	18	secretary	14
6	dod	19	19	executive	12
7	operations	18	20	military	11
8	director	17	21	systems	11
9	chief	16	22	cyber	10
10	continue	16	23	efforts	10
11	army	15	24	force	10
12	department	15	25	program	10
13	security	15			

Table 1.3.1 - *25 most frequent primary and secondary words from DoD transcripts Jan 2009 – Dec 2012, * = keyword*

#	**Word**	**Frequency**	#	**Word**	**Frequency**

1	cyberspace*	96	14	military	10
2	space	36	15	open	10
3	air	20	16	well	10
4	capabilities	17	17	cyber	9
5	new	16	18	defend	9
6	sea	13	19	those	9
7	security	12	20	threats	9
8	states	12	21	attack	8
9	united	12	22	defense	8
10	cooperation	11	23	know	8
11	department	11	24	national	8
12	operations	11	25	protect	8
13	domains	10			

Table 1.3.2 - *25 most frequent primary and secondary words from DoD transcripts Jan 2009 – Dec 2012, * = keyword*

#	**Word**	**Frequency**	#	**Word**	**Frequency**
1	network*	259	14	intelligence	36
2	networks*	252	15	Afghanistan	34
3	haqqini	76	16	attacks	34
4	those	66	17	support	34
5	defense	63	18	how	33
6	know	60	19	operations	33
7	security	52	20	get	21
8	think	48	21	information	21
9	military	47	22	pakistan	30
10	going	44	23	taliban	30
11	forces	39	24	just	29
12	well	38	25	threat	29
13	government	37			

Table 2.1.1 - *25 most frequent primary and secondary words from DoS press releases Jan 2009 – Dec 2012, * = keyword*

#	**Word**	**Frequency**	#	**Word**	**Frequency**
1	cyberspace*	91	14	united	11
2	security	33	15	working	11
3	international	22	16	cooperation	11
4	issues	19	17	address	10
5	space	17	18	build	9
6	challenges	16	19	continue	9
7	cyber	15	20	development	9
8	global	14	21	countries	9
9	threats	13	22	economic	8
10	norms	12	23	internet	8
11	national	11	24	governments	8
12	states	11	25	military	7
13	together	11			

Table 2.1.2 - *25 most frequent primary and secondary words from DoS press releases Jan 2009 – Dec 2012, * = keyword*

#	Word	Frequency	#	Word	Frequency
1	network*	557	14	global	104
2	networks*	439	15	social	100
3	new	213	16	through	99
4	united	169	17	development	91
5	women	146	18	countries	90
6	support	139	19	south	90
7	states	137	20	department	89
8	republic	128	21	world	83
9	people	113	22	security	81
10	international	110	23	oppertunities	80
11	business	108	24	access	77
12	networking	105	25	islands	76
13	state	105			

Table 2.2.1 - *25 most frequent primary and secondary words from remarks, testimony, speeches and briefings by DoS officials Jan 2009 – Dec 2012, * = keyword*

#	Word	Frequency	#	Word	Frequency
1	cyberspace*	181	14	policy	15
2	international	53	15	economic	14
3	law	41	16	global	14
4	issues	34	17	how	14
5	states	33	18	need	14
6	new	28	19	strategy	14
7	security	24	20	where	14
8	united	23	21	question	13
9	world	22	22	rules	13
10	norms	19	23	space	13
11	internet	18	24	work	13
12	information	16	25	behavior	11
13	cyber	15			

Table 2.2.2 - *25 most frequent primary and secondary words from remarks, testimony, speeches and briefings by DoS officials Jan 2009 – Dec 2012, * = keyword*

#	Word	Frequency	#	Word	Frequency
1	networks*	1393	14	networking	195
2	network*	1116	15	how	188
3	new	260	16	through	186
4	social	260	17	support	176
5	people	253	18	countries	169
6	illicit	247	19	security	168
7	states	228	20	help	166
8	global	224	21	information	162
9	criminal	218	22	work	153

10	united	207	23	well	150
11	women	203	24	across	148
12	international	201	25	those	145
13	world	201			

