A MEDIA ARCHAEOLOGY OF ONLINE COMMUNICATION PRACTICES

THROUGH SEARCH ENGINE AND SOCIAL MEDIA OPTIMIZATION

by

KAREN M. ESTLUND

A DISSERTATION

Presented to the School of Journalism and Communication
and the Division of Graduate Studies
in partial fulfillment of the requirements
for the degree of
Doctor of Philosophy

June 2021

DISSERTATION APPROVAL PAGE

Student: Karen M. Estlund

Title: A Media Archaeology of Online Communication Practices through Search
Engine and Social Media Optimization.

This dissertation has been accepted and approved in partial fulfillment of the
requirements for the Doctor of Philosophy degree in the School of Journalism and
Communication by:

| | |
|---|---|
| Dr. Kim Sheehan | Chairperson |
| Dr. Biswarup Sen | Core Member |
| Dr. Seth Lewis | Core Member |
| Dr. Colin Koopman | Institutional Representative |

and

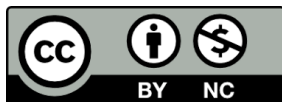| | |
|---|---|
| Andy Karduna | Interim Vice Provost for Graduate Studies |

Original approval signatures are on file with the University of Oregon Division of
Graduate Studies.

Degree awarded June 2021

DISSERTATION ABSTRACT

Karen M. Estlund

Doctor of Philosophy

School of Journalism and Communication

June 2021

Title:  A Media Archaeology of Online Communication Practices through Search
        Engine Social Media Optimization


    The control of information is embedded in the cultural politics and institutions

that regulate access to information.  In its most basic form, communication is a

practice of enabling the exchange of information. Websites have become one of the

primary ways that people access information; however, most of the access is

mediated through search engines and social media platforms.  Communication

research has explored the role of these platforms as gatekeeper and critical studies

have attended to the ideologies of search algorithms. From the advertising and public

relations industries, advice has emerged to communicators on how to make their

content accessible through these gatekeepers using optimization strategies. Critical

communication studies have not examined the relationship between these

optimization strategies that are used on actual webpages and access to information.

This dissertation seeks to fill that gap by asking how optimization techniques are

structured in online communications to increase access to information. How do the

techno-infrastructure of HTML and embedded assumptions shape communication

online? Where are points of resistance and opportunities for influence? How does this

differ from historic methods of preparing communications to be discovered and retrieved? This dissertation explores the history of search engine and social media optimization through a media archaeological approach to uncover the invisible infrastructures, habits, and assumptions that surround and shape communication online. By utilizing a media archaeological analysis, I will be able to situate the multi-layered practices in the form of optimization strategies. Critical histories are meant to be emancipatory. This dissertation is important for communication studies to develop an understanding of how we enable and influence discussions in our current digital cultural moment and to provide strategies for how communications are accessed.

CURRICULUM VITAE

NAME OF AUTHOR:  Karen M. Estlund

GRADUATE AND UNDERGRADUATE SCHOOLS ATTENDED:

University of Oregon, Eugene
University of Washington, Seattle
Reed College, Portland, Oregon

DEGREES AWARDED:

Doctor of Philosophy, Communication and Society, 2021, University of Oregon
Master of Library and Information Science, 2005, University of Washington
Bachelor of Arts, Classics, 2001, Reed College

AREAS OF SPECIAL INTEREST:

Communication and Information Technologies
Copyright
Information Access
Media Studies

PROFESSIONAL EXPERIENCE:

Dean of Libraries, Colorado State University, 2019-

Associate Dean for Technology and Digital Strategies, Penn State Libraries,
    The Pennsylvania State University, 2015-2019

Digital Scholarship Center, Head, University of Oregon Libraries, 2012-2015

Digital Library Services, Head, University of Oregon Libraries, 2011-2012

Digital Collections Coordinator, University of Oregon Libraries, 2007-2011

Digital Technology, Interim Head, J. Willard Marriott Library, University of
    Utah, 2006-2007

Adjunct Professor, Department of Communication, University of Utah, 2006-
    2007

Technology Instruction Librarian, J. Willard Marriott Library, University of Utah, 2005-2006

Graduate Staff Assistant, The Information School, University of Washington, 2003-2005

ACKNOWLEDGEMENTS

Thank you, to my husband, Eric, who in the ten years working on this degree, did not ask how it was going for the final three years of writing. Thanks for your love and support, listening when I needed it, and for holding back your own curiosity and anxiety so as not to spike my own anxiety.

Thank you to my committee. I especially thank my advisor, Kim Sheehan, who never faltered in the belief that I could finish and provided encouragement along the way, as well as helping me articulate the "why." Thanks for supporting me to pursue using a media archaeology analysis after that that philosophy course that blew my mind, and thanks to Colin Koopman for teaching that course on politics of information and introducing me to media archaeology. Thank you to Bish Sen for helping me see research as a way to bring about positive change and reminding me that there is more work to do. Thank you to Seth Lewis for taking on student that you had never met.

Thanks to Radhika Gajjala and Carol Stabile for recognizing my potential and reminding me that as much as is in my head, I haven't done the work until it leaves my head.

Thank you to Evviva Weinraub Lajoie, for friendship, helping me keep the librarianship (day job) field contributions in motion, and for creating space for me whether to write at your house or space to just be me. Thank you to Brandy Karl who kept me on task with daily reminders of writing encouragement. This dissertation would not have been completed without Brandy. Thank you for Carolyn and Scott

Cole for your friendship and Scott's eagle eye and advice in editing; however, I still like semicolons.

I thank my parents, John and Peggy Mahon, who instilled a curiosity in me and appetite to never stop reading and questioning. And to the memory of a family car ride in 1996 when we discussed the pluses and minuses of pursuing a PhD, as teenager me contemplated life goals. Took a while, but I did it!

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER I

INTRODUCTION

"[T]he overwhelming propensity of most people is to invest in as absolutely little effort into information seeking as they possibly can (Bates, 2002)."

As Americans increasingly cite the Internet as their primary way to keep informed and share ideas, [1] studies of new media communication in the online environment are increasingly important to understand the structures that govern how communication occurs online. The way to access much online content relies on large online ICT (information and computer technologies) commercial giants such as Google, Microsoft Bing, Facebook, and Twitter. The role of "gatekeeping" has transitioned from a print environment where news and publishers determined content and libraries and archives selected content to one where online services of search engines and social media sites become the "gatekeeper" for accessing information. The ease, speed, and amount of information available creates an illusion of direct access to information and lack of a gatekeeper. In attempts to unveil this illusion in the online environment, two primary socio-technical communication investigations have emerged: 1) critical analyses of the algorithms that the search and social media companies use to promote content, and 2) analyses of the strategies that users employ to work with the gatekeeping systems of the day to make their content more visible in these venues, search engine and social media optimization.

---

[1] A 2014 Pew Internet Center study found that 87% of Americans felt the Internet made them better informed. 75% felt better informed about national news. Only 49% felt it made them better informed about civic and local government activities. (*American Feel Better Informed Thanks to the Internet*, 2014).

In this second area of investigation, studies have focused on the active processes used for employing search engine and social media optimization strategies and the effectiveness of such strategies. SEO and SMO strategies are ways that content creators attempt to influence how and where their content appears in these gatekeeping tools. These techniques are often categorized under Search Engine Marketing (SEM). [2] Through these processes, various strategies and tools are employed within HTML pages with the goal that users will click on the search engine result or social media posting and go back to the organization's HTML site. Little attention in the communications field has been given to the institutions and technical constructs for content creation process and methods and tools used to either speak to or game the algorithms to further a message. A critical examination is needed to place the historical context of SEO and SMO strategies within the larger communication and online environment.

This dissertation seeks to identify the political and practical infrastructure surrounding access to communication of information online. By concentrating on the content creation and the optimization strategies used to make information available through online platform gatekeepers, this project hopes to identify opportunities for counter voices. This dissertation focuses on the optimization strategies of SEO and SMO

---

[2] This project does not largely include Search Engine Marketing (SEM) paid-for services, such as Google AdWords or Bing AdInsights, because they are governed by different technical structures than organic or native search engine results. With the paid services, different algorithms used to return results and additional mechanisms including payments and bidding processes are used to determine how and what content for the ads is displayed. Overlapping techniques or tools that are used for both paid-for and organic SEO and SMO techniques, such as with keyword placement or Google defined quality measures for prioritizing content, will be addressed.

as the primary tools for exposing content in these environments.[3] This history of

optimization strategies will explore the structures that enable optimization and that

control access to content.   The following research questions guide this inquiry:

**RQ1: What is the historical development of SEO and SMO strategies?**

**RQ1a:  What are the topoi in these practices?**

**RQ1b:  What is the interplay with changes in proprietary algorithms over time?**

**RQ2: How has the development of SEO and SMO strategies been actualized in HTML practices for major persuasive information industries?**

**RQ2a: How have the strategies been implemented in newspapers' online presence?**

**RQ2b: How have the strategies been implemented in political candidate websites?**

**RQ3: How have SEO and SMO strategies shaped communication online?**

To address these questions, this dissertation employs a historical media

archaeology approach. Media archaeological approaches are historical analyses that may

use quantitative or qualitative methods. A media archaeological approach is especially

useful in examining current phenomena and placing them within a larger historical

context to aid in understanding this current environment. Following from Foucault's

framework for an archaeological analysis, the structure and rules are emphasized over

content, intent, and the "creative subject" (Foucault, 1972).  By using a historical media

---

[3] This project is focusing on content that creators want to openly expose. There is a recognition that not all content on the Internet is available through search engines and social media platforms and also that not all content wants to be found and communicated through these intermediaries. Technological issues of actionable code/scripts and databases may also prevent content from being discovered through these services.

archaeology approach, I will investigate both the conceptual ideas and technologies surrounding search engine and social media optimization, examining approaches and strategies within the structures of the HTML webpage and the role of search engine and social media platforms to influence practices.

As an object of study, I will examine the SEO and SMO practices in HTML webpages. The scope of study is limited to on-page SEO and SMO techniques and for the article or issue page. For example, the *Los Angeles Times* will be examined for its practices to increase exposure within search engine results and social media platforms looking at specific HTML renderings of a newspaper article. Examples of off-page techniques that will not be examined include the number of external websites linking into pages and content created natively within social media platforms. Practices examined include application of `<meta>` tags and semantic web, hyperlink analysis, and structured content. The media archeological examination will include three sources of data: 1) instruction manuals and guidebooks on SEO and SEO strategies; 2) select *Los Angeles Times* article webpages harvested and available from the Internet Archive's Wayback Machine;[4] and 3) U.S. Senate political candidate webpages harvested and available from the Library of Congress United States Elections Web Archive.[5]

---

[4] https://archive.org/web/

[5] https://www.loc.gov/collections/united-states-elections-web-archive/

## Optimization Overview

"Optimization" is defined as finding the best and most efficient process as close to "fully perfect" as possible.[6] Through techniques of SEO and SMO, the content creator employs strategies to promote access to their content. Although optimization techniques are intended to identify the most effective way of making content accessible, the techniques are heavily regulated by the search engine and social media corporations. Because it is in the interest of these corporations to have content structured for their services, they usually provide helpful and detailed guidelines on techniques and standards for their systems.[7]

There is also the practice of extreme optimization, called Black Hat, which occurs when web content creators game the systems developed by the gatekeepers in order to promote webpages using the rules of the structure that may have little to do with any actual content. These tactics are touted by the web and news industries as illegal and/or ineffective (Boutet & Quoniam, 2012). There is debate in the web community about the appropriate levels of optimization to employ; however, the side of the "right" is typically associated with the search engine and social media corporations. The context for how the coding structures, tools, guidelines, and institutions have historically enacted these regulations of information is important to understanding how information can be accessed through these modern gatekeepers.

---

[6] http://www.merriam-webster.com/dictionary/optimization

[7] Google, *Search Engine Optimization Starter Guide*: http://www.google.com/webmasters/docs/search-engine-optimization-starter-guide.pdf; Bing, *SEO Analyzer*: http://www.bing.com/toolbox/seo-analyzer; Facebook, *Content Sharing Best Practices*: https://developers.facebook.com/docs/sharing/best-practices; Twitter, *Twitter Cards*: https://dev.twitter.com/cards/overview

Optimization is different in search engine and social media platforms. Search engines attempt to provide access to the "right information" that matches a user's query; whereas social media platforms are less concerned about "right information" and seek to provide a good user experience. Despite these separate goals, because of the structure of the web and HTML documents and similar work involved on the content creator's end, which often overlaps, it is useful to examine the practices as a holistic set of techniques for content creators to make their content available through the primary gateways to information on the web.

## Search Engine Optimization (SEO)

Search Engine Optimization is the set of strategies and practices used to influence placement and ranking on search engine results pages (SERPs) for indexed web content. Indexing content to make it readily accessible has been a central feature of the Internet. The first search engine, Archie, was designed in the late 1980s to search content in ARPANET (Savetz, 1993). Since then, the web search engines have continued to evolve and act as gatekeepers to the content on the open World Wide Web (Introna & Nissenbaum, 2007). The indexed web, including Google, Yahoo, and Bing, comprised around 50 billion pages of web content in 2014, and Goolge has 92% of the market share(*The Size of the World Wide Web (The Internet)*, n.d.). Although not all web content is available via the indexed web and listed on SERPs, a high ranking result on a SERP is an essential part of making web content accessible with many websites finding up to 64% of traffic coming from organic search results (Zeckman, 2014). Strategies to increase the likelihood of a high-ranking result for web content have changed in conjunction with

changes to search engine algorithms and media formats. Search engines have also

responded to SEO strategies that they consider harmful, such as Black Hat strategies, and

modified their algorithms to retain control of what appears in search results (Malaga,

2008a).

**Brief SEO History**

In order for search engines to take advantage of the information provided on the

web, they selected elements and practices in HTML to query and return. Each search

engine needs an algorithm to function. The most famous of these is the Google

PageRank® algorithm:

$$PR(A) = (1-d) + d \ (PR(T1)/C(T1) + ... + PR(Tn)/C(Tn))$$

**Figure 1.1.** Snippet from original PageRank algorithm.[8]

In the late 1990s search engines were still trying to categorize the web as well as

provide search functionality. Google's launch in 1998 changed this behavior, and the

search box that is now ubiquitous became the primary tool of search engines.  A new

industry arose around helping web content creators receive a better ranking on search

engines around 1996, and the term "Search Engine Optimization (SEO)" was coined to

describe these strategies (Sullivan, 2004).

Many of the changes in SEO strategies have been subtle but important and are in

direct response to rules set by search engines. Each search engine has different algorithms

---

[8] (Brin & Page, 1998)

and may promote slightly different HTML elements and practices, yet function out of similar principles. The most significant public changes to search engine algorithms regarding SEO all involve Google because Google is the only engine that makes major changes public.  Most Google changes are primarily in response to what it considers illegal or unethical behavior. For example, in 2009, Google discontinued emphasis on `<meta>` keyword tags because they deemed that too many content creators were using them to mislead the search engine and plagiarize or subvert the work of competitor information or commercial sources ("Google Does Not Use the Keywords Meta Tag in Web Ranking," 2009).  Some changes are also in response to technology and device developments.

In 2012, an editorial on *Forbes.com* entitled, "The Death of SEO: The Rise of Social, PR, and Real Content," created a storm of comments, speculation, denial, and support. The article received over 525 comments and was the #1 trending article for 2012 before Forbes decided to lock the comments (Krogue, 2012). This editorial and many of the changes that followed for SEO was not a "death of SEO" but rather an integration in SMO and a realization that SERPs started to give preference to social media content in results, as well as the drift of users accessing content directly through social media platforms.  SEO practices and the SEO industry continue to thrive.

**SEO Basics**

Although not all SEO will affect each search engine in the same way, general SEO expert advice recommends following Google SEO strategies, which will thereby affect rankings with other search engines, but paying attention to the subtle differences

**Table 1.1.** Adapted from "Snapshot of Major Changes in Google Algorithm History".[9]

| Date | Updates | Purpose |
|---|---|---|
| 2015- | Unnamed; referred to by the SEO industry as Phantom 2 | Rank by the quality and "truthfulness" of a webpage (Dong et al., 2015) |
| 2015 | Mobile Friendliness | Increase rank on main Google SERP if webpage is mobile-friendly |
| 2014 | "In the News" Box | Blogs and non-traditional news media included in News Search Results and "In the News" box on main Google SERPs |
| 2014 | **Pigeon** | Focuses results on a local geographic level to provide more relevant results to users |
| 2013 | **Hummingbird** | Builds on earlier knowledge graph integration and allows for semantic web and knowledge graph search |
| 2012 | **Penguin** | Address web spam and sites not following Google's Webmaster quality guidelines |
| 2011 | **Panda** | Address content and link farming and high-ad sites. In direct response to actions from Overstock.com and JC Penney's, which took over Google Search results for many consumer goods. |
| 2010 | Social Signals | Customize search results based on social media and network of user |
| 2010 | **Caffeine** | Google search infrastructure redesigned for fresher content; no effect on SEO. |
| 2009 | Real Time Search | Emphasize news and social media |
| 2009 | Keyword Trust | `<meta>` tags for keywords no longer factored in results |
| 2008 | Google Suggest | Shows user popular search string options as they type in the search box |
| 2005 | **Jagger** | Targeted at poor quality links and link farms |
| 2003 | Boston – **Fritz** (monthly updates) | Changes to index, supplemental index, treatment of links and hidden links and text |

---

[9] See ("Google Algorithm Change History," 2015, "Timeline of Google Search," n.d.)

that Bing and Yahoo may utilize increased access to content (Sherrod, 2010; Smarty,
2009; *The Differences Between Google & Bing SEO Algorithms*, 2014). The historic and
current framework necessitates that SEO activities take place with the HTML framework,
which means that much multi-media content such as images and videos are not the focus
of SEO activities beyond the tags available within HTML code.[10]  Most SEO strategies
can be either manually applied or automated / scripted.

In addition to the techniques outlined in Table 1.2, integration on social media
sites and structural considerations, such as Google's and Bing's design and mobile-
friendly preference rankings are also used for SEO. It is also important to notify search
engines to crawl your site. Many SEO experts recommend creating a sitemap that lists all
the pages and links in your website and submitting that to each search engine (West,
2012).