3.3 Analysis

Presented above are the results of the quantitative-extensive analysis, using the keywords “cyberspace” and “network(s)” applied to the various categories of text produced by the DoD and the DoS during the Obama Administration. The overall picture that emerges from the results is the relatively high occurrence of words that indicate the presence of securitizing speech-acts. The central word to the inquiry – and indeed the word that is imperative to the construction of security itself according to the prevailing understanding of securitization – “security”, is present amongst the most frequent words used in association with “cyberspace” in every single corpus of text from both the DoD and the DoS, as seen in tables 1.1.1, 1.2.1, 1.3.1, 2.1.1 and 2.2.1. In the publications from the DoD (tables 1.1.1, 1.2.1 and 1.3.1), the word “security” was uttered in 15% of all instances where “cyberspace” was mentioned. More surprising perhaps, given the division of labor between the DoD and the DoS, is that in the DoS press releases (table 2.1.1) “security” is the word most frequently associated with “cyberspace”; “security” was spoken by representatives of the DoS in 21% of all instances where “cyberspace” was mentioned.

One of the potential weaknesses of the method, as discussed previously, is its inability to provide context beyond a quite rudimentary level. The analysis shows that “security” is mentioned frequently in the context of grammatical proximity to the word “cyberspace”, but of course, there could well be instances where the intention of the author was not at all to indicate any relation between the two. In fact, the opposite might even be true; the sentence “efforts to increase security should not include cyberspace” would be an example of a sentence where a securitizing speech-act clearly is not present, but that the method would none the less include in the resulting report.

However, this interpretation becomes less likely when considering some of the other words that the tables indicate are frequently associated with cyberspace. To illustrate, either or both of the words “defend” or “defense” are present in conjunction with “cyberspace” in all tables associated with the DoD (1.1.1, 1.2.1 and 1.3.1), “threats” are included in two of the three DoD tables (1.1.1 and 1.3.1) and in one of the two DoS tables (2.1.1) and “protect” is also present in two of the three DoD tables (1.1.1 and 1.3.1). As noted, in order for a referent object to be successfully securitized, the securitizing agent must point to the presence of a critical threat to the referent object and to the need to take extraordinary measure to ensure its safety (Williams, 2003, p. 514). The frequent occurrence of words

like “threat”, “protect” and “defend” lends credibility to the interpretation that the ubiquity of the word “security” is indicative of securitizing speech-acts.

The results of the inquiries using the keyword “network(s)” are less pronounced, however. Looking at the tables portraying the usage of the keyword by the DoD they seem to be inconclusive as to their meaning. On the one hand, table 1.1.2 showing the references made to “network” and “networks” in speeches made by representatives of the DoD during the Obama Administration, is quite consistent with the portrayal of “cyberspace”. “Security” is frequently mentioned alongside “network(s)”, as are other words that would indicate securitizing speech-acts, such as “threat”, “defend”, “attack” and even “critical”. In the speeches, links are also seemingly established both between “network(s)” and “cyber”, and between “network(s)” and to a human collective, represented in this case by the “government” indicating the portrayal of a constellation of referent objects. Hints of a similar trend can also be deduced from table 1.2.2. On the other hand, the results presented in table 1.3.2, showing DoD transcripts, the primary word “network(s)” seems not at all be used in the cyber context.

Indeed, it becomes clear from the presence of “haqqini” along with “taliban” and “pakistan” that “network(s)” in the context of the DoD transcripts does not pertain to the cyber domain, but rather to the Haqqini network, an insurgent group operating in Afghanistan and Pakistan (Mazzetti et. al, 2011). The fact that the automated process in this instance cannot differentiate between a reference to “the Sopranos of the Afghanistan war” (Mazzetti et. al, 2011) and references to networks as it pertains to the cyber domain, reveals a limitation imposed by the method. The same weakness is echoed in the results from the DoS. Even though “security” is present both in the press releases (table 2.1.2) and in the “remarks, testimony, speeches and briefings” (table 2.2.2), it is not clear in what context “network(s)” appear. There are indications that “network(s)” are linked to human collectives, as evidenced by the presence of the terms “women”, “people”, “business” and “state”, but there are no clear indications that “network(s)” are understood in the context relevant to this inquiry.

The results from the attempts using the primary word “network(s)” to capture both the construction of constellations of referent objects, and the indications of securitizing speech-acts fall into two categories. In the first case, the method does a reasonably good job both at capturing the links between “network(s)” and a human collective, indicating a constellation of referent objects, and at capturing the association made between this constellation and “security”. In the second case, the polysemic quality of the word “network(s)” proves an obstacle. In the case of DoD news releases, the contextual use of “network(s)” is obvious (and obviously not one that bears any relevance to this investigation) and in the case of the DoS results, its use is too ambiguous to provide any real insight.

4 Evaluation and Conclusions

4.1 Evaluation

While conclusions presented here are meant to speak both to the specifics of this investigation and to the general nature of extensive approaches, they are by no means exhaustive or representative of all forms of quantitative-extensive inquiries. The purpose is rather to point to specific strengths and limitations present in this inquiry and relate them to a broader discussion of the strength and limitations of different methodological approaches.

Overall, the quantitative-extensive investigation into the securitization of cyberspace produces results of varying quality. When used to investigate the prevalence of securitizing speech-acts uttered in reference to a well-defined understanding of the relevant referent object, it provides quite useful and relatively unambiguous results that to a reasonable extent reflect the characterization provided by the quantitative inquiry. The ubiquity of words that would indicate a securitization process, such as “security”, “threat”, “defense” and “protect”, lends credibility to the utility of the extensive method as a mean to provide a broad characterization of such a process. While its ability to provide contextualized and specific insight is limited compared to the qualitative-intensive approach, the extensive approach can provide a value by supplementing intensive analysis seeking to examine the reliability of its results. Due to its ability to process text corpuses that would have otherwise remained inaccessible in their entirety, an automated extensive analysis can provide insight into the broader trends and themes that emerge from the portrayal of a referent object by a group of securitizing actors.

To a lesser extent, the results also indicate a possibility that an extensive method could add to the understanding of a securitization process even if the concept of the referent object is understood as a constellation of linked referent objects. Results do not always speak for themselves, they have to be evaluated and put into the context of a broader understanding that can only be provided by a human input. The fact that when the method used in this investigation failed to capture the intended phenomena it did so quite obviously could in this regard be seen as a redeeming quality. For this reason, one should be careful not incorporate too much of the analysis into any method that relies upon an automated process. A text analysis must be able to show a reasonable amount of validity and a good amount of reliability (Teorell & Svensson, 2007, p. 269). By infusing too much complexity into the process in order to seek greater validity in the results, any

weaknesses inherent to the method are at risk of being amplified. In the case of this particular investigation, the compromise struck between complexity and simplicity – and in extension between validity and reliability – provided results that when apparently valid could rely upon quite significant reliability, but when proved invalid allowed for easy recognition of this fact.

4.2 Conclusions

The qualitative investigation of the texts by Obama and Panetta show that a securitization of cyberspace is enabled by presenting cyberspace as a series of connected referent objects, bound together by a network. This constellation of referent objects is presented as under a constant threat of attack from omnipresent adversaries. By invoking images of catastrophes in the past such as Pearl Harbor and 9/11, the securitizing actors can relate previous catastrophes to hypothetical disaster scenarios involving cascading effects that present existential threats to a range of referent objects linked to cyberspace. By establishing the critical vulnerability of cyberspace and the referent objects it is connected to, Obama and Panetta can go on to proclaim the need for urgent and decisive action to combat the threat posed to the sovereignty of the nation.

Elements of the characterization provided by the qualitative inquiry are also found in the quantitative inquiry. When basing the inquiry on a one-dimensional and well-defined understanding of the referent object, the results of the quantitative inquiry displayed a high prevalence of words that would be indicative of securitizing speech-acts. The quantitative method lends support to the idea that the elements that make up the securitizing speech-acts in the qualitative analysis are mirrored in the corpus of texts produced by the DoD and the DoS. When additional dimensions were added to the understanding of the referent object however, and its meaning became less well-defined, the qualitative inquiry was less successful at providing conclusive and unambiguous results.

The quantitative-extensive method is best utilized as a complimentary tool to an intensive qualitative inquiry. Where an intensive approach might fail to recognize how an observation relates to a broader pattern, the extensive approach can provide a more overarching map. It can provide the broad strokes and outlines to a representation of a phenomenon, but the coloring and details are best left to a more intensive approach.