**SEO – The Dark Side**

In the web industry, these strategies and techniques also have variant practices that are
labelled as "white hat" – correct, good, honest, and proper methods – and "black hat" –
malicious, sneaky, and false methods. The notion of "fully perfect," "good," and "right"
permeates how web content creators are supposed to act and follow the rules set by the

---

[10] With increasing search engine and social media queries for things like color search and facial
recognition, it will be interesting to see if these techniques affect content creators in their quest to have
information be found. The same use-case for enhancing content to be found does not appear to have
modified current strategies. See Google color image search (Tanguay, 2009); Facebook Facial
Recognition, (Taigman et al., 2014).

**Table 1.2.** SEO Techniques.[11]

| SEO Technique | Description |
|---|---|
| Link Building | The process of encouraging relationships for others to link back to your site (i.e., inbound links). – *Off-page technique* |
| Link Farming | Linking to other sites (i.e., outbound links). |
| `<title>` tag | Craft a succinct title with keywords toward the front that is less than 54-75 characters. Note: this title is not the same as a title of the content. |
| `<meta>` description tag[12] | The description is less used in ranking and more often a tool displayed as part of the SERP. The description tag should include keywords in the title and be less than 160 characters. |
| `<meta>` tags[13] | `<meta>` keyword tags were once highly utilized but are less used. Additional metadata elements may be put into customized tags for semantic web or specific applications such as Google Scholar, with information such as the academic journal from an article's citation.[14] |
| Keywords | Ensure good keywords are in all the headings `<h1>`, `<h2>`, `<h3>`, etc. tags on a page and through `<body>` text. |
| Rich Snippets | Utilize schema.org and semantic web markup embedded in content on the page, such as news, events, and media. |
| Cloaking and Doorway Pages | Create pages for indexing that redirect to a different page of content. *Black Hat Technique* |
| Designed URLs | Use URLs that are expressive and descriptive of the page content, are short, and use hyphens between words. Assign "canonical" URLs when duplicate page content exists, such as a print version. Use full URL addresses for internal site links. |

search engines and social media corporations. Activities that do not conform to the rules

are labelled as "black hat." These activities include cloaking, link farms, hidden code,

---

[11] Adapted from: (Fishkin, 2015b; Killoran, 2013; Malaga, 2008; West, 2012; Yalçın & Köse, 2010)

[12] E.g., `<meta name="description" content="This dissertation is about SEO and SMO." />`

[13] Basic HTML metatags in addition to description are "author" and "keywords" http://www.w3schools.com/tags/tag_meta.asp; e.g., `<meta name="keywords" content="SEO,SMO,search engines,media archeology,gatekeeping" />`; `<meta name="author" content="Karen Estlund" />`

[14] https://scholar.google.com/intl/us/scholar/inclusion.html#indexing; e.g., `<meta name="citation_author" content="Estlund, Karen" />`; `<meta name="citation_journal_title" content="The Journal Annual Review" />`

and door pages among others (Killoran, 2013; Malaga, 2008). "Black hat" activities are also often associated with SEO strategies that utilize automation or scripting.

There is little interrogation of what actually constitutes "black hat" methods and what are simply ways of making content more accessible that wouldn't be otherwise (such as creating doorway pages for heavy multimedia content which cannot be as easily queried as text).[15] Black hat techniques are often conducted by spammers and other counterfeit companies (Israel et al., 2013; Lu & Lee, 2011; Wang et al., 2011, 2014); however, going against the algorithms isn't always malicious. A well-researched content on a webpage, with many citations and links to those sources, but those sources having pre-existed don't link to the item, would be considered link farming, and the page could be banned from search engine results. Advocates of these types of activities have tried to encourage active SEO that does not necessarily adhere to all of the rules defined by search engines and challenges the assumptions that not following the rules is an unethical activity (Boutet & Quoniam, 2012; Fishkin, 2008).

## Social Media Optimization (SMO)

Social Media Optimization (SMO) for web HTML pages is the process of making HTML content social-media ready so that it can be integrated into social media feeds. The integration is typically started by either a representative for a content creator or a

---

[15] One of the landmark changes in Google search algorithms occurred in 2006 after BMW's German website created doorway pages which had the search terms and content that a search engine would retrieve but then redirected to a multimedia site. Google banned BMW from its search results for a time due to the action, as well (Malaga, 2008).

content recipient linking to the HTML information. The relationship between what is web content and what is social media content is also blurred through this integration as content is formulated for both traditional websites and social media platforms and the user's role in selecting and contributing the content plays an important role in the distribution of information (Gerlitz & Helmond, 2013). Like search engines, social media platforms can act as gatekeepers to providing access to information; however, unlike search engines the information is provided by a combination of social relationship recommendations from one own's circles and groups and the algorithms from the social media platform feeds. The highly localized and personalized information in these feeds has been studied as noteworthy of the ever-narrowing exposure to diverse content for people using these platforms (Hermida et al., 2012; Messing & Westwood, 2012).

Unlike SEO, SMO strategies are less about getting the algorithms to process and rank the content highly and more about making the content the correct format and, most importantly, creating content that appeals to a person so that they are likely to link to it in their social media feeds (Foster, 2015; Rayson, 2013). The emphasis on providing metadata, rich-formatted content, efficient sharing methods, and viral content are the central precepts of social media optimization.

**Brief SMO History**

SMO strategies came to the forefront of marketing and communication efforts around 2006. Several marketing websites point to a blog post by Rohit Bhargava from the Influential Marketing Group as the start of the SMO usage. In this post, Bhargava outlines five tenets that should be considered for SMO: 1) increase your linkability, 2)

make tagging and bookmarking easy, 3) reward inbound links, 4) help your content travel, and 5) encourage mash-up (Bhargava, 2006).  SMO strategies were quickly taken up by the SEO community and added to many SEO guides (Fishkin, 2015a). SMO strategies have not changed dramatically over time but have aggressively continued to focus on basic structural information to make linking and mash-up easy and content that tries to appeal to users to get them to forward the HTML content to social media. These strategies have shifted toward automated efforts and the work of bots to increase prevalence in social media platforms (Allen, 2016).

The changes that have precipitated most SMO modifications are a result of social media platforms prioritizing the type of content they choose to highlight and hiding content that they determine is too aggressively marketed. In a different manner than search engines, the social media platform is more concerned with advertising interfering with the social experience and not undercutting their advertising revenue rather than a focus on providing relevant or accurate results in a feed. Social media platforms hide content from feeds that they consider too promotional or spam. Social media platforms have also attempted to reduce "fake news," but efforts to eliminate the biases have largely been ineffective (Levin, 2017). Because social media optimization strategies are about structure and helping the social media platforms consume and display the HTML content, they can be used in any type of page or content. The existence of such strategies also does not have a bearing on whether the content is accurate, relevant, or fake news. The social media platforms rely heavily on user interactions and judgement on content formatted to feed well into their systems.

**Table 1.3.** Basic SMO Strategies.

| SMO Technique | Description |
|---|---|
| Open Graph `<meta>` tags[16] | Use the semantic web Open Graph protocol to catalog information on page. Use primarily for Facebook. |
| Twitter Card `<meta>` tags[17] | Use twitter specific metadata schema to format elements and provide metadata for twitter formatted "card" to include media and rich text elements with shared content from HTML pages.  Use for Twitter. |
| Social buttons (like and share) | Use HTML buttons on site that forward user to social media platform to like and or share content. Use for multiple social media sites. |
| Headline / Title Optimization for Click Through Rate (CTR) | Test multiple titles using automated tools or A/B tests with users to identify titles that will have high CTRs with social media users to return users from social media to home HTML pages. |
| Share-able Image | Create an image and share on HTML page formatted specifically for social media platforms.  Size the image(s) depending on the platform and reference in the Open Graph and/or Twitter card `<meta>` tags. |
| Bots | Bots are automated methods of imitating user behavior to post or communicate with users. |

**SMO Basics**

The two primary social media platforms that provide guidance and are generally

thought of as best practices to follow are for Facebook and Twitter, who have the largest

current social media shares. SMO on-page techniques for HTML webpage content focus

in a few main areas. Tools and content management and publishing platforms often assist

in the creation of the structured metadata that may be needed for easy integration into

social media platforms. However, individualized author strategies may also be required.

The editor-at-large of *Upworthy* notes making writers come up with 25 titles for each

---

[16] See http://ogp.me/; e.g. `<meta property="og:title" content="SMO Open Graph" />`

[17] See https://dev.twitter.com/cards/markup; e.g., `<meta name="twitter:creator" content="@estlundkm" />`

post and then running them through CTR (click through rates) tools to test effectiveness (Mordecai, 2014).

## Signficance of the Study

Communication scholars have studied the gatekeeping function of search engines (Granka, 2010; Introna & Nissenbaum, 2007; Mager, 2012) and the targeted and selected content choices available in social media feeds and reception (Hermida et al., 2012; Khang, Ki, & Ye, 2012; Lovejoy & Saxton, 2012).[18]  These studies primarily focus on the receiver of information or the actor as the gatekeeper. They have been significant in exposing the bias in content selection and autonomy of the receiver; however, they have not explored the everyday strategies to challenge these barriers.

Critical studies of search engines and social media have focused on the ideology of search engines (Fuchs, 2012b; Mager, 2013, 2014; Noble, 2013; Rieder, 2012). These studies range from interrogating algorithms within the institutional contexts of larger economic and technology cultures to feminist critiques that explore how the opinions are proliferated through the lens of a western white male perspective.  Fuchs' work has also focused on the content creator for search engines and social media platform through a political economic examination of how search engines and social media platforms utilize user-created content and labor for their profits whereby a Marxist exploitation of surplus

---

[18] Communication studies have also focused on the role of the user as activist and news generator and social relations using social media (Khang et al., 2012). These studies are an important contribution to both user creation and reception of content in the social media environment; however, they do not address the optimization and retrieval strategies, which are the focus of this project.

value ensues (Fuchs, 2010, 2012a). Gerlitz and Helmond have contributed work exploring the role of Facebook and social media in the larger online environment, particularly the use of Open Graph metadata[19] and relationships with the Facebook "Like" button to demonstrate the economic ideology embedded in each user interaction (Gerlitz & Helmond, 2013). Noble's work is noteworthy for examining the type-ahead feature in Google's search box with both the limitations and the prejudices enhanced through this search feature (Noble, 2013) and has been used as a call for more government regulation of these technical gatekeepers.

Some research exists on the effects of mass SEO link building to receive a particular search result to a Google query, known as Google Bombing (Bar-Ilan, 2007; Tatum, 2005). This research, however, is not extensive and focuses primarily on motivations. The limitations of these studies are in attributing the power in these actions to the user alone. The studies do not provide a critical investigation in the socio-technological and cultural structures that constrain and enable these activities. Without understanding the broader institutional approach, Google Bombing becomes simply a one-off act.

In the communications field, research on SEO and SMO is focused primarily in advertising and public relations research focusing both SEO and SMO strategy and studies on ROI (return on investment) and customer loyalty and perception (Berman & Katona, 2013; Lipsman et al., 2012). SEO and SMO are also examined within the

---

[19] Metadata is a set of descriptive data about another data or content source. Metadata can be descriptive, technical, or administrative in functionality and applied manually or automatically to the original content source.

computer science field where SEO is referred to as "adversarial web search" and is treated as hostile to the algorithms and a problem to be fixed (Castillo, 2010; Malaga, 2008). On-page social media optimization, within computer sciences, is typically studied within the context of a particular set of tools for the semantic web and network analysis (Kinsella et al., 2011; Sizov, 2010).

This inquiry is limited by U.S. and English-language focused webpages and manuals in order to continue the early studies on gatekeeping within the context of U.S. politics and newspapers. This study is also primarily concerned with the premise of communicating more widely and making content accessible amid the gatekeeping technologies of search engine and social media sites online. Some countries around the world govern the types and characteristics of content that can be returned in search engine results or displayed on social media sites. Recent court cases in the United Kingdom and the European Union have focused on aspects of privacy and the "right to be forgotten," which may specify the content that search engines are allowed to display (European Comission, 2014). Additionally, countries such as China have been restricting content for decades (Goldsmith & Wu, 2006). The U.S. does not at this time and historically has not legally mandated what can be displayed in search and social media results, which provides an unobstructed base for this analysis. A future study should seek to examine techniques for search engine and social media optimization within these legal and governing frameworks in non-U.S. contexts.

**Dissertation Overview**

The remainder of this dissertation is structured as follows: two background chapters, CHAPTERS II and III; a communication systems and diagramming overview of the technologies, CHAPTER IV; an overview of SEO and SMO strategies from instruction manuals, CHAPTER V; a chapter focused on SEO and SMO in newspaper articles online using the *Los Angeles Times*, CHAPTER VI; a chapter focused on SEO and SMO in U.S. Senate political candidate websites on election issues, CHAPTER VII, and a concluding chapter, CHAPTER VIII.

CHAPTER II reviews the interdisciplinary theoretical background of communication and information system modeling, politics of information, and gatekeeping studies to examine structural and institutional practices used to create, inform, and react to optimization practices, as well as the sociotechnical systems for communication and access to information. CHAPTER III provides an overview of the methodology and reviews the use of and rationale for a media archaeological analysis as the framework with a historical document analysis, as well as the selection of content for analysis. This chapter also provides background on web archives and the process for accessing the archived webpages needed for the analysis.

CHAPTER IV: The communication system environment for SEO and SMO provides a diagraming analysis of the communication processes used in the context of search and making information "accessible." As topoi are explored in a media archaeological analysis, part of the process is to place cultural phenomena in the context and evolution of pre-existing media and mechanisms. This chapter provides the context

to be drawn on for the base of the inquiry into search engine and social media optimization, as well as outlines of the processes used in the current technologies.

CHAPTERS V through VII present the major findings of the dissertation through a historical method and document analysis employed in the media archaeological process. CHAPTER IV: How-To Manuals and Instructions for SEO and SMO will use instruction manuals and how-to guides for the strategies historically recommended to employ. The optimization strategies will each be discussed based on changes over time and with specific attention to the structure of HTML and the expertise needed to comply with the strategies. CHAPTER V: News Stories use of SEO and SMO Strategies in the *Los Angeles Times* will review archived HTML webpages of newspaper articles for the presence of suggested SEO and SMO strategies.  CHAPTER VI: U.S. Senate election political candidate webpages will review election issue webpages in close election races for the presence of suggested SEO and SMO strategies. These chapters will also address points of success and failure in the adoption of optimization strategies and evidence for points of transition in techniques.

The conclusion will review the topoi identified in the media archaeological analysis and the impacts on writing and communication on the web. It will also compare the expected outcomes from the instruction manuals with evidence found in news and political candidate webpages. Recommendations for future study will also be addressed.

CHAPTER II

THEORETICAL FOUNDATIONS & LITERATURE REVIEW

This dissertation takes a critical historical approach to examine institutional structures within communication practices of search engine and social media optimization. To answer the research questions, this chapter provides a contextual and theoretical review that draws on an interdisciplinary cross-section of theories from communication, philosophy, sociology, and information science. Beginning with new media / digital communication theory, grounded in foundational theories of communication models, this section also provides a historical lens for the construction of technical systems that form web communication technologies on which search engine and social media optimization practices are enacted.

Following the discussion of new media and digital communication theory, this chapter takes a critical overview through politics of information and critical code studies to examine the institutional and power structures built into systems of information. Gatekeeping then provides a framework for how media organizations have historically shaped and provided access to information and what can be communicated. The chapter concludes with a review of gatekeeping studies in the online environment that provide a basis for understanding the power relationships that govern content exposure and access to information in the digital environment and information retrieval systems. These theories and background are important for understanding the historical context and implications for how search engine and social media optimization practices are actualized in historical and contemporary online environments.

**Communication and Information Theory Models**

In order to understand digital media as an object and a set of processes, it is necessary to review early theories of communication technology models that provided a basis for systems of communication, as well as theoretical models of communication exchange. This section reviews the historical roots of new media theory through information and communication models developed in post-WWII telecommunications and cybernetics and concludes with communication models for the processing of information from the Internet. The communication models of the mid-twentieth century that were developed alongside early computing systems aimed at increasing the quality of communication transmissions. They illustrate an emphasis on inputs and outputs for information that is transmitted and received.  As such, these models propagated a theory of information as an object of transmission. The quality of the transmission for effective communication was something that could be engineered and configured within the constructs of the mathematical and engineered system. These theories laid the groundwork for how digital information is perceived, valued, and regulated, as well as the architecture that informed the initial building blocks of the Internet and World Wide Web.

**A Mathematical Model of Communication**

During WWII in the U.S. and Great Britain, mathematicians, engineers, and scientists worked together on war and anti-war technologies. One of the major research and development institutions in the United States was Bell Labs in New York City, which was the research hub for the Bell telephone system. During the Cold War, Bell Labs

continued to function as a hub for ground-breaking research with the increased federal

government funding for scientific research, which created an influx of new information

technologies (Rogers, 1997). Claude Shannon, a researcher in Bell Labs during these

periods, proposed a new model for communication publishing two papers, "A

Mathematical Theory of Communication" later combined into a book with introduction

by Warren Weaver that became the basis for modern digital systems (Nahin, 2013;

Rogers, 1997). For Shannon, communication systems and the reproduction of an accurate

message sent at one point and received in another occupied his research (Shannon, 1948).

He developed a one-way model of communication and information theory of sender and

receiver, which because of the easily generalizable categories and the introduction from

Weaver to expand uses of the model, was quickly adopted across disciplines as a way to

explore communication (Rogers, 1997).
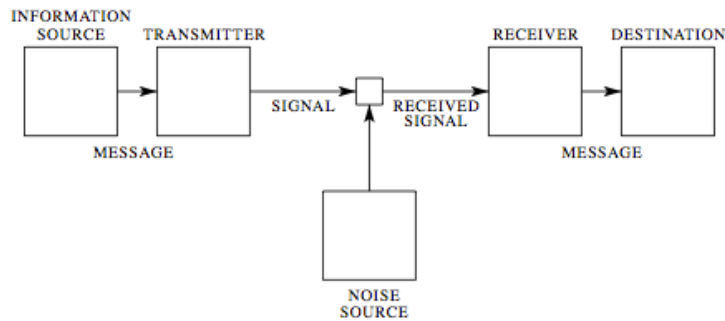


Fig. 1 — Schematic diagram of a general communication system.

**Figure 2.1.** Shannon's Mathematical Model of Communication (Shannon, 1948).

This linear model of communication is a sender/receiver model where the goal of

the communication system is to avoid errors, reduce the "noise," and produce as clear a

message as possible (Shannon, 1948). In order to accomplish his task of noise reduction,

Shannon introduces three concepts: 1) In Shannon's model, he reduced information to a binary set of 1's and 0's that could theoretically be transmitted along electrical current (Shannon, 1948, p. 395). Although the terms bit and binary had been used previously with systems, Shannon's interpretation of 1's and 0's moving through the channel as information was first introduced in a "Mathematical Theory of Communication" (Nahin, 2013). 2) The use of Boolean logic and error detection through redundancy is central to the structure suggested by Shannon. This includes parity bit checking at the source end and parity bit checking at the receiver end of the channel with the Boolean exclusive OR (XOR)[20] (Nahin, 2013). 3) Shannon also provides an encoding of language where a "27-symbol 'alphabet;'" is used for English that adds the space as a character (Shannon, 1948). This leads to his discussion of relative entropy and redundancy. As Shannon explains for ordinary English, in words with eight letters or less, the chance for redundancy is roughly fifty percent, "This means that when we write English half of what we write is determined by the structure of the language and half is chosen freely" (Shannon, 1948). The transmission of binary values in communication systems, exclusive OR gates for error detection, and redundancy and relative entropy greatly influence the apparatuses of future communication and information technology systems

As a mathematical model, Shannon was explicit in the separation of meaning from communication. For Shannon, meaning is "irrelevant" to the problem of the engineering system (Shannon, 1948). Shannon sought to reduce noise in his system,

---

[20] The XOR presents a true result only when there is a difference in inputs is detected, i.e., one is true and one is false.

which would more allow for the more accurate message to be received; however, what is considered noise to the system becomes a fundamental question in modern information retrieval and digital systems. The reliance of the structure of the system as agnostic to meaning presents an interesting problem for new media studies. Tiziana Terranova argues Shannon's adaption of the information processing model in the system of communication causes a crisis of the meaning of information, whereby a group like "conscientious journalists" prioritize accuracy of information, but the engineer reduces information to a ratio of signal to noise (Terranova, 2004). This is important for this study as there is a clear tension of what constitutes the accuracy of information and where meaning is derived within the structures of communication systems. This is manifested in how search engines and social media platforms filter content and content creators use optimization strategies to make their content accessible.

**Cybernetics**

Cybernetics is a second model of communication that is integral to understanding the context of optimization strategies and algorithmic retrieval. The cybernetic model is a continuously evolving model of a communication system that closely follows a biological model (homeostasis) of learning and growth toward a more efficient process of communication (Wiener, 1961). In order to grow, the cybernetic system processes positive and negative feedback for its learning mechanism. Criticism of cybernetics point to the a-priori nature of negative feedback for growth as a major practical limitation for system design (Sutherland, 1975). The feedback requires an amount of self-recognition for the systems of how messages exchanged between two or more units influence each

other (Rogers, 1997). In Norbert Wiener's model of a cybernetic communication system, the system was self-learning and adapted to refine the message and output. Within cybernetics, because of the dependency on feedback in order to understand the information theory, the apparatus for information flow must be examined (Guilbaud, 1959).

Like Shannon's model, cybernetics does not consider the meaning or semantics of the message. Cybernetics involves a series of probabilities and likely selections but is less concerned with the accurate or correct message than Shannon. "What is of interest to our theory is the *choice*, the range of possible messages" (Guilbaud, 1959). Wiener noted that, in the case of information retrieval with large amounts of information, special effort was required to make that information available that required a familiarity of previous information for relevancy of any future retrievals (Wiener, 1961). Important for this inquiry is the reliance on previous information and choice are foundational to early models of search engine and social media algorithms. Through the ability to store massive amounts of information in the memories of computing machines, Wiener saw a way to use the outputs to do work in the world; for communication to benefit medicine and mental health (Conway & Siegelman, 2009). Cybernetics' description and ambition for information retrieval is useful for understanding the tactics of SEO and SMO, as search engines and social media platforms change the algorithms. From cybernetics, we gain an understanding of communication technologies relying on negative feedback and based on previous conditions of receiving the message with vast stores of information for ranking and making visible content according to their conception of an evolving and perfecting system.

**Digital Communication Models and the Internet's Foundations**

Shannon and the work of Cyberneticians are often pointed to for the beginnings of the Internet where their models of communication influenced the Internet's design. The structure of the communication between network nodes established the rules that online communications begin to take place. The early development of ARPANET, which set the foundation for the Internet, was a communication network between universities across the United States and resulted from a combination of military and academic interests that, although developed in the 1970's, was not widely used by the public until the 1990's (Schröter, 2012, p. 302). In order to facilitate communication, protocols were defined for communication across this network.

Although several alternatives were developed for communication across a global network of computers that established the beginnings of the Internet, the TCP/IP protocols designed by Vincent Cerf and Robert Kahn were adopted as the means for global communication. In their model, communication is transferred from HOSTS, which are composed of both source and destination computers, packet switchers and processes for the information to travel defined within the HOSTS (Cerf & Kahn, 1974, p. 637). Packet switching is used to enable information to travel in defined chunks and be reconstituted at the receiving end of a communication network. Cerf and Kahn describe communication processes between different networks through the use of GATEWAYS, which enables the communication between different networks through agreed protocols (Cerf & Kahn, 1974, p. 638). The introduction of gateways allowed for networks to maintain their own local protocols but provide a way to transmit standard expected formats, e.g., through an *internetwork header*, intercepted at the gateway, which allows for

communication between networks.  The TCP (transmission control program) handles the

processes of transmission at the level of the HOSTS; TCP enables breaking up of

information into processable chunks, with error checking and (e.g., checksum), and

reconstitution of messages at the receiving HOSTS. The IP allows for addressing of HOST

machines within the network (Fall & Stevens, 2011).

This TCP/IP protocol established by Cerf and Kahn set out the basis for

communication across the Internet; the flexibility of which parts of the standards are

communication networks that were necessary between machines in an external network

(i.e., Internet) vs. a local network were essential to the communication design system. As

the Internet developed into a network across global nodes, a layer-network approach was

adapted where different systems could implement different parts of the protocol but

essential features are shared between communications across the network.

| | | | |
|---|---|---|---|
| HOSTS | Application | Internet-compatible applications, e.g., the Web (HTTP), DNS, | |
| | Transport | Provides exchange of data between "ports" managed by application. May include error and flow control (e.g., TCP) | |
| ALL INTERNET DEVICES | Network (Adjunct) | Unofficial layer that helps accomplish setup, management and security of the network layer (e.g., ICMP, IPsec) | Network Layer |
| | Network | Defines abstract datagrams and provides routing (IP) | |
| | Link (Adjunct) | Unofficial layer used to map addresses at the network to those used at the link layer on multi-access link-layer networks (e.g., ARP) | Driver |

**Figure 2.2.** Layered Network Architecture from (Fall & Stevens, 2011, p. 14).

28

The layered network protocols do not specify how to present information on the Internet, however, and the World Wide Web provides the Internet with a way to communicate through a presentation layer, e.g., web pages. The beginnings of web development and protocols were defined by Tim Berners-Lee in initial definitions for HTML (Hyper-text mark-up language). Berners-Lee was a researcher at CERN who wanted to solve the problem of accessing and finding information on the Internet, and he proposed a "universal linked information system" (Berners-Lee, 1989). Berners-Lee was concerned primarily with staff turn-over at CERN and the loss of information from single experts that couldn't be shared with a wider community (Berners-Lee, 1989). The history of HTML mark-up is based on these strategies for finding information, communication between networks, and the focus on linking as central to the information knowledge environment and communication between communities. As the global network expanded, the World-Wide Web Consortium (W3C) was created to define and manage the protocols of online communication on the Web. HTML (Hyper-text mark-up language) was defined in documents on the early w3c.org site proposing a simple, yet expandable set of tags (mark-up) for documents on the Web. As HTML2 was rolled out, it became the standard for communication across the Web.

These early communication models and networking set the stage for communication studies to investigate how communication occurs through this new medium. The development of new media, digital, and Internet studies is a subfield of communication and media studies that developed alongside the technologies that it seeks to interrogate.

**Critical Approaches to Understanding Communication and Information Models**

In understanding these models of communication and information, this project takes a critical approach that ties culture and technologies as interrelated and dependent on each other. Contrary to a technological deterministic model, where technology is often viewed as neutral, follows a sort of natural evolution, and the technology directly effects society, in a critical view the elements of culture and society embedded within the constructs and infrastructure of technologies (MacKenzie & Wajcman, 1999). Central to the critical paradigm is the emphasis on social construction. This project examines the implementations of technological strategies, in the forms of SEO and SMO, and looks for linkages to prior media with the understanding that the technological cannot be separated from the cultural.

Through Raymond Williams' *The Long Revolution,* the role of changing technologies is viewed from a historic and contextual perspective and the shaping as a result of societal conditions toward a social construction rather than purely a technological deterministic model of communication technologies (1961). In Williams' discussion of technical changes in media, such as newspapers and books, both the technological advances of the printing and presses and transportation via railway led to increased distribution. However, the distribution cannot be viewed as separate from actors in society and cultural processes. "A large part of the impetus to cheap periodical publishing was the desire to control the development of working-class opinion, and in this the observable shift from popular educational journals to family magazine (the latter the immediate ancestors of the women's magazines of our own time) is significant" (Williams, 1961, p.56-57). In this example, Williams illustrates that the advances in

technology are not the only motivating factors of changes in periodical publishing. One

of the goals of this project is to provide an illustration of the technical changes so that

they can be further interrogated for cultural and societal influences and motivations.

Another important aspect of the critical approach is the possibility for change.

John Dewey saw the role of mass communication as a tool that could be used for

increased public participation and democratic ideals (1946). This approach is

characterized by a questioning and interpretive framework and also a sense of optimism

of change that could be possible through understanding, and in the case of Dewey,

pragmatic action. In Dewey's view of communication, the act of conversation and inquiry

is a necessary part of communication; communication does not exist outside of the social

needs to communicate and opinions are formed only in discussion as part of active

community life (Carey, 1989, p.81).

From James Carey's analyses of communication, we also the concept of ritual

communication in addition to and transmission of communications. In a ritual

communication environment, communication is embedded in institutions of society and

is continually re-inscribed yet adapts and evolves with periods of social change (Carey,

1989). Ritual communication as a socially constructed view should be explored for

significance and implications for communication. One of the needs for the social

construction and ritual communication view is the allowance for social change. Part of

the work in this examination is to examine the organizing principles of communication

and "to try to find out what other people are up to, or at least what they think they are up

to; to render transparent the concepts and purposes that guide their actions and render the

world coherent to them" (Carey, 1989, p.85). This project seeks to understand the

technical applications of SEO and SMO strategies and to identify the ritual re-inscriptions from previous forms of communications, ways information is organized and exposed, in order to identify places for change within the structures of online communication practices. In examining the implementations of SEO and SMO, this project looks to describe the "constellation of practices that enshrine and determine those ideas in a set of technical and social forms" (Carey, 1989, p.86). The questions posed, in this project, seek to first identify the practices as employed, in order that they may be further examined in terms of culture and society.

### Encoding / Decoding

Thirty years after the publication of Shannon's Mathematical Theory of Communication, Stuart Hall contends in his 1970's essay "Encoding/Decoding" that a problem with Shannon's model is that it assumes an equality of conditions on both the sending and receiving end of the message (Hall, 2006). Instead, he proposes that how the meaning is interpreted , the accuracy and what constitutes "meaningful discourse", however, still depends on a set of conditions and codes that may not be the same at both ends of the message. In Hall's encoding/decoding model, the codes that affect meaning include: frameworks of knowledge, relations of production, and technical infrastructure. The institutional structures, networks of production, organization, and technologies, are essential components to transmitting and receiving meaning from communications. The model of communication, thereby, necessarily begins in a cultural frame to send the message and is received in another cultural frame in order to be understand.

In examining the historical practice of search engine and social media optimization, Hall's model is especially useful to overlay the structural influences at the ends of the sending and receiving models that aid to a critical historical investigation. Hall contributed the idea of contextual interpretations for both the sending and receiving of communication, which is embedded in the social constructs and conditions of production on either end. In looking at online communications and interactive communication technologies, cultural studies and the Hall's model of encoding/decoding allows important connections to better understand new media (Shaw, 2017). In this project, the constructs and conditions are exposed for SEO and SMO strategies in the hopes that the cultural influences can be explored to ask the question of who defines the structures on what is good and how and what communications are accessible. "All activity is not resistive, of course, but neither is it complicit" (Shaw, 2017, p.600).

## Digital New Media Studies

Within the discipline of communication, digital new media theory has been used to explore the contemporary communication technologies, including digital and internet communication technologies, brought to the field (Morley, 2007; Silver, 2004; Sterne, 2005). The importance of "new media" in this study is to understand the roots of the digital as object and apparatus that both enables and limits contemporary methods of communication. The study of digital new media takes two main forms in the field of communication. One approach focuses on the study of the Internet as a transformative or transgressive medium, which allows for new forms of communication and interaction. An alternative approach explores the transition of specific communication media to the

33

Internet as a parent medium, such as television, video games, multi-media art, news, and advertising. Because of the interdisciplinary nature of communication, the definition of what makes something *new* varies, and there is not a universally agreed upon definition that permeates the field (Silver, 2004; Sterne, 2005). Where definitions of new media have succeeded in the argument of *newness*, scholars have concentrated on characteristics such as the ability for increased personal connection and social groups (Baym, 2010), sociotechnical systems (Haraway, 1987), re-usability and re-mixing (Deuze, 2006; Lessig, 2006), and values embedded in format (Sterne, 2012). Although new media has emerged over the centuries from the telegraph to television,[21] the crucial concept that defines these as *new*, and that I fix my definition of new media on, is when the medium elicits transformative views of reality and social practices. This type of newness is helpful in investigating the role of search engine and social media optimization by examining what makes the communication methods *new* and how that affects our understanding of communication technologies in society. Examining the structural issues, the conceptualization of communication via HTML documents and the interplay with search engine and social media platforms to surface content allows for an in-depth look at changing conceptualizations of communications provided by Internet technologies and standards, as well as the aspects that persist through technologies, "topoi."

"Topos" (topoi, plural) was originally developed by Ernst Robert Curtius for literary studies and adapted by Erkki Huhtamo for media studies. Essential to the idea of topos is that rather than emphasizing what is "new" with new media technology, topos

---

[21] See (Williams, 1975)

present the recurring cultural formulae built into systems. It's a way of looking at how what is new is shaped by what is already known (Huhtamo & Parikka, 2011). Topoi are useful for exposing the social and culture ways of knowing built into our systems that replicate over new forms of media. In this project, HTML pages will be examined for topoi that persist from previous forms of media and enforce functional attributes and information access in online communications.

**Hidden Mechanisms**

Another unique aspect of the digital media environment and its *newness* which is relevant to this study is the role of code and the digital medium. Code is sometimes viewable, sometimes readable, sometimes not. Sometimes, it can be viewed by using additional tools. Sometimes, the code is hidden by design of the program or, in the case of HTML, for security purposes to prevent things like code injection by hackers into JavaScript. Unlike a traditional print medium, where the code and technology in the inks and paper may be visible and yet still unknown to the user, the code that underlies digital media content may be completely hidden (Hayles, 2004; Kittler, 1995). Parts of the hypertext environment could be viewed if source code is rendered in a browser, but that is dependent on the exposure of the code as written (e.g., includes, database logic, and additional scripts may not be viewable). The full HTML new media environment is dependent upon a web browser on a hardware device to render the content. The hidden values, constructs, and structural framework that are working in this interaction between code, browser, and hardware are *new* in the digital media context.

In this dissertation, the code, within the HTML structure of webpages for optimization techniques, is a central object of study not solely as a processing tool but also to explore the embedded practices and ideas within the code framework. Critical code studies within digital new media studies have a theoretical precept that asserts a "performative, transformative, and mediating" function to code rather than merely an instrumental function (Marino, 2006). Code in its structure, ordering, and rules is an ideological expression that cannot be separated from the ideology in which it operates (Marino, 2006) and is inseparable from its operational context in a capitalist economy (Berry, 2011). The function of critical code studies is to make visible what has been made invisible and to demonstrate its cultural significance (Berry, 2011; Kitchin & Dodge, 2011; Mackenzie, 2006). By critically examining the rules of code and structured content, part of the invisible is made visible and can be evaluated for contributions to the overall communication practices.

Within the structure of code and particularly relevant to the study of search engine and social media optimization, is the role of algorithms in search engine and social media platforms that expose and promote content. "All code, formally analyzed, encapsulates an algorithm" (Mackenzie, 2006). Algorithms as a processing structure of code are akin to code with embedded ideologies and act much like institutions with a regulatory function on an individual's behaviors (Napoli, 2014). This interplay between code, algorithms, and how they are translated with hardware and software are points of investigation of this project that must be examined in order to expose to socio-technological infrastructure that shapes communication practices and access to information.

In this dissertation, I am concerned with the structure of the elements in the HTML, its roots, and its interaction with rendering applications and their algorithms. This approach is content meaning agnostic and focuses on application and presence of codes and values which are allowed within the HTML standard that can be used to promote and manipulate the logic of the search engine and social media platforms code and algorithms.

**Search Strategies and the Networked Document**

The hyperlink is an essential component within the study of new web media, as it increases the remix of content, removes or adds contextual interpretations, and defines a network of communication and relationships that exist both on a traditional theory of society and network communication through knowing and relationships and an algorithmic process by which scripts define the network and hyperlinked relationships between communicative content. "[S]ome Web pages work as electronic documents…while same pages more importantly point to document" (Gitelman, 2006, p. 128). This networked system of information with the use of hypertext, links, and efforts toward creating the semantic web as envisioned by Tim Berners-Lee have the effect of transitioning the Internet from a "Web of Documents" to a "Web of Data" (Park, Jankowski, & Jones, 2011, p, 147). The relationships between documents in the online environment provide an additional and novel approach to search that builds upon structures of cataloging, categorization, and indexing and elevates the position of the pointing document from previous media formats.  The purposes between these different

types, electronic documents and pointer documents may not be distinguishable until the page is examined.