5 References

Balzacq, Thierry 2011 "A theory of securitization" in Balzacq, Thierry (ed.) *Securitization Theory – How Security Problems Emerge and Dissolve.* New York: Routledge (Amazon Kindle Version), location 249-1116.

Buzan, Barry – Hansen, Lene, 2007. "Editors' Introduction" in Barry Buzan & Lene Hansen (ed.) *International Security – Volume I, The Cold War and Nuclear Deterrence.* SAGE Publications, p. xvii-xl

Buzan, Barry – Wæver, Ole, 2003. *Regions and Powers – The Structure of International Security.* New York: Cambridge University Press

Hansen, Lene – Nissenbaum, Helen, 2009. "Digital Disaster, Cyber Security and the Copenhagen School", *International Studies Quarterly.* Vol. 53, No. 4, p. 1155-1175

Mazzetti, Mark – Shane, Scott – Rubin, Alissa, J., 2011. "Brutal Haqqini Crime Clan Bedevils U.S. in Afghanistan". *New York Times, 2011-09-24.* Available 2013-01-06 via http://www.nytimes.com/2011/09/25/world/asia/brutal-haqqani-clan-bedevils-united-states-in-afghanistan.html?pagewanted=all&_r=0

Obama, Barack, 2012. "Taking the Cyberattack Threat Seriously". *Wall Street Journal, Opinion* 2012-07-19. Available 2013-01-06

Panetta, Leon, 2012. "Defending the Nation from Cyber Attack". U.S. Department of Defense. Available 2013-01-6 via http://www.defense.gov/speeches/speech.aspx?speechid=1728

Perc, Matjaz, 2012. "Top 100 most frequently used 1-grams between 2006 and 2008 in the English corpus" available 2013-01-06 via http://www.matjazperc.com/ngrams/1G2006Y2008E2.html

Silver, Nate, 2012. *The Signal and the Noise, Why So Many Predictions Fail – But Some Don't.* New York: Penguin Group.

Strizel, Holger, 2007. "Towards a Theory of Securitization: Copenhagen and Beyond", *European Journal of International Relations.* Vol. 13, No.3, p. 357-383.

Teorell, Jan – Svensson, Torsten, 2007. *Att fråga och att svara*. Edition 1:1. Malmö: Liber AB

U.S. Department of Defense, 2012a. *Secretary of Defense Speeches, 01 January 2009 – 20 December 2012.* Available 2013-01-06 via http://www.defense.gov/speeches/SecDefArchive.aspx

U.S. Department of Defense, 2012b. *News Releases, 01 January 2009 – 20 December 2012.* Available 2013-01-06 via http://www.defense.gov/releases/archive.aspx

U.S. Department of Defense, 2012c. *Transcripts, 01 January 2009 – 20 December 2012.* Available 2013-01-06 via http://www.defense.gov/transcripts/archive.aspx

U.S. Department of Defense, 2012d, *About the Department of Defense.* Available 2013-01-06 via http://www.defense.gov/about/#mission

U.S. Department of State, 2012a. *Press Releases, 01 January 2009 – 20 December 2012.* Available 2013-01-06 via http://www.state.gov/r/pa/prs/ps/index.htm

U.S. Department of State, 2012b. *Remarks, Testimony, Speeches and Briefings by Department of State Officials. 01 January 2009 – 20 December 2012.* Available 2013-01-06 via http://www.state.gov/r/pa/ei/speeches/index.htm

U.S. Department of State, 2012c. "Fiscal Year 2012 Agency Final Report".

U.S. Strategic Command, 2012. "U.S. Cyber Command". Available 2013-01-06 via http://www.stratcom.mil/factsheets/Cyber_command/

Williams, Michael C., 2003. "Words, Images, Enemies: Securitization and International Politics", *International Studies Quarterly*, Vol. 47, No. 4, p. 511-531.

World Bank, 2012. "Internet users (per 100 people)". Available 2013-01-06 via http://data.worldbank.org/indicator/IT.NET.USER.P2/countries/1W?display=graph

Wæver, Ole, 2007. "Securitization and Desecuritization" in Barry Buzan & Lene Hansen (ed.) *International Security – Volume III, Widening Security*. SAGE Publications, p. 66-98.

6 Appendices

6.1 Appendix 1

Below is the source code for the (python) program used in the quantitative inquiry:

```
# WordWebs v2.121221

from datetime import datetime
import re
user_path = raw_input("Input the path to the file to be analyzed:
")
workfile = open('user_path', 'r').read()

workfile = workfile.replace('\n', ' ')
class TextWeb(object):

    common_words = ['the', 'of', 'and', 'to', 'a', 'in', 'that',
'is', 'was', 'i', 'for', 'as', 'with', 'it', 'The', 'be', 'on',
'his', 'he', 'by', 'not', 'at', 'are', 'or', 'you', 'from', 'had',
'have', 'which', 'this', 'her', 'but', 'an', 'they', 'were',
'all', 'their', 'one', 'we', 'him', 'she', 'would', 'so', 'been',
'will', 'my', 'who', 'more', 'them', 'can', 'has', 'me', 'In',
'He', 'when', 'no', 'It', 'there', 't', 'out', 'into', 'if',
'its', 'said', 'up', 'other', 'time', 'than', 'about', 'what',
'A', 'may', 'some', 'your', 'do', 'only', 'our', 'could', 'any',
'these', 'such', 'two', 'like', 'This', 'very', 'then', 'But',
'also', 'should', 'And', 'first', 'over', 'made', 'see', 'man',
'most', 'now', 'us', 'must', 'before']

    saved_rows = ""
    keyword = ""
    def __init__(self, raw, printing, tofile, preprocessed,
hardkeyword, recovery):
        # if true, raw data will be provided, if false, common
words will be omitted
        self.raw = raw
```

```
        self.printing = printing
        self.tofile = tofile
        self.preprocessed = preprocessed
        self.hardkeyword = hardkeyword
        self.recovery = recovery

    def getKeyword(self):
        if self.hardkeyword:
            user_keyword = raw_input("Input (hard) keyword:\n")
            self.keyword = user_keyword
            keyword = re.compile(user_keyword)
        else:
            user_keyword = raw_input("Input (soft) keyword:\n")
            self.keyword = user_keyword
            keyword = re.compile(user_keyword.lower() + "[\S]*")
        return keyword

    def cleanUp(self, rawText):
        print "cleanUp"
        # removes anything that is within brackets including the
brackets (assumed to be html)
        rawText = re.sub("\[[^\]]*]", "", rawText)
        rawText = re.sub("http://[^\)]*\)", "", rawText)
        rawText = re.sub("\(/\w*/\)", "", rawText)
        rawText = re.sub("\S\.htm[l]?", "", rawText)
        rawText = re.sub("/\w*/", "", rawText)
        rawText =
re.sub("\w*\.[com|net|org|mil|gov|fr|se|dk|no|co\.uk]", "",
rawText)
        rawText = re.sub("--", " ", rawText)
        rawText = re.sub("\w*javascript\w*", "", rawText)
        rawText = re.sub("\(http[s]?|[^a-zA-Z0-9. ]{2}", "",
rawText)
        rawText = re.sub("\S*%\S*", "", rawText)
        ## low priority cleanup
        rawText = re.sub("[^\s|\w|.]|\||_", "", rawText)
        rawText = " ".join(rawText.split())
        return rawText

    def splitToRowList(self, rowText_in):
        print "splitToRowList"
        # returns a list of sentences from cleaned text
        rowText = self.cleanUp(rowText_in)
        # matches new sentence
        rowRegex = re.compile("[^A-Z]\. [^$]")
        returnList = []
```

```
        maxN = str(len(rowRegex.findall(rowText)))
        count = 0
        for i in range(len(rowRegex.findall(rowText))):
            cm = rowRegex.search(rowText)
            returnList.append(rowText[:cm.end()-1])
            rowText = rowText[cm.end()-1:]
            count += 1
            if self.printing:
                print str(count) + "/" + maxN
        self.saved_rows = returnList
        saved_rows_file = open("saved_state", 'w')
        saved_rows_file.write(str(returnList))
        return returnList

    def recoverSavedState(self):
        recoveredList = open('saved_state', 'r').read()
        recoveredList = recoveredList[1:-1]
        recoveredList = recoveredList.replace("'", "")
        recoveredList = recoveredList.replace('"', '')
        recoveredList = recoveredList.split(",")
        return recoveredList

    def genWordList(self, textList_in):
        print "genWordList"
        # generates a list containing all the words used in the
text
        if self.preprocessed:
            textList = textList_in
            #re-establishes saved_rows as textList lest it be