**Remix, Variability, and Mutability**

Disintegration and remix are defining characteristics of digital new media digital that meet the threshold for reconceptualization of communication. The ability to break up, re-purpose, re-construct, and disable again in an efficient manner is a *new* feature of the online digital environment. This is fundamental in the design of the communication from the basic transmission of messages as defined in the packet switching protocols of the Internet. The implications of deconstruction and remixing have affected how a digital new media object is to be taken as a holistic object and re-conceptualized as a process (Deuze, 2006; Hayles, 2004; Jenkins, 2004; Landow, 2006; Manovich, 2005). Copying music and cultural media products in the online environment, and indeed the ease of copying and making a near duplicate of an original in a digital environment, is a historic change (Sterne, 2012). However, the more interesting question to me is how the copying and remixing aspect affect not only the legal framework but how content is seen, distributed and integrated into society as a communicative form and beyond the new creative cultural works[22] as lobbied by Lessig, Jenkins, and others.

Disintegration, remix, and linking has led to significant changes in how society views communication practices where the content is no longer whole but is part of a process and in the context of search results (search engines, data mining, and other

---

[22] Early pre-digital examples of this include Two Live Crew lawsuit for copying music within one of their releases.

activities): 1) it is inherently networked, and 2) exposed as part of algorithmically defined mechanisms. Assumption of context and attribution may be faint or completing missing in this new model of communication.

Another defining characteristic of new media for digital media content is the transitory and unstable nature of digital media content. Communication in previous contexts was typically either fixed (printed form, recorded, etc.) or transitory (speech on the phone). Although not all fixed communication content was integrated into an archival environment, the possibility was there.[23] Digital media content is both fixed and transitory. The ability to change content to modify and the tools needed to render content all lead to a new conception of what it means for communication to be finished. What does versioning look like in the digital environment? What is the "official message?" These questions are even further complicated by the ability for digital communications to provide tailored personalization of content. There are interesting questions for communication studies about which content is delivered to users based on these personalization measures (e.g., through newspaper home pages, search engine results, and more). The technological ability also exists to deliver different versions of the same content based on an individual user's computer and/or browsing settings. What is somewhat unique is that other than efforts to preserve web content by libraries and the Internet Archive, there is no check and balance on the historicity of the content. This problem was illustrated quite well by the George W. Bush administration and re-writing

---

[23] Certainly, degradation of nitrate film, fire, and other hazards have affected the ability for traditionally fixed content to be archived.

of White House webpages.[24] New theories and methods need to be developed to deal with the issue of fixed, yet variable communications. At what point are versions archived and how does that version affect the interpretation of the archival record?

## Politics of Information

### What is "Politics?" and a Politics of Information

For the purposes of this project, I follow a definition of "politics" akin to James Paul Gee in *An Introduction to Discourse Analysis* where that which is political is where human relationships and actions affect how social goods are and should be distributed (Gee, 1999), and information is the social good under investigation in this project. The production of knowledge and share of information is a social and historical process that precipitates a notion of public good (Fuchs, 2008). This project operates under the assumption that communication of information is a social good. The tension and debate within a politics of information focuses on the emancipatory and good process and this controlling element of the good. A mechanism of control in the information economy is the logic of the protocols that define, structure, and implement code (Galloway & Thacker, 2007).

---

[24] In response to that problem, the University of North Texas embarked on a major web archiving initiative of government websites. Yet, they are only able to harvest part of the online government environment and are an as a public institution in the state of Texas still subject to their state oversight.

**Politics of Information Organization**

As information is organized it becomes further integrated into political systems that determine how and what information is available according to a specific model of information organization. The organization of information becomes a necessity as quantities of information increase. As a result of the increased amount and complexity of printed information available over the past 200 years, which is too difficult for manual searching, additional mechanisms have been implemented to help with searching (Bates, 2002). Prior to onset of the printing industry, the transition to formats that enabled searching to find specific passages began with the book and vertical files (Gitelman, 2014; Vismann, 2008).

Information sets are organized to facilitate two kinds of access methods to the content: browsing and search. Browsing is enabled in print media through a structured product with sections or chapters and use of headings and layout in order for the reader to quickly peruse, browse, and identify information to consume. In this way, the "typed copy worked as a sort of natural language code" (Gitelman, 2014, p.70). Layout and font choice are integral to allowing for browsing within a document or corpus. This system of browsing in print media for information seeking and retrieval is usually limited, relying on the existing terms within the content that may be augmented by design and layout to "catch the eye" and are sometimes aided by a Table of Contents for quick location finding. With the printing revolution, the constructed object became more standardized (Febvre & Martin, 1976). The mechanisms that enable more direct searching include both revisions of the format and the technique of indexing.

When collections of objects in these formats became too unwieldy, indexing was employed to facilitate access. Indexing is the process of creating a short cut based on identified terms (e.g., subjects, dates, and people) to enhance access to certain portions of text or content organized in a taxonomy of set standards, terms and rules to facilitate finding information. Index catalogs provide an index across multiple works or collections. The earliest index catalogs of collections focused on personal, specific professional, or specific institutional contexts. The onset of more generalized indexes presented a transformation to a highly controlled political context (Krajewski, 2011). To facilitate search in print media, a supplemental information organizational guide had to be created. The resulting search guide takes the form of an index or catalog. Early collections of documents may have been arranged chronologically; however, around 1500, the process of arranging documents according to subject was introduced (Vismann, 2008). Indices and catalogs provide an externally applied set of vocabularies or taxonomies that define the subjects and ways of access. The function of an index to select and retrieve information becomes a censoring device by the nature of the selection of what and what not to include (Krapp, 2006).

The organization and selection of subjects and ordering is inherently political and steeped in traditions of particular social groups and social forces (Ranganathan, 1973). Subject categories also informed classification and ordering of materials in libraries and archives, such as the Dewey Decimal System and Library of Congress Classification System that made it easy to locate items on a shelf using a scheme of ordinal numbering. Classification is "the process of translation of the name of a specific subject from a natural language to a classificatory language" (Ranganathan, 1973, p. 31). The act of

classifying information and using taxonomies was the work of a skilled indexer or librarian who was trained in the formal standardization provided by a classification scheme. The effect of printing, democracy, and the availability of public libraries to the masses resulted in new methods of ordering that were broadly accessible to the populace and necessarily transparent (Krajewski, 2011; Ranganathan, 1973; Vismann, 2008). Subject remains the dominant form of classification to print materials through the development of electronic databases of information with a necessary transparency to subject and arrangement taxonomies and schemas that facilitate finding information.

The limits of the effects of a particular social group and a critical example of eschewing the politics of the organizational system can be seen in the work of Howard University librarian Dorothy Porter in the 1930's. Porter defied the Dewey Decimal System, the organization system widely in use at the time and still today, which specifies ordering and classification of books by subjects by creating systems that allowed for the integration of materials by and about Blacks. "Against an information landscape that exiled black readers and texts alike, Porter's catalog was a site where radical taxonomy met readerly desire" (Helton, 2019, p.101). Porter ordering of information was an act of activism that redefined how and what information was made accessible.

**Politics of ICTs (Information and Communications Technologies)**

Instead of a set of organizational and institutional priorities collecting and collating information, the Internet provides a vast networked and de-centralized distribution mechanism for information. In this network, however, information must still be collected and collated to make it findable. This is where an index, particularly for

search engines, provides the mechanism. This decentralized nature of the online environment furthered speculation that the Internet could act as a tool for public good and democratization free of the structures that defined and restricted access to information in earlier forms of media. Many scholars, however, point to the neoliberal ideologies and current systems of capitalist control only enhanced by online communications (Castells, 2000; Dean, 2009) and dominated by a few mega ICT companies. Technological development and online communication have helped the formation of an information economy where the key commodity of exchange is control. Wendy Chun points to a dualism of the Internet as a tool of freedom and a "dark machine of control" (Chun, 2006).  In this environment, which Terranova describes as "informational milieu", political intervention is only possible with an engagement of distribution and access to information such as "opening up channels, selective targeting, making transversal connections" (Terranova, 2004). The politics of information frame the base for investigating the institutions that control commodities for the public good and regulate access to information.

### *Politics of Search Engines and Social Media Platforms*

Search engines and social media platforms are a central locus of the institutions of control in the online environment. Google and Facebook have fought against their characterization as media companies in the political discourse in the U.S., minimizing the perception of them as institutions in need of policy intervention and regulation (Napoli, 2014).  The success of these services is that they "have become indispensable to the political economy of citation indexes, online public relations and marketing, knowledge

production, and NGO advocacy activities" (Franklin, 2013). Political discourse without exposure and discovery through these services can make access to information and communication of ideas a tricky business.

In investigations of the political role of search engines and social media platforms and looking at the role of the algorithms used in these processes, the algorithms cannot be examined without understanding of the social and cultural context of their creation. "Algorithms are inevitably modelled on visions of the social world, and with outcomes in mind, outcomes influenced by commercial or other interests and agendas" (Beer, 2017, p.4). It is important to understand the way that decisions are made to render the content via search engines and social media platforms are influenced by the algorithms and how the content creators attempt to subvert, play the game or manipulate the algorithms for access to their content. "The power of algorithms here is in their ability to make choices to classify, to sort, to order, and to rank. That is, to decide what matters and to decide what should be most visible" (Beer, 2017, p.6). In this project, the role of SEO and SMO strategies are examined in conversation with the algorithms that provide or prevent opportunities for access to information. Though code is examined and technical strategies, those strategies should be thought of in context of the social conditions and structures in which they were created and enacted.

## Gatekeeping

This dissertation uses Gatekeeping theory to surface the role that various structures, norms, and subjectivity play in preventing and allowing access to information in the online environment and how those mechanisms may be subverted by an individual

or organization through the strategies of search engine and social media optimization. In this analysis, the historic role of gatekeeping through various media are important for understanding how gatekeeping may function in both similar and different ways from the Internet and online documentation.

**Gatekeeping and Mass Media**

The communication theory approach of gatekeeping began with David Manning White's study of the choices for a newspaper story from its inception of what was newsworthy to the decision to print and distribute through a chain of gatekeepers (White, 1950). It is through many decision points in the chain that determine what information is communicated and considered newsworthy. In this landmark study for gatekeeping, White emphasized that it is when looking at the stories that are rejected by a newspaper editor for printing that the subjectiveness of the decision-making process is revealed and the emphasis on the experiences of that gatekeeper are dominant in the gatekeeping process (White, 1950, p. 386). The newspaper editor is the "terminal gatekeeper" for White, as the person who ultimately decides what information is available to the broader public. As part of White's argument, he discusses the premise from psychology that "people tend to perceive as true only those happenings which fit into their own beliefs concerning what is likely to happen" (White, 1950, p. 390). This concept is foundational to a role of gatekeeping not only as a means for determining information available but noting that that process is inherently biased and coded within the norms of expectations from the gatekeepers that are making those decisions.

Subsequent studies of mass media and gatekeeping have refined gatekeeping models of "agenda-setting," noting the impact of gatekeeping in mass media to set the political agenda. In McCombs and Shaw's study of the 1968 presidential campaign, new content, and voters, they identified correlations between the issues of importance to voters and those emphasized in the news media (McCombs & Shaw, 1972).[25] In this study, they explored how agenda-setting could have an important influence on the social and political spheres. Following these defining studies for gatekeeping, communication scholars have further investigated the role of gatekeepers across mass media and expanded to identify the influence of the organizational impact on the gatekeepers. The organization functions within "input-output relationships" with its environment (Dimmick, 1974, p.2).

As a result of these studies on gatekeeping and the role of knowledge production, one part of remediation suggested is to assert a separation from the producers of content and disseminators in order to reduce the impact of organizations on mass-media gatekeepers (Hirsch, 1972). As we move into the online environment, that separation may be more striking than traditional mass media; however, the balance of information available may sway less favorably for society after all.

---

[25] An interesting comment form McCombs and Shaw that bears note for the study in the online environment is that the values of readers and news producers are strikingly different (McCombs & Shaw, 1972, p. 185).

**Gatekeeping Online**

One of the earliest problems identified with the Internet was that there was so much information that finding a specific piece of information became problematic (Berners-Lee, 1989). Early attempts at categorizing the web by Yahoo and even the Librarian's Index to the Internet attempted to reproduce historic archival indexing within the online space. Some strategies, such as the use of `<meta>` tags for categorization and information management were adopted by both search engines and social media applications. The digital world, however, also made it easier for full text indexing, and these services are not limited by the space of time needed for such activities as historically were required by print.

*Gatekeeping through Search Engines*

In recent years, the influence of search engines as a gatekeeper has been a frequent object of studies. Search engines have a basic function where they crawl the World Wide Web and index webpages based on the content of the HTML where they then use proprietary algorithms to rank the search results by relevance. There are many strategies and tools to make this easier for search engines. As discussed in the introduction, elimination from a search engine can, in effect, make that information inaccessible, including submitting URLs to search engines for indexing. Search not only limits to a subset of information, but it also functions like institutions and sets the criteria for information seeking by individuals (Napoli, 2014). The criteria that are used for relevancy of rankings is then pivotal in the gatekeeping function for search engines and

has resulted in some scholars calling for a public demand to release the algorithms for

transparency (Introna & Nissenbaum, 2007).

### *Gatekeeping through Curation of Links in Online News*

Another important angle of gatekeeping in the online environment is the role of

content creation and selection for integration within online pages themselves.

> Online journalism can be functionally differentiated from other kinds of
> journalism…The online journalist has to make decisions as to which media
> format or formats best convey a certain story (multimediality), consider options
> for public to respond, interact or even customize certain stories (interactivity), and
> think about ways to connect the story to other stories, archives, resources and so
> forth through hyperlinks (hypertextuality) (Deuze, 2003, p. 206).

Even in the online environment, these activities are not dissimilar from the role of the

editor in what get printed in the newspaper page except the decision is now how to relate

and or not relate other content online with your content. In studies of newspapers online,

the practical concerns of longevity of links, the authority / trust in content, and

competition with other outlets may limit the use of linking to online content (Cui & Liu,

2017).  The attitudes of journalists toward linking and what to include or not within the

webpage content are aligned with "classic journalistic principles" (De Maeyer, 2012).

When news sites have used links to sources, those links are "directed toward sources that

were within mainstream media (often internal), political neutral, undated, and reference-

based" (Coddington, 2012, p. 2020). As the hyperlink is one of the defining

characteristics in digital new media, the extent to which journalists consider linking

strategies or incorporate the practice or not, points to how disruptive or not online

communications practices have been for news media. The motivation to include links

within traditional journalism in online environment is focused on providing context for curious readers, and there is general agreement that it is a good practice to inform readers; however, it is not typically employed (Coddington, 2012; De Maeyer, 2012). This also outlines a tension in how the decisions of gatekeeping within a webpage may have a direct impact with the other gatekeepers on the web, through search engines and social media platforms, and control access to content.

Consistent with these attitudes, in looking at the evidence of linking within news articles, studies of the activities of journalists have shown little time in the journalists work spent on considering or curating links (De Maeyer, 2012; De Maeyer & Holton, 2016). "The confrontation of the bright theoretical promises usually related to hyperlinks and the more nuanced picture showed by empirical research about the actual linking behaviour of news sites underlines a stimulating gap" (De Maeyer, 2012). One of the transitions to within online journalism also requires a recognition and acceptance that "[journalism] it does not function as sole provider of content" (Deuze, 2003, p. 218). Where and to whom the power of gatekeeping information online is more nuanced an complicated in the layering of gates, in order to make content accessible.

One of the goals of this project is to identify technical areas of consideration that should be included into the communication discourse and creation process as essential to providing access to information and where considerations of gatekeeping can be further interrogated. These questions of connecting content online are not unique to news media. News media provides a unique look into the practices of connecting to other sites In this project, the role and decision of curating links, as an SEO and SMO strategy, is a way of influencing gatekeeping online; however, the act of curating the links is a gatekeeping

function itself, as well. As content is created for the online environment, these various layers of gatekeeping should be kept in mind and be part of the decision-making of the content creation process for online communications.

### *Gatekeeping through Social Media*

Social media platforms are inherently different than search engines in that, although they also may link to external content, they are also fully independent applications where content is largely user-created and that information is fed back into the application. The application could theoretically exist without external content. As search engines and social media platforms began to act as an intermediary between the media and other content creators and readership, the role of gatekeeping switched to one of regulation through algorithms and code.

Research studies of the American public's behavior in seeking information online and sources of online access increasingly point to an increased percentage of people finding their information online.[26] The perception of the role of the gatekeeper is important in identifying the importance of the gatekeeping and to help expose the invisible actors (human and machine) at this level of gatekeeping. A 2012 Pew Research Center Study found that 2/3 of adults said "search engines are a fair and unbiased source of information" (Purcell et al., 2012). The public perception of search engines is that they are neutral (Pan et al., 2007). Although social media platforms do not have the same perception of neutrality, most users believe that they are receiving content in their feed

---

[26] See several studies from the Pew Research Center: (*American Feel Better Informed Thanks to the Internet*, 2014, *Internet Use Over Time*, 2014, *State of the News Media 2015*, 2015)

based on friend's recommendations and neutral algorithms functioning on factual and

non-ideological data (Light & McGrath, 2010). The gatekeeping role is essential to the

argument of this dissertation where competing ideologies are at work to expose and

provide access to information, so these institutional structures and behaviors will be

examined.

CHAPTER III

METHODOLOGY

There are many challenges with writing a contemporary history of online communications, and the methodological approach must account for the influence of the structure of the presentation. Gitelman asks, "How is doing a history of the World Wide Web, for instance, already structured by the web itself?" (Gitelman, 2006). This chapter is organized into four sections and reviews strategies and processes for studying webpages as historic online documents. The first section will review the research questions that drive this project. The second section will present both the methodological framework and an overview of strategies of a media archaeological analysis in the context of analyzing SEO and SMO. The third section will discuss the method of historical document analysis applied in the media archaeology framework. The final section will review the selection, collection and analysis procedures employed to answer the research questions.

**Research Questions**

Previous critical history research on the gatekeeping function of search engines and social media platforms focuses on the notion of the hidden and proprietary algorithms that are used to determine content exposure and have critiqued the results of these algorithms as the algorithms themselves remain hidden. These studies often call for an emancipatory transparency of the algorithms; however, the impetus for the corporations to employ transparency is unknown, and the likelihood of government regulation is even further unknown, as deregulation of corporate America has been the

ongoing trend for some time.  What incentive is there for companies to expose these

algorithms?  Computer science research also focuses on adaptations of search engine and

social media platform algorithms in order to prevent the use of "adversarial strategies"

that seek to jump the gates of the search engine and social media platforms. These

computer science technical research projects are aimed at perfecting the gatekeeping

function. Advertising and marketing materials address SEO and SMO for a practical

application, but rarely look critically at the usage over time and interplay with search

engines. This project is focused on the interaction of SEO and SMO strategies and looks

at opportunities to influence with SEO and SMO due to the structure of the web and

online content. By investigating the role of SEO and SMO, this project seeks to identify a

history of this interaction between platform gatekeeping algorithms and the SEO and

SMO strategies, in order to provide attainable outcomes. This study is guided by the

following research questions:

> RQ1: What is the historical development of SEO and SMO strategies?
>
>> RQ1a:  What are the topoi in these practices?
>>
>> RQ1b:  What is the interplay with changes in proprietary algorithms over
>>             time?
>
> RQ2: How has the development of SEO and SMO strategies been actualized in
>             HTML practices for major persuasive information industries?
>
>> RQ2a:    How have the strategies been implemented in newspapers' online
>>                 presence?
>>
>> RQ2b:    How have the strategies been implemented in political candidate
>>                 websites?
>
> RQ3: How have SEO and SMO strategies shaped communication online?

The first research question situates the SEO and SMO strategies in context over time by looking at technical guidebooks and manuals with instructional content for the SEO and SMO strategies. To address this question, a historical descriptive study of the strategies is used. It also involves looking at the relationships and interplay between the strategies for SEO and SMO and the changes in search engine algorithms over time. In addition, this historical overview provides a reflection on how these strategies do or do not differ from previous strategies in other media formats for selecting and making information accessible. Are there topoi present that can be identified as indifferent to the media of the web and HTML structures?

The second research question takes the distilled strategies from the first question and explores how these strategies were implemented in webpages in major persuasive industries. In the media archaeological analysis used in this project, it is important to examine the structure and usage of coding. In order to identify the norms enforced in the technology, HTML pages from industries where the everyday impact of availability of information is essential to the existence and mission of those domains are explored. The major persuasive industries of the newspaper articles and political candidate webpages were selected for examination. In communication studies, newspapers and candidate platforms have historically been assessed as newspapers as the gatekeepers to candidate platforms in the print medium and selection of news stories (Cui & Liu, 2017; De Maeyer & Holton, 2016; Fink & Schudson, 2014; White, 1950). Both news media and political platform candidate pages have been widely examined in online communication practices, and the role of search engines in gatekeeping content (Ali et al., 2019; De Maeyer & Holton, 2016; Diakopoulos, 2015; Nechushtai & Lewis, 2019). This complementary

technical and media archaeological analysis provides further tools for decision-making by producers of news and political content when and what strategies to incorporate on SEO and SMO and the pluses and minuses of the strategies coupled with other practices and or author intent.

In addition, by focusing on newspaper article and political candidate webpages, these two sectors present an opportunity to explore communication practices, which engaged in early web activity. Because of the early engagement in web activity, archived versions of this content over time have been captured and are available for analysis of practices over time. By examining archived webpages, the goal of this question is to validate the strategies found in the first question and understand the implementation and changes in SEO and SMO strategies over time.

The final research question explores how SEO and SMO strategies within the HTML structure and page content affects communication strategies in an online environment through the examination of newspaper article and political candidate webpages. This question does not delve into writing for the web strategies[27] as a whole, focusing instead on the SEO and SMO strategies and how their implementation may or may not change communication practices in newspaper article and political candidate webpages. This examination will reveal a snapshot of how actualized strategies of SEO and SMO may have influenced communication practices.

---

[27] Writing for the web is a set of strategies focused on the usability, tone, use of white space, and recommendation such as using "you" and "we" in text (Assistant Secretary for Public Affairs, 2016).

As in many historical studies, the research questions began as a "guided entry" and are refined as the study progresses and discoveries occur within the investigative process (Smith, 1981, p. 307). A historical method qualitative approach was used to address these questions with document analysis. Rather than merely a descriptive exploration, these questions attempt to explore the important relationships that are involved in communicating online and the complexities of communicating through the gatekeeping technologies of search engines and social media platforms.

To answer these research questions, two media types are used. The first set of media will be instructional guides and how-to manuals books centered on applying SMO and SEO strategies in HTML. A historical document analysis will be used to examine these sources and the recommendations they assert. These books are limited to professionally published print books and do not include self-published books. The second set of media are archived webpages, accessed through the Internet Archive's Wayback Machine and the Library of Congress web archives collections.

## Methodological Approach: Applying a Media Archaeological Method

This dissertation presents a recent history of contemporary communication and technology practices with HTML webpages and their exposure through search engines and social media platforms. In order to perform a historical analysis for a contemporary practice, this dissertation employs a historical media archaeology approach. Media archaeology is particularly suited for this investigation by employing a techno-cultural approach and that is both "self-reflexive" and examines media as "archival objects of research" (Ernst, 2005, p. 587) and aspires to expose the invisible and alternate histories

to make sense of digital media and the political and cultural institutional contexts of the present (Parikka, 2012).  The media archaeology approach exposes the invisible rules and structures that serve the discourses available through the content results and carries an important cultural effect. A media archaeology investigation, as expressed by Foucault, differs from a traditional historical analysis or a strictly textual analysis by:

> 1) examining rules of discourse rather than the content of the discourse,
> 2) defining the specificity of discourses and the rules they enact,
> 3) not championing a creative subject, and
> 4) not attempting to recreate intent but rather is "systematic description of a discourse-object" (Foucault, 1972, p. 140).

For this project, the place of the media within the discourse and the processes that both frame the communication and the creation of HTML documents becomes the object of study rather than the discourses, content analysis, or effects of the communication. This dissertation adheres to an approach that emphasized the functionality of the technical architecture, operations, and processes that exist within cultural norms and powers relations for a specific medium.

A media archaeology analysis provides insight and contextual bearing on the question of whether a cultural practice was an effect of new media or the new media was created because of the "epistemological setting of the age demanded them" (Ernst, 2005, p. 587). In the revealing of the contexts and architectural frameworks and operations, one of the tasks of the media archaeology project is to interrogate newness and look for what is already known and the recurring cultural formulae, "topoi," that permeate the media apparatus (Huhtamo & Parikka, 2011).  This dissertation also seeks to identify those topoi in the development of optimization practices and access to information.

As a critical method, media archaeology adopts many practices from discourse analysis but focuses on structures and materiality over content (Foucault, 1972). Operationally, this means that the structural rules are examined in much the same way as discourses and how the rules of discourses are embedded in the media (Parikka, 2012). Many of the traditional components of a qualitative discourse analysis are used in media archaeology and historical method best practices should be followed for a rigorous study. A media archaeology analysis should be empirical, systematic, and rigorous. Specific, precise, and thoughtful decisions need to be made in selection of content for analysis and examinations. An essential component of a media archaeological analysis is an infrastructure approach to the investigation rather than an interpretation (Ernst, 2013; Foucault, 1972; Kittler, 1995).

In a traditional historical document analysis, the identification, authentication, and verification of documents is essential to an empirical study (Scott, 1990). These concepts are complicated in the digital environment. Where once things like handwriting and type of paper could be used to authenticate documents, the digital does not have such affordance and documents are continually made new and take the form of different representations (Brügger, 2012). Internet studies that utilize the internet in the archival historical search must also be conscious that the act of doing that historical work is structured by itself (Gitelman, 2006). Throughout the data gathering and analysis for this dissertation, a recognition of utilizing search while studying search will be acknowledged.

The media archeological analysis uses a document analysis but at the level of the discourse-object and its materiality. The digital media discourse-object has a layered

materiality consisting of multiple components that make up the transitory and variable

digital object (Parikka 2012).  In a traditional media archaeological analysis, the

materiality is examined in terms of the physicality of the structural components that

compose a medium. An important component of the structural materiality is how the

pieces connect and function together. In discussion of the role of the first microprocessor

from Intel, Kittler notes, "…computing, whether one by men or by machines, can be

formalized as a countable set of instructions operating on an infinitely long paper band

and the discrete sign thereon" (Kittler, 1995, p.148). On top of this formalized set of

instructions for the hardware, software creates another layer of hidden instructions upon

the media, and yet it is reliant on the hardware and to materiality of the components to

which "are built into silicon and thus form part of the hardware" (Kittler, 1995, p. 150).

The materiality of the computing system is part of the hidden mechanisms to be exposed

through a media archaeological analysis.

In this project, instead of focusing on the materiality of the components of the

proliferation of hardware and software devices that exist to render HTML content, the

layers of hidden mechanisms to be exposed focus on the functionality and the processes

that occur between the HTML, SEO and SMO strategies, and search engines or social

media platforms. Beyond a counting or basic content document analysis, the media

archaeological analysis explores the references, functionality, and intertextuality of the

documents in context. It does not search for meaning or intent, and in examining through

the relationship with search engines and social media platform, the text in action becomes

the focus. The implication in this type of analysis is that the technical is inherently social

and cultural and built on structures that create and reinforce power relations, and in the

case of this project, access to information. The media archaeological analysis, in this project asks, what is it that actors are *doing* with the words [and code] (Prior, 2003)? The important conception of this work is not the intent of the coding but the effect that the code has on the access to information through the gatekeeping applications of search engine and social media platforms. The choices in the coding rely on the sociotechnical structures within HTML and online communications. In addition, the decisions of the choices in application and use of code create pathways and barriers that are part of a larger sociocultural consumption of information and reception of communication.

This project also attempts a critical history. There is an opportunity to identify guidance for future use and possibilities of change (Winthrop-Young & Wutz, 1999). By using a media archaeology approach, the intent is to dissect the systematic and structural attributes that form as gatekeepers to information and to identify opportunities for influence the accessibility of information, as well as to identify places for activism within the structural forces that determine the standards and allowances of communication methods online.

**Historical Documents**

With a media archaeology approach, a historical research is the primary investigation tool for media artifacts (primary sources).  In this project, historical documents will be the basis of the investigation. In using historical documents in research, one of the most critical aims is to provide evidence that is not selected merely to prove an existing conclusion. "[B]y examining document content in terms of a strictly defined set of procedures, researches can produce robust and reliable conclusions" (Prior,

2003, p.149). This study will examine the HTML code of archived webpages for evidence of SEO and SMO strategies embedded in the code. To provide context and inform the identification of topoi, instruction manuals for SEO and SMO strategies will inform the document analysis of the archived webpages.

### *Use of Instruction Manuals*

Instruction manuals are how-to guides examined to provide context and comparison between the recommended strategies in the manuals and the actualization of strategies in the webpages later examined. Manuals present an illustration of standards and practices, which make it possible to identify complete relevant strategies (Prior, 2003, p. 151). In the study of web communications and technology, manuals and how-to-books are not expected to be representative of how a typical user may implement tactics and strategies but rather well-thought-out and intentional articulation of strategies based on author expertise (Owens, 2015, p. 33). Because manuals and how-to guides are also written in order to develop a skill, the tactics and strategies outlined in the texts can be extracted but also need to be matched with an audience and goals.

### *Use of Archived Web Documents*

Studying historic web documents provides particular challenges in that the pages may change over time and different versions may be captured. The device and hardware used to render the content may not be available or the particular organization of content could be constructed based on personalized information (Park et al., 2011, p. 286). Studying the web document is particularly a problem in archived web pages where a reconstructed and incomplete copy is viewable, because all code used to render the page

and media content may not have been captured in an archival harvesting process (Brügger, 2012). The use of web documents in this study is advantageous and should not encounter that issue, as the code needed for SEO and SMO strategies must be embedded in the rendered HTML in order for web crawlers to harvest the pages properly. Although some code may be written as part of an include or script, it is rendered viewable in the HTML. The use of "View Source" web browser function and "save as HTML" or "save as Complete Webpage" allow the HTML for SEO and SMO strategies encoded to become visible. In this project, archived webpages accessed through the Internet Archive's Wayback Machine and Library of Congress web archives collections will use the built in Chrome Browser function, "Save Page As…Webpage, Complete" to capture pages for analysis. This results in capturing several files associated with the webpage in an accompanying folder.
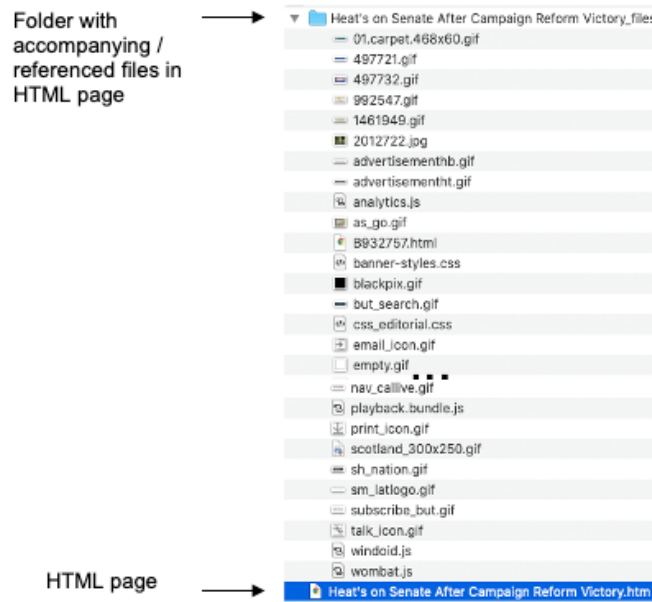


**Figure 3.1.** Example "Save Page As…Webpage, Complete" artifacts.

Adobe Dreamweaver was used to examine the HTML code. Dreamweaver was used for the following three reasons, 1) it automatically detects and inserts accompanying files in the view, 2) it allows for a split view of HTML code and a sample rendered page through a browser at the same time, and 3) it has quick code find, replace, and strip tools that allowed for efficient removal of additional HTML inserted by the web harvesting tools. Both the Internet Archive's Wayback and the Library of Congress' web archives collections use a version of the Open Wayback Tool[28] to harvest webpages, which inserts clear notifications of non-native HTML added the pages that are added when the webpages are viewed through the online web archives:

```
<!--
   =====================================
   BEGIN Wayback INSERTED TIMELINE BANNER

   The following HTML has been inserted
   by the Wayback application to enhance
   the viewing experience, and was not
   part of the original archived content.
   =====================================
-->
......
   <!--
   =====================================
   END Wayback INSERTED TIMELINE BANNER
   =====================================
-->
```

**Figure 3.2.** Example comments surrounding HTML code inserted in Wayback applications.

---

[28] https://github.com/iipc/openwayback

Additionally, the Wayback application adds path directory code for image and internal

webpages, in order to point to the archived versions and not seek the live web for these

artifacts.

```
background-
image:url(https://webarchive.loc.gov/all/20161011232019i
m_/https://drjoeheck.com/wp-content/uploads/2015/08/az-
subtle.png) ;

<a title="Meet Joe"
href="https://webarchive.loc.gov/all/20161011232019/http
s://drjoeheck.com/meet-joe/">Meet Joe</a>
```

**Figure 3.3.** Example directional code inserted by Wayback application to direct to
archived versions of referenced files.


Because this process retains the original link after the inserted archival web location link,

it does not affect the interpretation. It is important to look at the files archived at the time

of the web harvest instead of current live links, in order to imitate the original

presentation. When links are broken and/or images are missing, often the web harvester

application was not able to harvest those files and add to the directory of accompanying

files. The code used to point to those missing files, however, are not altered and the

original code is able to be examined even if the design is not.

Interpretation as a part of historical analysis must involve five types of control: 1)

evaluation of sources; 2) context; 3) historiographical changes over time; 4) generalizing

with concrete and specific evidence; 5) self-awareness to minimize bias (Startt & Sloan,

1989, p. 146-47). The following Data Collection and Analysis section will review these

controls for each set of sources examined in this project.

## Data Collection and Analysis

Before addressing the specific data collection and analyses used in this project, I will address the control of self-awareness and minimizing bias control for validity concerns. As a professional librarian, my job is about connecting users to information. Search and discovery is fundamental to this process, and I have taught credit level university courses on web design. I became a librarian when the Internet was still fairly new as a widespread tool, in the early aughts, and Google was only a couple of years old. In my graduate program, we spent significant time reviewing the Librarian Index to the Internet, which was a manually curated site of links to "reputable" sites on the Internet. Yahoo's categorized home page was in a rivalry with Google's single search box at that time. We also were trained in DIALOG search procedures, which was an aggregated pay-by-search query tool for scholarly articles that was developed in the mid 1960's. I taught myself how to be a web application programmer with several O'Reilly manuals and IT resources provided by the University of Washington. My experience in web application development and library science aligns me with a typical or advanced user for most the manuals examined. I am also able to easily read and understand some of the more technical materials and quickly parse HTML and/or use tools for identification of tags and structure. I can fully read and understand the source code of the HTML archived pages without needing to render through a browser or the rendering preview tools provided in Dreamweaver.

**Instruction Manuals and How-to Guides**

The selection of manuals and how-to guides was "purposive and pragmatic" (Prior, 2003). In order to identify relevant texts, I searched for "search engine optimization" and "social media optimization" in Worldcat.org; the international catalog of library holdings. These terms are not official subject terms for library uses, as is often the case with newer concepts, so I used the subject listings in the records of the first couple of identified texts to further identify texts and continued this pattern. Subject headings for the books in the analysis included: Internet Marketing, Social Media, Web Search Engines, Internet Searching, Search Engines, Electronic Information Resource Searching and Electronic Commerce.