forgotten in the next ev. recursion
            self.saved_rows = textList
        elif self.recovery:
            textList = self.recoverSavedState()
            self.saved_rows = textList

        else:
            textList = self.splitToRowList(textList_in)
        keyword = self.getKeyword()
        wordList = []
        for line in textList:
            if keyword.search(line):
                for word in line.split():
                    # checks if list is to be raw or cleaned for
common words
                    if self.raw:
```

```
                        wordList.append(word.replace(".",
"").lower())
                    else:
                        inList = False
                        for item in self.common_words:
                            if word.replace(".", "").lower() ==
item:
                                inList = True
                        if not inList:
                            wordList.append(word.replace(".",
"").lower())



        return wordList

    def wordFrequency(self, wordList_in):
        print "wordFrequency"
        # returns a list of lists in the form [['word_i',
'frequency_i']]
        wordList = self.genWordList(wordList_in)
        # returns word frequency list in the form ["word",
"frequency"]
        sortedList = sorted(wordList)
        singlesList = ["PLACEHOLDER"]
        for i in range(len(sortedList)):
            if sortedList[i] != singlesList[-1]:
                singlesList.append(sortedList[i])
        frequencyList = []

        for i in range(len(singlesList)):
            placeHolderList = []
            placeHolderList.append(singlesList[i])

placeHolderList.append(str(sortedList.count(singlesList[i])))
            #print str(sortedList.count(singlesList[i]))
            frequencyList.append(placeHolderList)

        return frequencyList

    def reportFrequency(self, frequencyList_in):
        print "reportFrequency"
        # reports frequency table of words in the order of highest
occurrence (formatted for readability)
        frequencyList = self.wordFrequency(frequencyList_in)
```

```
        highest = 0
        # finds the word with the highest occurrence
        for i in range(len(frequencyList)):
            if int(frequencyList[i][1]) > highest:
                highest = int(frequencyList[i][1])

        # uses highest to loop through and print the words from
highest frequency to lowest
        count = highest
        #print count
        if self.tofile:
            outfile =
open("word_webs_report_{0}_{1}.txt".format(str(datetime.now())[:19
].replace(" ", "__").replace(":", "_"), self.keyword.replace(" ",
"_")), 'a')
        while count > 0:
            for i in range(len(frequencyList)):
                if int(frequencyList[i][1]) == count:
                    if self.printing and self.tofile:
                        print "Word: '%s' - frequency: %s" %
(frequencyList[i][0], frequencyList[i][1])
                        outfile.write("Word: '%s' - frequency: %s
\n" % (frequencyList[i][0], frequencyList[i][1]))
                    elif self.printing:
                        print "Word: '%s' - frequency: %s" %
(frequencyList[i][0], frequencyList[i][1])
                    elif self.tofile:
                        outfile.write("Word: '%s' - frequency: %s
\n" % (frequencyList[i][0], frequencyList[i][1]))
            count -= 1
        if self.tofile:
            outfile.close()
        newKeyword = raw_input("Exit?\n")
        if newKeyword.lower() != "yes":
            TextWeb(False, True, True, True, True,
False).reportFrequency(self.saved_rows)

## Raw, Printing, tofile, preprocessed, hardkeyword, recovery
TextWeb(False, True, True, False, True,
False).reportFrequency(workfile)
```

6.2 Appendix 2

The following example bash-script can be used within a UNIX-environment to automate the process of creating the preprocessed file, using html2text.py[1]:

```
   #! /bin/bash

   for item in *.htm
   do
     python2 /path/to/html2text.py "$item" >>
/path/to/destination/file.txt
   done
```

[1] Created by Aaron Swartz (http://www.aaronsw.com), available at http://www.aaronsw.com/2002/html2text/ under the GNU GPL 3.0.