In order to identify texts with particular influence, I selected texts that had over 200 holdings in libraries worldwide. This number is a bit tricky, in that, in an attempt to combine language and similar editions (e-formats and print editions), WorldCat has combined some of these together. For the purposes of this project and looking for relative popularity and spread of titles, that limitation does not affect the general spread of the text. Another limitation in holdings for manuals and guidebooks is that outdated texts are usually withdrawn from library collections. In order to address this problem, I did not use the 200 holding threshold for books published in 2011 or earlier. In these instances, the availability of the texts affected my ability to include it in this project, and primarily texts were selected based on their availability for analysis. As an additional mark of popularity of the texts, I also looked each up on Amazon.com and noted the number of ratings for each text. These also may be problematic numbers, as I did not filter out paid or robot

ratings for each text. This measure was employed as a check and understanding of popularity and not a selection factor. The most popular texts on Amazon.com for search engine optimization and social media optimization are self-published texts promoted specifically by Amazon. I decided not to use these texts and focus on those with established publishers as a criterion for quality. It would be interesting, in a future study, to see what differences, if any, exist between the manuals published by known technology publishers and the self-published titles that are highly used from Amazon.com. I identified 15 manuals spanning publication from 2005 to 2018 for the analysis. A little less than half of the titles are published by John, Wiley & Sons. This is a high percentage of the titles due to both the focus of John, Wiley, & Sons on producing technical manuals and their acquisition of many technical publishers.[29] Table 3.1 presents the manuals selected for the analysis and includes the number of Amazon Ratings and WorldCat holdings, as well as the target audience, which was identified during the examination of the texts.

Each manual was around 200 pages. For this historical document analysis, I used an iterative approach consisting of skimming and then close readings of the texts. The initial texts used to create the sample data collection sheet were: *The ABC of SEO: Search Engine Optimization Strategies* (2005), *Search Engine Optimization Bible* (2009), and *Introducing SEO: Your Quick-Start Guide to Effective SEO Practices* (2016). During the data collection and analysis, I refined the data I gathered from the texts, adding more depth to the questions about the presentation and strategies related to Black Hat

---

[29] See: https://www.wiley-vch.de/en/about-wiley/the-publishing-house

**Table 3.1.** Chronological Listing of How-to Guides and Instruction Manuals.

| Title | Publisher | Year | # of Amazon Ratings | WorldCat Holdings | Target Audience |
|---|---|---|---|---|---|
| *Digital Branding A Complete Step-by-step Guide to Strategy, Tactics, Tools and Measurement* | Kogan Page | 2018 | 16 | 282 | Marketing professionals; public relations professionals |
| *Introduction to Search Engine Optimization: A Guide for Absolute Beginners* | A Press | 2017 | 5 | 324 | Interns; college students; self-paced learners; journalists |
| *Introducing SEO: Your Quick-Start Guide to Effective SEO Practices* | A Press | 2016 | 1 | 302 | Web designers; website managers |
| *Win the Game of Googleopoly: Unlocking the Secret Strategy of Search Engines* | John Wiley & Sons / Skillsoft | 2015 | 62 | 606 | Small businesses |
| *Search Engine Marketing, Inc: Driving Search Traffic to Your Company's Web Site* | IBM Press | 2015 | 38 | 218 | Marketing professionals |
| *Social Media Optimization for Dummies* | John Wiley & Sons | 2015 | 15 | 621 | Marketing professionals |
| *Letting Go of the Words: Writing Web Content That Works* | Morgan Kaufmann imprint of Elsevier | 2014 | 109 | 1718 | Marketing professionals; graduate students; technical writers |
| *Search Engine Optimization: Your Visual Blueprint for Effective Internet Marketing* | John Wiley & Sons | 2013 | 45 | 1332 | Marketing professionals |
| *Optimize How to Attract and Engage More Customers by Integrating SEO, Social Media, and Content Marketing* | John Wiley & Sons | 2012 | 59 | 660 | Marketing professionals; public relations professionals; small to medium sized business owners; large company marketing executives |

**Table 3.1. (continued).**

| Title | Publisher | Year | # of Amazon Ratings | WorldCat Holdings | Target Audience |
|---|---|---|---|---|---|
| *Search Engine Optimization Bible* | John Wiley & Sons | 2009 | 12 | 501 | Website managers; web application programmers |
| *The Findability Formula: the Easy, Non-Technical Approach to Search Engine Marketing* | John Wiley & Sons | 2009 | 55 | 173 | Marketing professionals |
| *Mastering Web 2.0: Transform Your Business Using Key Website and Social Media Tools* | Kogan Page | 2009 | 3 | 648 | Marketing professionals; small and medium sized business owners |
| *Marketing through Search Optimization: How People Search and How to be Found on the Web* | Elsevier | 2008 | 3 | 522 | Marketing professionals |
| *The ABC of SEO: Search Engine Optimization Strategies* | Lulu Press | 2005 | 12 | 11 | Website managers; web application programmers |

techniques and merged the specific category of mobile techniques into the category of

"Other" notes. The primary research questions did not change with any discoveries

during that process. A sample data collection sheet is listed in appendix A.

**Archived Webpages**

One of the challenges of studying the variable nature of websites and webpages

has been addressed by using web archives, which take a snapshot of the webpage at a

particular point in time. These archives, which are associated with a harvest date and

time, may be incomplete and render differently in web browsers that exist at the time of

analysis compared to web browsers used at the time they were created. However, the

HTML code leaves traces of what is missing and scars for where content should be, such as with a missing image or broken Adobe Flash. The technologies that are used in web archiving are not too dissimilar from search engines. A robot / spider (code) goes out and crawls the webpages tracing links and grabbing content as it goes. Where the search engine takes that data into a cache that is indexed and returns results, the web archive packages the files in a mirror of its original formation and copies the files to be stored within the archives in a format called ARC.[30] The process of searching the archives is limited by the content that has been captured well within the application. Just like print archives, "[w]e may say that archives are the manufacturers of memory and not merely the guardians of it" (Brown & Davis-brown, 1998). Research in web archives is limited to what has been successfully harvested by the web archiving application.

The Wayback application has four basic components that make up an archival web service: 1) **Query UI,** which allows users to search against the Resource Index in the collections, 2) **Resource Store,** which stores copies of the web pages and associated files, 3) **Resource Index,** which allows full text search of the archived pages and other search queries, and 4) **Replay UI,** which presents the content, usually with archival citation information inserted, and inserts the Wayback code to maintain links to files harvested at the same time (Tofel, 2007). For this project, all four components were used to find, examine, and save documents. The Query UI was used in the Internet Archive's Wayback machine to query a primary URL string stored in the Resource Index, while the

---

[30] ARC file format specification from the Internet Archive:
http://archive.org/web/researcher/ArcFileFormat.php

Query UI at the Library of Congress web archives cataloged websites. Therefore, I was able to query the Resource Index for characteristics based on geography and level of U.S. election. The Replay UI was used to save the webpage and accompanying files that are stored in the Restore Store. In addition, I took screenshots of particular code and Replay views in the browser of notable aspects found during the analysis.

Analysis within web archives is based on retrieval of a particular URL and archives are grouped around a URL. Tools used to search these web archives are lacking and "require a substantial human effort" (Costa et al., 2017). The Replay UI was used to select webpages with content, primarily through the calendar browse interface.
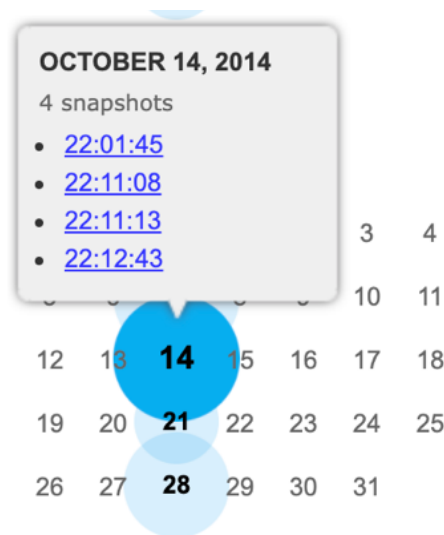


**Figure 3.4.** Calendar browse interface of Open Wayback application displaying number of snapshots of the webpage created by harvests.

Existence of a snapshot does not guarantee retrievable content. Each snapshot had to be selected and often snapshots were eliminated due to URL resolution errors the harvester encountered, such as "302: redirect" and "404: content not found." Another group of websites eliminated from consideration contained content that appeared to be harvested

into a snapshot, but resulted in a blockage in retrieving the webpage content due to a pop-up or a log-in screen.  The process by which webpages were identified that had content and could be analyzed was a long process of trial and error. This process is extremely time consuming, as the load time on each webpage from the archival services is significant, taking up to 5 minutes for a partial page load from the Resource Store.

### *Newspaper Articles Archived Webpages*

In order to identify webpages for this project related to newspapers, I made several assumptions: 1) online newspaper articles for major dailies are created through content management systems, 2) one or more individuals may have been involved in the creation of the file content for an online article, 3) automated scripts from the content management system may or may not have been used to populate aspects of the pages, 4) due to the complex enterprise content management systems usage, articles will have the same basic structure from the same paper around the same time period, and 5) those content management systems are unwieldy and unlikely to be changed frequently.

For this project, I selected the *Los Angeles Times* for their online articles. The *Los Angeles Times* has a robust online history, is archived by the Internet Archive's Wayback Machine and has maintained a constant URL since its initial harvesting by the Internet Archive, *latimes.com*. Unlike many other national newspapers, the subscription gateways to archiving online content were only prohibitive from 2003-2004; see Figure 3.5 for overview of archived content available. It is also a newspaper that has both a national and a local audience and devotes significant resources to articles on national politics. In addition to the scope, audience, and availability of content, the Los Angeles Times was

selected because of its history of integrating technological advances and innovations and for influential articles published during the time period of available archived content, including five Pulitzer Prize winning articles in 2014 alone. (*Los Angeles Times | History, Ownership, & Facts*, 2019). In selection of the *Los Angeles Times*, it was also important to identify a news site where over 70% of the content was created by the news organization. As online newspapers have attempted to stay afloat, many have integrated mass amounts of click bait and including third-party content that may or may not be relevant to the news content. In looking to identify the SEO and SMO practices as integrated into the content and context of the online news articles, it was important to eliminate the Ad and click bait concentrated publications.

Saved **116,098 times** between December 21, 1996 and December 30, 2020.
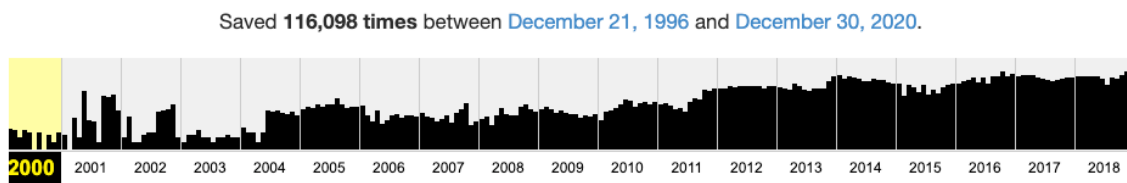


**Figure 3.5.** Chronological graph of latimes.com website harvests on the Internet Archive's Wayback Machine, which spans 2000 to 2018 of publicly available content.

Due to the *Los Angeles Times'* use of a large content management system, the structure and templates for articles changed infrequently, and I was able to select an article per year to analyze for specific structural changes. To confirm this assumption, I did checks of two to three harvests during a particular year and skimmed for structural changes. To identify the articles for analysis, I looked for snapshots of *latimes.com* over a couple of weeks preceding or following an election. Because of the variability of the presence of harvests, the dates of the analyzed articles are not consistent. From the

**Table 3.2.** Newspaper articles selected from *Los Angeles Times* on the Internet Archive.

| ID | Harvest Version (YYYY-MM-DD-HH-MM-SS) | URL | Article Title |
|---|---|---|---|
| **la00** | 2001-01-07-19-45-00 | http://www.latimes.com:80/news/politics/decision2000/upd_election001109b.htm | Florida Recount May Go Into Next Week |
| **la01** | 2001-10-08-03-32-58 | http://www.latimes.com:80/news/politics/la-082401tommy.story?coll=la-headlines-politics | Scarce Funds Imperil Bush Health Goals |
| **la02** | 2002-02-15-21-46-20 | http://www.latimes.com:80/news/nationworld/nation/la-021502finance.story | Heat's on Senate After Campaign Reform Victory |
| **-** | 2003 | N/A | N/A |
| **-** | 2004 | N/A | N/A |
| **la05** | 2005-12-20-18-25-03 | http://www.latimes.com:80/business/la-102405econ_lat,0,7536501.story?coll=la-home-headlines | Bush Names Bernanke to Replace Greenspan as Fed Chief |
| **la06** | 2006-10-16-11-27-58 | http://www.latimes.com/news/nationworld/world/la-fg-planb16oct16,0,4775251.story?coll=la-home-headlines | Panel to Seek Change on Iraq |
| **la07** | 2007-11-06-02-22-46 | http://www.latimes.com/news/local/la-me-carona31oct31,0,786373.story?coll=la-home-local | An unsettling portrait of 'America's Sheriff' |
| **la08** | 2008-10-28-13-34-35 | http://www.latimes.com:80/news/local/la-me-mailvote27-2008oct27,0,2952582.story | Popularity of mail-in voting surges in California, elsewhere |
| **la09** | 2009-10-27-11-36-32 | http://www.latimes.com:80/news/nationworld/world/la-fg-obama-afghan27-2009oct27,0,7820767.story | Push for Afghanistan troop increase continues on deadly day |
| **la10** | 2010-10-27-02-06-04 | http://www.latimes.com/news/nationworld/nation/la-na-conservatives-endgame-20101026,0,7304435.story | Conservatives struggle to unify for voter outreach |
| **la11** | 2011-10-27-23-15-31 | http://latimesblogs.latimes.com/technology/2011/10/obama-2012-campaign-starts-a-tumblog-tumblr.html | Obama 2012 campaign heads to Tumblr |
| **la12** | 2012-10-30-12-40-03 | http://www.latimes.com/news/politics/la-pn-biden-clinton-romney-jeep-ad-20121029,0,6637512.story | Biden on Romney Jeeps-to-China claim: 'Have they no shame?' |

**Table 3.2. (continued).**

| ID | Harvest Version (YYYY-MM-DD-HH-MM-SS) | URL | Article Title |
|---|---|---|---|
| **la13** | 2013-10-29-02-37-29 | http://www.latimes.com/world/la-fg-spying-phones-20131029,0,3235295.story | White House OKd spying on allies, U.S. intelligence officials say |
| **la14** | 2014-10-21-05-43-34 | http://www.latimes.com/world/middleeast/la-fg-iran-nuclear-20141021-story.html | Report says U.S. may OK more centrifuges in Iran nuclear talks |
| **la15** | 2015-10-21-13-30-57 | http://www.latimes.com/local/california/la-me-tribal-law-enforcement-20151020-story.html | In Humboldt County, tribe pushes for bigger law enforcement role on its lands |
| **la16** | 2016-10-15-11-26-13 | http://www.latimes.com/politics/la-na-pol-clinton-fundraising-20161014-snap-story.html | Hillary Clinton keeps fishing for big money while lagging behind with smaller donors |
| **la17** | 2017-11-02-02-04-16 | http://www.latimes.com/politics/la-na-pol-immigrant-abortion-20171023-story.html | How long can the Trump administration prevent a 17-year-old immigrant from getting an abortion? Case tests limit |
| **la18** | 2018-10-22-21-36-43 | http://www.latimes.com/nation/la-na-trump-cruz-texas-20181022-story.html | No more 'Lyin' Ted' — Trump heading to Houston to support Texas senator |

archived *latimes.com* homepage on a particular date, I selected an article that was of relevance to national politics. Part of the selection criteria was that an article must be linked from the homepage, an indicator of importance. I was unable to collect any newspaper articles for 2003 and 2004, as the *latimes.com* had all of its content behind a paywall that the open-source harvester could not penetrate for those years. Table 3.2 outlines the 17 articles, their corresponding URLs, harvest version /snapshot date and time, and research assigned ID used for this analysis.

Two copies of the webpage and assets were saved to a cloud folder for the researcher. One was a straight copy from the Replay UI service, and the second was used

in the analysis and was edited to remove the additional Open Wayback code. A sample data collection sheet is in appendix A.

### *Political Candidate (U.S. Senate) Issue Archived Webpages*

In order to focus on political candidate webpages where access to the online content for candidates may have contributed to the success of an election, I limited the webpages to U.S. Senate elections, then reduced to those where the final margin of victory was under 5%.[31] The assumption in these close races is that they may have been more motivated to provide increased access to the candidate webpages. The Library of Congress web archives provides access to the "United States Election Web Archive." Five candidate webpages were used per election cycle, as exploration of further candidate pages provided little to no new insight in the structure after looking at five sites. The selection and review of five items of investigation per year also eliminated any tendency to focus on the unique or obscure. Candidate webpage content selected for the analysis had a topic, issue, or priority; i.e., not a biography or slogan only. Some candidate webpages were eliminated from analysis due to the lack of available content on topics, URL resolution errors, or pop-up blockers. For the 2012 election cycle, in order to have five candidate websites, I analyzed Bob Casey Jr's website from the Pennsylvania election, where his margin of victory was 9.1% over Tom Smith. There were fewer U.S. Senate election victories under 5% margin of victory in 2012.

---

[31] The U.S. House of Representatives provides publicly accessible records of U.S. elections including the House of Representatives and U.S. Senate: https://history.house.gov/Institution/Election-Statistics/.

Due the additional cataloging information that the Library of Congress provides with their web archives, I was able to initially limit the content to U.S. Senate elections and then find those identified on the close margin analysis based on the state. Because U.S. Senate candidates may have often repeatedly run for another office, such as the U.S. House of Representatives, the calendar browse view of the Replay UI was used to select the correct election year. Another advantage of the cataloged resources in the Library of Congress collection was direct relationships between sites and various URLs that belonged to the same candidate. From the calendar view of the appropriate election year and URL, I selected a snapshot as close to October 15th as possible. This date was selected both because of its proximity to the national election date in November and its appearance as a frequently available harvest date. From the main page of the candidate website, I selected the first article available on an issue or priority for the analysis. Because some sites prioritized issues where others used an alphabetical order, there is no significance across websites to the first article that was available. The following table includes 40 webpages analyzed, from elections 2002-2016, with the researcher assigned ID, candidate name, harvest version / snapshot date and time, URL, and page title for the article/topic.

Two copies of the topic webpage and assets were saved to a cloud folder for the researcher. One was a straight copy from the Replay UI service, and the second was used in the analysis and was edited to remove the additional Open Wayback code. A sample data collection sheet is in appendix A.

**Table 3.3.** U.S. Selection of political candidate archived webpages on issues.

| ID | Candidate | Harvest Version (YYYY-MM-DD-HH-MM-SS) | URL | Article Title |
|---|---|---|---|---|
| **pc02a** | Jean Carnahan | 2002-10-05-10-29-11 | http://www.jeancarnahan.com/news/releaseview.cgi?prtid=16 | Carnahan Launches Ads Against Social Security Privatization Privatization Schemes Force Reductions in Social Security's Guaranteed Benefit |
| **pc02b** | Tim Johnson | 2002-10-12-12-59-44 | http://www.timjohnsonforsd.com/workinghard/agriculture.php | AGRICULTURAL ECONOMY |
| **pc02c** | John Thune | 2002-10-13-05-58-12 | http://www.johnthune.com/issues.asp?formmode=issue&id=3 | Agriculture |
| **pc03d** | Mary Landrieu | 2002-10-15-10-47-51 | http://www.marylandrieu.com/issues_adoption.html | Adoption |
| **pc02e** | Suzanne Haik Terrell | 2002-10-15-10-36-44 | http://www.suzieterrell.com/plan_economic.html | Providing Economic Security |
| **pc04a** | Mel Martinez | 2004-10-09-16-12-05 | http://www.melforsenate.org/ | Fighting for Florida Families |
| **pc04b** | Betty Castor | 2004-10-30-02-58-04 | http://www.bettynet.com/site/pageserver?pagename=iss_economy | A Plan to Move Florida's Economy Forward |
| **pc04c** | Tom Coburn | 2004-10-22-02-19-36 | http://www.coburnforsenate.com/prescription.shtml | Dr. Coburn's Five Point Prescription for Better and More Affordable Health Care |
| **pc04d** | Brad Carson | 2004-10-10-03-43-43 | http://www.bradcarson.com/agriculture/ | Growing A Strong Oklahoma |
| **pc04e** | Pete Coors | 2004-10-12-01-17-06 | http://petecoorsforsenate.com/issues1.htm | On The Issues - Jobs and the Economy |
| **pc06a** | Jon Tester | 2006-10-18-18-01-11 | http://testerforsenate.com/issues | Jon Tester on the Issues |
| **pc06b** | Conrad Burns | 2006-10-11-21-56-05 | http://www.conradburns.com/issues/details.aspx?id= | Agriculture |
| **pc06c** | Jim Webb | 2006-10-04-19-08-42 | http://www.webbforsenate.com/issues/issues.php#iraq | Iraq |
| **pc06d** | George Allen | 2006-10-18-18-14-02 | http://www.georgeallen.com/site/c.hgITL5PKJtH/b.1528127/k.B841/Taxes.htm | Taxes |
| **pc06e** | Jim Talent | 2006-10-18-18-25-29 | http://www.talentforsenate.com/issues/default.aspx?id=1 | Agriculture |

**Table 3.3. (continued).**

| ID | Candidate | Harvest Version (YYYY-MM-DD-HH-MM-SS) | URL | Article Title |
|---|---|---|---|---|
| **pc08a** | Mark Begich | 2014-11-04-22-30-41 | http://www.markbegich.com/priorities/fiscal-responsibility/ | Fiscal Responsibility |
| **pc08b** | Ted Stevens | 2008-10-16-03-33-26 | http://tedstevens2008.com/issues/access-to-federal-lands/ | Access to Federal Lands: Making Traditional Use of Public Lands |
| **pc08c** | Jeff Merkley | 2008-10-15-21-00-05 | http://www.jeffmerkley.com/2008/09/growing_rural_o.php | Growing Rural Oregon |
| **pc08d** | Gordon H Smith | 2008-10-16-01-32-07 | http://www.gordonsmith.com/issues/details.aspx?id=27 | Ensuring Our Communities Are Safe |
| **pc08e** | Frank Lautenberg | 2008-10-29-21-52-16 | http://www.lautenbergfornj.com/issues-homeland-security-and-combating-terrorism.php | Homeland Security and Combating Terrorism |
| **pc10a** | Michael Bennet | 2010-10-15-01-18-54 | http://www.bennetforcolorado.com/issues/details/2010-09-building-a-21st-century-economy | Building a 21st Century Economy |
| **pc10b** | Ken Buck | 2010-10-08-18-02-51 | http://buckforcolorado.com/social-security | Social Security |
| **pc10c** | Pat Toomey | 2010-10-14-22-27-08 | http://www.toomeyforsenate.com/content/jobs-and-economy | JOBS AND THE ECONOMY |
| **pc10d** | Joe Sestak | 2010-10-14-23-25-40 | http://joesestak.com/Economy.html | ECONOMY |
| **pc10e** | Patty Murray | 2010-10-14-22-20-12 | http://www.pattymurray.com/issues?id=0005 | Agriculture |
| **pc12a** | Dean Heller | 2012-10-17-19-46-47 | http://deanheller.com/issues/ | Growing the Economy |
| **pc12b** | Rick Berg | 2012-10-17-20-47-52 | http://www.bergfornorthdakota.com/view/featured/issues/jobs-and-the-economy?ref_v=2 | Jobs and the Economy |
| **pc12c** | Richard Carmona | 2012-10-03-17-48-22 | http://www.carmonaforarizona.com/priorities/creating-jobs | Creating Jobs |
| **pc12d** | Jon Tester | 2012-10-17-19-36-04 | http://www.jontester.com/issues/creating-jobs/ | Creating Jobs |
| **pc12e** | Bob Casey Jr | 2012-09-06-01-03-23 | http://bobcasey.com/pennsylvania-jobs | PENNSYLVANIA JOBS |
| **pc14a** | Scott Brown | 2014-10-07-23-42-54 | https://www.scottbrown.com/issues/ | Issues |

**Table 3.3. (continued).**

| ID | Candidate | Harvest Version (YYYY-MM-DD-HH-MM-SS) | URL | Article Title |
|---|---|---|---|---|
| **pc14b** | Mark Begich | 2014-11-04-22-30-41 | http://www.markbegich.com/priorities/fiscal-responsibility/ | Fiscal Responsibility |
| **pc14c** | Ed Gillespie | 2014-10-15-01-27-03 | http://edforsenate.com/eg2/replacing-obamacare/ | Replacing Obamacare |
| **pc14d** | Dan Sullivan | 2014-10-14-22-06-22 | http://www.sullivan2014.com/jobs_the_economy | Jobs &amp; The Economy |
| **pc14e** | Jeanne Shaheen | 2014-10-14-22-29-36 | http://jeanneshaheen.org/priority/womens-rights/ | Women's Rights |
| **pc16a** | Maggie Hassan | 2016-10-19-00-37-16 | http://maggiehassan.com/priority/combating-substance-abuse/ | Combating the Heroin &amp; Opioid Crisis |
| **pc16b** | Kelly Ayotte | 2016-10-11-23-33-26 | http://www.kellyfornh.com/media-center/get-the-facts/college-affordability/ | Kelly is working to make college more affordable |
| **pc16c** | Pat Toomey | 2016-08-17-01-08-31 | https://www.toomeyforsenate.com/iran_isis | [On Iran & Isis] |
| **pc16d** | Katie McGinty | 2016-10-11-23-17-27 | http://katiemcginty.com/issues/#jobs | Issues |
| **pc16e** | Joe Heck | 2016-10-11-23-20-19 | https://drjoeheck.com/on-the-issues/ | JOBS &amp; ECONOMY: |

## Summary

A media archaeological method is used for this project with a historical methodology involving document analysis. This framework is especially useful in investigating contemporary histories. Instruction manuals and guidebooks are used as artifacts from the time they were published with specific strategies of recommended SEO and SMO tactics. Two sources of primary documents were used in verifying the actualization of these strategies by using web archives and selecting articles from the *Los Angeles Times* and topic pages from U.S. Senate candidates in close election races. This

81

combination of manuals and archived webpages allows for a close investigation of

specific trends and applications of SEO and SMO strategies over time.

CHAPTER IV

COMMUNICATION SYSTEMS FOR INFORMATION RETRIEVAL

The context for a media archaeological analysis needs to be established in an exploration of the systems that it exists within and its relationship to prior communication and media systems. The starting point for a media archaeological analysis should involve a diagram of the systems and information (Parikka, 2011). This step in the analysis is important in order to capture the complex operations in the communication system and areas of study. The term "information retrieval" is introduced with computerized information systems. However, the goals are consistent in earlier forms of media, where indexing provided a universal means of "search" and finding content within a large corpus (Krajewski, 2011). Information retrieval is the basic process in which communication of information is exposed and made accessible. This diagraming step in the media archaeology method is important in order to capture the complex operations in the communication system and areas of study and identify the processes in historical context. The following series of diagrams illustrates how search and retrieval can be conceived through various media.

**Information Retrieval in Print Mediums**

Figure 4.1 is a search / information retrieval diagram, which presents the system of indexing for information retrieval with print materials. This diagram is attempting to abstract to the level where it could be used for a library catalog or other standard index used in order to find information across a corpus or collection.
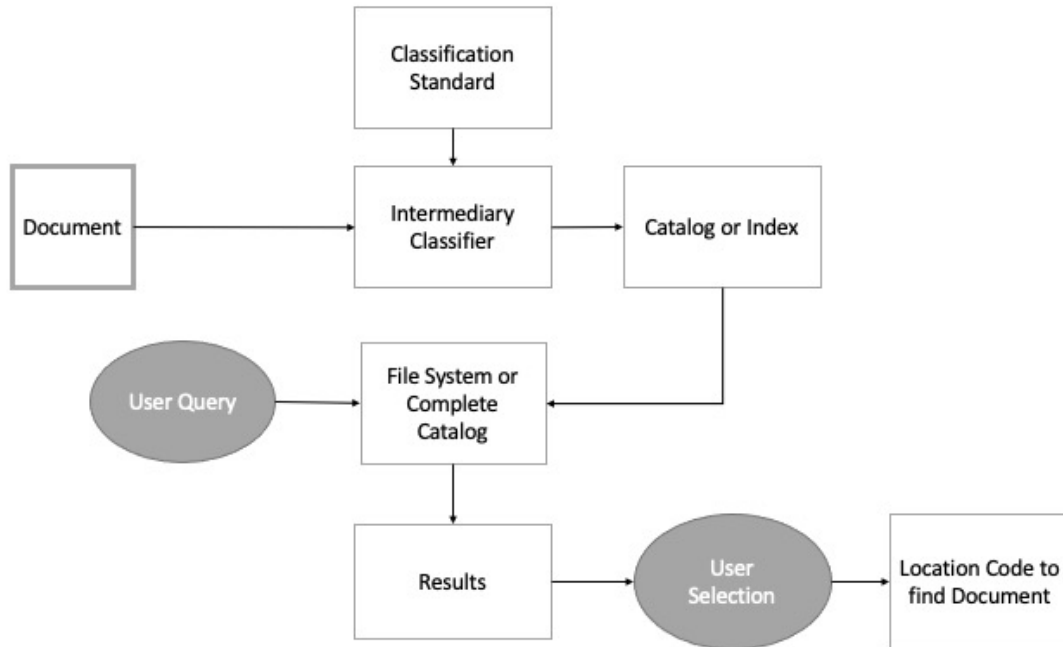
**Figure 4.1.** Diagram of search in a print catalog or filing system.

In this model, there are two particular points where the standard classification is used, in order to facilitate accessing information: classification standard as used by the intermediary classifier and the files system or complete catalog through which the user searches. In order for the user to be able to employ the catalog, the system must be transparent and use terminology that is understandable to the user. The intermediary classifier as an actor in the process could be a person or an automated process, both of which follow the rules in the standard or classification system in order to categorize and organize documents. In this system, the creator of the document has little to no control over how the document will be classified in this process. Once it is given to the system, the structure of the system and its principles guide the next steps in the classification and

categorization. The user and information seeker are also limited by the rules and conditions of the classification and categorization system.

## The "Memex" for Information Retrieval

As new forms of media were developed to store documents, such as microfilm, conceptual ideas of how to search vast amounts of information began to be developed. In the essay, "As We May Think" by Vannevar Bush, the idea of a "memex" machine is imagined. This information retrieval system is perfectly situated for the individual researcher / scholar. The information is stored on microfilm and queried through a device on a desk. Additionally, in this system, the researcher is able to define relationships between documents and integrate their own notes into the information storage system.
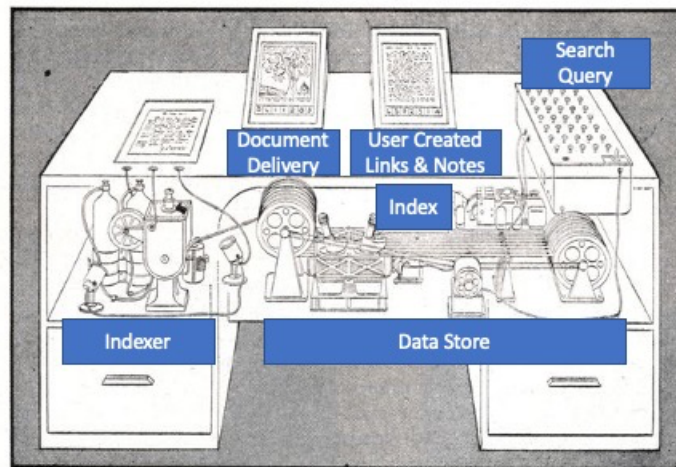


MEMEX in the form of a desk would instantly bring files and material on any subject to the operator's fingertips. Slanting translucent viewing screens magnify supermicrofilm filed by code numbers. At left is a mechanism which automatically photographs longhand notes, pictures and letters, then files them in the desk for future reference.

**Figure 4.2.** Annotated diagram of the Memex conceptual communication and storage and retrieval machine from "As we may think" (Bush, 1945).

Although never built, the Memex is considered to be an early inspirational model that was used by early Internet designers in the creation of the World Wide Web (Houston & Harmon, 2007). In this system of information retrieval, that addition of linkages and relationships as an essential part of the structure introduced a new component to be considered in retrieving relevant information.

## Information Retrieval in Databases

Figure 4.3 illustrates a generic system of information retrieval in a textual database of documents. Similar to the print information retrieval system, the primary difference in this database model is the ability to query data through a query engine and automated functionalities and the significantly increased capacity to query at scale through computerized systems. The diagram is not intended to replicate a technical diagram of a database infrastructure but rather to highlight the places of interaction between the document, user, query and results. In this model of information retrieval, both the document creator and user are limited to classification standard applied by a third party (machine or person) in order to retrieve relevant documents.

The diagram is assuming that the database can store electronic copies of the documents, in addition to the indexed data in the database; however, instead of document delivery of an electronic document, this system could also return a locator in a classification system, which is needed to then retrieve the document from another system. The basic system does not change, although, the user experience may be greatly enhanced by the capacity to deliver electronic documents.
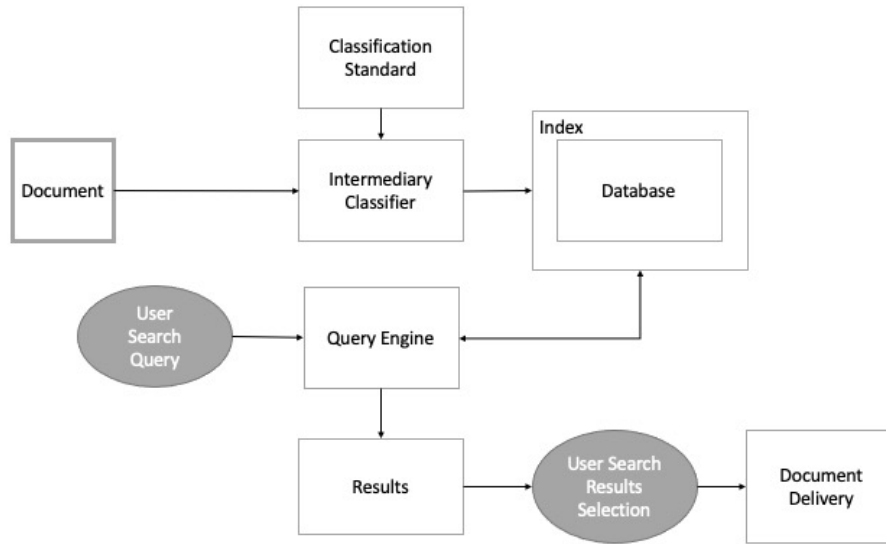
**Figure 4.3.** Generalized diagram of text information retrieval systems and search queries.

### Information Retrieval on the World Wide Web

With the World Wide Web and the creation of search engines, the system for information retrieval evolved. The structure of the HTML document itself as machine readable content changed the way that content could be classified and categorized. The creator of the HTML document could put code in the document HTML `<head>` section to call out to the search engine, `<meta name="robots" content="all" />`, or to not index the page `<meta name="robots" content="noindex" />` and give additional directions to the search engine.[32] These specific mechanisms and extra processes are represented in Figure 4.4 through the double-arrows between the HTML

---

[32] See: Google Search Central, "Robots meta tag…,"
https://developers.google.com/search/reference/robots_meta_tag.

document code and the search engine web crawler. The creator of the HTML document can also provide information through the website directory to the search engine through a "robots.txt" file whose the primary purpose is to specify what should not be crawled by the search engine.[33] The content creator has an additional mechanism to send information to the search engines to crawl their content through the creation and submission of sitemaps.[34] Sitemaps are like an architectural map or guide to important content on your website. For small or simple websites, a sitemap is often not needed. The multiple methods that the document creator has to communicate to the web crawler / indexer are unique to this communication and information retrieval system.

The HTML document's data is processed by the search engine web crawler. If SEO has been applied, then it becomes an additional way for the creator to influence how the content is indexed within the search engine and ideally increases the chances of the content being found by the search engine users. In this stage, many search engines also provide tools to help document creators select keywords based on trends and user searches, which in some ways becomes like a stand-in for the taxonomy and vocabulary guides in prior information retrieval systems. SEO strategies applied also increase the communication between the HTML code, page content, and the web crawler.

Once the web crawler has parsed the code on the website, the rules of the search engine begin the gatekeeping function. Typically, search engines publish their primary

---

[33] See: Google Search Central, "Create a robots.txt file," https://developers.google.com/search/docs/advanced/robots/create-robots-txt.

[34] See: Google Search Central, "Build and submit a sitemap," https://developers.google.com/search/docs/advanced/sitemaps/build-sitemap.

rules and best practices in order to encourage good behavior and also prevent Black Hat techniques. For Google, this is where mobile friendliness, accessibility, and well formatted HTML code are evaluated. During this phase, Black Hat techniques are also assessed, and a webpage/site may be banned from Google or not indexed depending on how the site's algorithms interpret adherence to the rules. There are many reasons why a webpage may have what seem like Black Hat techniques that are not. A notable conflict between good communication practices and web search engine rules is the practice of citations. If an essay on a webpage has many citations that link to other webpages, and those webpages do not link back, which may be part of the medium within the webform (e.g., linked webpages were published earlier and now static), then the webpage may be flagged for link farming and not indexed. Once a webpage passes the search engine rules gate, it is then indexed within the search engine data store. A cached version of the webpage may also be stored in the search engine index at this time.

Proprietary algorithms process the data and determine when to retrieve content for user searches. This is a black box of proprietary information. The process is not visible to the document creator or the user. Page content, functionality added through SEO, linking and relationships, and keyword searching is evaluated. In addition to the search and retrieval algorithm, the search engine may also employ language translation, localization data, user habits, and other filtering at this stage. Critical communication and code studies have called for transparency surrounding this black box in order to expose the restrictions, influence and/or change biases that the search engine employs at this stage.

The results from the algorithms are sent to a search engine results page (SERP) [35] where organic search content may appear alongside paid search content and knowledge panel information. The SERP may vary depending on the hardware (device) and software (browser) that the user employed to conduct their search. The SERP usually contains a short description, which may be taken from a metadata tag in the `<head>` or generated from the initial page `<body>` content. Once a search result is selected by a user, the rendering of the HTML document opens in their browser. This resulting view of the HTML document is also dependent on the user's hardware and software.
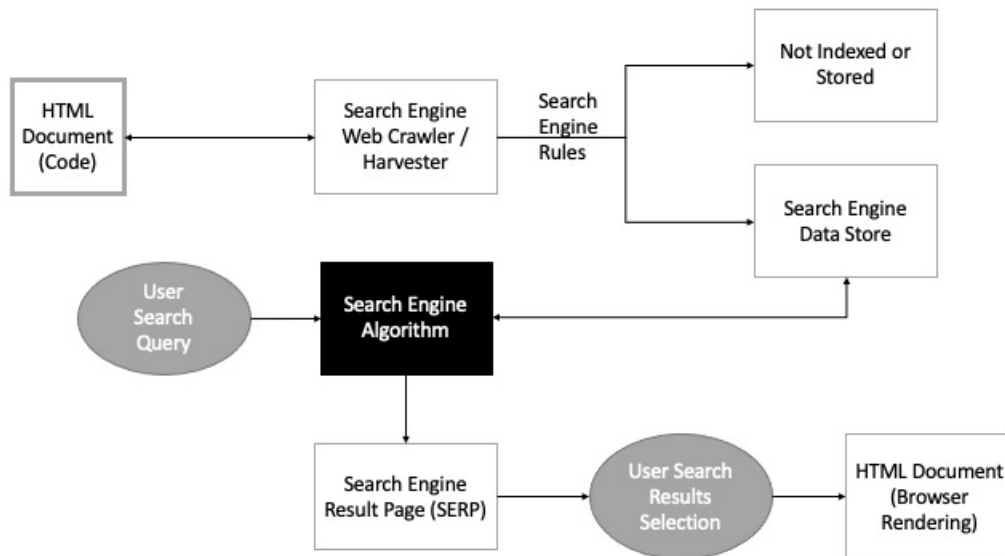
**Figure 4.4.** Internet search and retrieval using a search engine.

---

[35] Knowledge panels in Google are for people, places, organizations, and things. Open graph metadata code must be provided, in order to support knowledge panels in Google. See: https://support.google.com/knowledgepanel/

Overall, the search engine information retrieval system within the World Wide Web has four distinguishing characteristics from prior information retrieval communication systems: 1) the dialogue between the document creator and the indexing service, 2) the closed algorithm box, 3) addition of evaluation relationships between documents in the retrieval, and 4) the impact of the hardware and software of the user on the delivery of the document.

## Information Retrieval in Social Media Platforms

Through Figure 4.5, the process of on-page social media optimization for information retrieval is illustrated. The HTML page is initially exposed through a social media platform, which is an operation that can be either manually or programmatically instigated. In the case of social media operation, the use of user data and targeted streams rely on this initial user-mediated share of content through the social media platform, although that user could be the same person as the document creator. The ability for the document creator to affect the social media integration is enabled through the SMO on-page techniques that add specific metadata to the `<head>` of the HTML page, in order for content to display well within the social media application's interface.

The social media platform's algorithms then define additional exposure through the platform and render a display of the content. The proprietary algorithms that promote content in the social media platform feeds are a black box and unknown to the user or document creator. At this stage, the social media platform algorithms also review content posts that link to external HTML pages and can evaluate for localization, user habits, fake news, copyright violations, and content that otherwise violates the social media

platform's policies. Interestingly, most of the strategies to automate or manually address violations of content have failed in social media platforms whether that is fake news that persists on Facebook or false-violations from copyrighted bots that remove content from Facebook and YouTube.

The display rendering in a user's feed will depend on the hardware and software of the user. This pattern then repeats with additional social media shares, reactions (i.e., likes), and comments for the content. In some cases, the strategies of the social media platform can also be manipulated using spam accounts and bots to increase content exposure through the platform. Popularity and targeted results to an individual's search and browser history are a features of this method of retrieval.

In the model of information retrieval within social media platforms, the HTML document creator has little ability to influence the alignment of information provided on the HTML page with the social media feed results generated by the platform. Although most social media platforms do have a search feature, the majority of content will be viewed through this feed, which is more of a browse for information seeking behavior. (This does differ for multi-media dominant sites like YouTube.) The result of HTML content shared through social media platforms is that instead of social media optimization structures, although they are important, the primary method for increasing content shares is through construction of images and text that encourage interaction of users, i.e., clickbait.
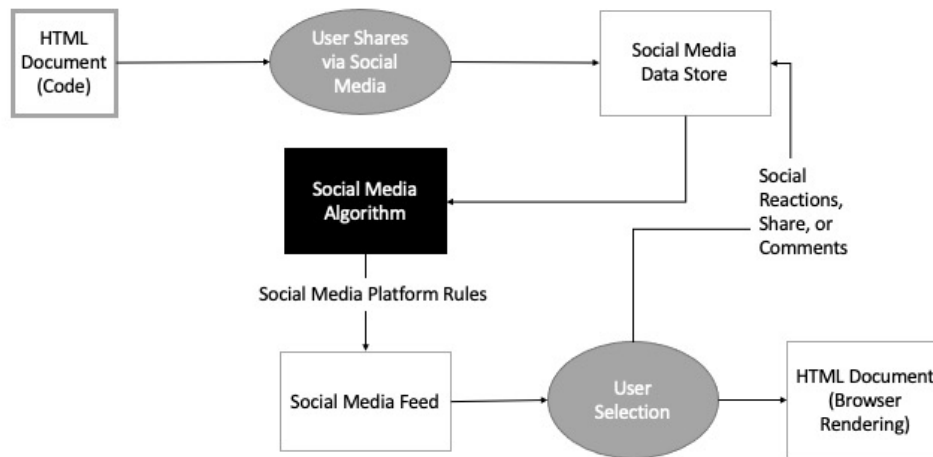
**Figure 4.5.** HTML content found through social media platforms.

## Summary

By reviewing abstract diagrams of information retrieval through communication systems of various media, search and information retrieval on the Internet can be contextualized within technologies and practices that came before. The overall goal throughout these systems is to make document content accessible to a user. In the early models of print information retrieval, the size of the corpus and time needed to search print catalogs was a limiting factor. Across all the models, the role of politics of information and gatekeeping is a determining factor in access to the content. In the earlier models, the taxonomies and vocabularies were necessarily transparent to the user, whereas, in the later models, there is some transparency with best practices and tools published by search engines and social media platforms, but the primary function that returns information is hidden in proprietary algorithms. Within these algorithms, the

criteria used to retrieve documents is expanded from the more traditional subject, author, place, time retrieval to include relationships, format, and other characteristics defined as important to the search engine.

As search engine and social media optimization strategies are evaluated in the next chapters both in terms of recommendations in manuals and actualization of strategies in newspaper article and candidate web pages, the broader context of these communication systems for information retrieval is necessary. Chapter V will look at search engine and social media optimization strategies over time as recommended through instruction manuals and guidebooks, all within the communication information retrieval systems illustrated in Figures 4.4 and 4.5 in this chapter.

CHAPTER V

HOW-TO GUIDES AND INSTRUCTION MANUALS FOR SEO AND SMO

Using technical instructional manuals and how-to guides in order to explore

historic communication practices is an efficient and effective method for analysis (Prior,

2003). The analysis in this chapter has two functions. First, it provides a look at the

changes in SEO and SMO recommendations over time while looking for topoi in the

context of pre-existing media and mechanisms. As the Internet becomes more important

to communication, the role of search engines and social media applications take a further

hold as the gatekeeper to information. Secondarily, it provides a basis for comparison of

the SEO and SMO strategy recommendations in the manuals to the subsequent HTML

web-page analyses of chapters VI and VII. The recommendations contained in the how-

to-guides and instructional manuals for SEO and SMO strategies include HTML code-

specific format and structure suggestions, writing, design, and content categorization.

This chapter presents an analysis of how-to guides and instruction manuals for

SEO and SMO strategies published from 2005 to 2018. Because technical manuals are

often discarded when they become obsolete or replaced by new versions, the first

available manual for the study was from 2005.  Selection of the guides was limited to

availability through libraries and used bookstore platforms. The last manual examined

was published in 2018, the year this analysis began in earnest. The analysis is structured

into two sections. The first section includes analysis of the goals, audiences, and authors

of the texts, in order to set the context for the recommended strategies in the texts. The

second section reviews various SEO and SMO strategies and tactics and is broken into

subsections of structural advice, tags, tools, and composition and design of webpages and

analyzes the strategies for identification of topoi and media and communication practices that transcend or evolve with webpage content guarded by the Internet's online gatekeeping tools which control access to information.

## Goals of SEO and SMO Manuals

In examination and analysis of the SEO and SMO instruction manuals and how-to-guides, the expertise of the authors and the audience and aims of the texts are important to contextualize.  The overall aims of the texts focused on making information on web pages findable online and had similar characteristics, since the texts were identified in Worldcat.org using the following subject heading terms: Internet Marketing; Social Media; Web Search Engines; Internet Searching: Search Engines; Electronic Information Resource Searching; and Electronic Commerce. The biggest differences in the texts are the amount of text devoted to SEO and SMO strategies specifically within the code and technical instructions. Some texts devoted a large space to tools and coding, while others noted things that you might want to hire someone else to do for you.

### Authors

The majority of authors' expertise was established within the marketing and advertising professional fields. The most common expertise touted was web marketing and/or SEO consulting by firms with clients such as Disney, Nike, L'Oreal, and the BBC. A few were accompanied by technical editors and co-writers or focused on web skills, in particular. The length of experience was a continual theme in expressing expertise; however, the form of that experience differed greatly. From one of the earlier manuals

(2005), one of the author's listed qualifications was that they had been using the internet since 1987. In a more recent manual (2018), twenty years of experience in digital marketing is also listed among the author's qualifications. Of course, digital marketing had not existed for twenty years when the 2005 manual was written. The newer manuals also tended to include speaking and conference engagements as author qualifications, as well as industry awards. One author self-claimed to have "pioneered the video search engine optimization phenomena" (Bradley, 2015). The expertise presented by the authors, as a whole, was constructed in ways to demonstrate concrete impact in the industry and helped to set up a practical and useful approach to the recommendations with the texts.

**Audiences**

The majority of the manuals were primarily aimed at a small business audience, with the focus on marketing and getting products to appear in search results. Additionally, a few of the texts noted how the strategies were also very useful for non-profits and journalists (Kelsey, 2016), college and graduate students (Redish, 2014; Rowles, 2018), or terminology in order to communicate with members on a team within a larger organization who would do this work (George, 2005; Lincoln, 2009; Odden, 2012). A few of the texts were also aimed at web designers and web system architects or those who may be interested in starting their own SEO business (George, 2005; Ledford, 2009; Shenoy & Prabhu, 2016). One manual, *Search Engine Optimization: Your Visual Blueprint for Effective Internet Marketing,* had a complete section of the text devoted to those who wanted to make money by selling ads through their websites and how to find a topic and content that would entice others to then pay you to include their ads on your

webpage. One of the most common examples of success using this approach is within the online mattress sales industry. This audience deserves further study in another context where the information on the webpage is desired to have top results in search engines specifically in order to court advertisers. The lack of more technical and code-focused audiences for the manuals may be due to the terms used to select the manuals. However, it is also interesting that part of the strategy of these texts was to work with audiences that may not have been technical and still take advantage of the technical aspects of SEO and SMO strategies for the findability of content. It's only noteworthy that more advanced guides were not retrieved with the search criteria in Worldcat.org or Amazon.com for the terms "search engine optimization" and "social media optimization."

**Approaches**

The manuals were divided along two primary approaches of the texts: 1) those that focused on tools, and 2) those that focused on adapting skills from non-digital arenas, such as communications and marketing to the digital environment. "Give Google exactly what it wants and needs. In return, Google will give you exactly what you want...page one domination!" (Bradley, 2015, p. xv) is an excellent exemplar of the first approach. Whereas, other texts made it explicit that focusing on technology and tools were not goals of the manual: "Communications and selling are the keywords. Not technology" (Lincoln, 2010, p. xvii), and "*Letting Go of the Words* is about strategy and tactics, not about tools. Technology changes too fast to be a major part of the book – and the principles of good writing transcend the technology you use" (Redish, 2014, p.xxvi). Interestingly, there was not a clear connection between texts that focused on tools and

technology with a more technical audience. The difference in goals and approach to SEO and SMO optimization varied even with the same audience, such as small businesses.

In the area of search engine optimization, all of the texts focused on Google as the primary search engine. Some manuals did include additional search engines, such as Yahoo and Bing. As noted in several of the manuals, because the search engine algorithms are proprietary and generally use similar principles, working on SEO strategies for Google should be translated to competing search engines as well. Interestingly, even with the focus on Google, many of the texts did not attempt to describe the various significant algorithm changes that have affected how SEO and SMO work within Google. The exception to this was *Win the Game of Googleopoly: Unlocking the Secret Strategy of Search Engines*, which was one of the more tools and technology focused texts and cited the changes of Googles' Penguin, Panda, and Hummingbird releases in particular. Although, the algorithms were not especially called out in the majority of manuals, the recommendations shifted over time in accordance with the algorithmic changes, particularly with the `<meta name="keywords">` tag, as we'll see in the next section.

One of the major challenges of social media optimization as a whole field is the need to select and adapt content to the various social media platforms. To address this issue, several texts included goals on how to select the appropriate social media venue for one's content. Many, of course, are no longer major social media venues, such as MySpace. In this project, on-page social media optimization in webpages is the focus because of the ability to influence social media optimization through the HTML webpage code itself, rather than within each different social media platform. Eight of the fifteen

manuals included specific advice for social media optimization. Because Twitter and Facebook are two of the only social media applications that provide specific criteria for on-page social media optimization, it is not surprising that the manuals focused on these two platforms. The goals described in the texts for these sections were more akin to the tool and technology focused texts and about crafting the format in the social media optimization particularly for the platform.

## SEO and SMO On-Page Strategies

The following section will review and analyze the on-page strategies and recommendations for SEO and SMO with HTML pages found within the instruction manuals and how-to guides in order to identify topoi. Critical to the recommendations within the books is the recognition that search engines and social media applications filter and promote web content based on webpages not websites. Because of this, each page within an organization's or individual's website should include SEO and SMO strategies particular to the content on that page. An important distinction of new media communication on the Internet is this disaggregation of content and context. Much as Apple's iTunes changed the relationship to music and the listener from albums to individual songs, the work of search and social media tools as a primary gate to web content has had the effect of separating the pages from its website as the primary unit of access and consumption.

In addition to the focus on webpages as distinct entities, search engines specifically look at the relationships between links both within websites and to / from external webpages. Those connections as coded and described within a webpage are also

important on-page SEO and SMO factors. "Most SEO efforts are focused on web pages. Effective web page optimization includes a consideration of the individual page as well as its relationship with other pages on the overall website" (Odden, 2012, p.133). This presents an interesting challenge in the context of the webpage and the website, where connections need to be explicit in order to establish the relationships with other pages on the website. As on-page optimization strategies are reviewed in this section, recommendations are included both for how a page declares and describes itself and how the relationships with other web entities and content are described. "Google admits that there are over 200 signals (factors) that it looks for when determining how to rank your website" (Bradley, 2015, p. 59). Part of the work of the manual and how-to-guide authors is to identify strategies that can make a significant difference. As a result of what has been exposed by the corporations that own search engine and social media platforms, there is much alignment in the strategies presented throughout the manuals.

**URL Optimization**

Each webpage has a URL, which is the address / locator for the page on the Internet, and its primary component is a domain. Optimization strategies for the URL of a webpage are consistent through the search engine and social media optimization guides in the importance of a domain name that is named appropriately with content and ideally matches most likely search terms (Lincoln, 2009; Odden, 2012; Redish, 2014; Rowles, 2018; Shenoy & Prabhu, 2016). A manual from 2008 stresses the importance of limiting the usage of a sub-domain in domain strings (Michael & Salter, 2008) as interfering with search engine optimization. However, that advice is not consistent throughout the guides,

but it does appear in a 2015 guide warning that subdomains lack the credibility with

search engines and social media platforms that a primary domain holds (Bradley, 2015).



**Figure 5.1.** Basic anatomy of a URL (*Technology for Librarians 101: Anatomy of a Web Address*, 2014).

The domain acts as a foundation for web content, and value ads from content or

other structures on a webpage are credited back to the domain in search engine systems

(Bradley, 2015). Tips on domain creation and naming are fairly consistent throughout the

publication dates of the manuals examined in this project. Strategies suggested include

using short, clear, and descriptive (not technical) wording in domain names. Formatting

considerations include not using any special characters, other than the hyphen, and

always be lowercase. "Never, ever in a million years allow your URLs to have uppercase

characters" (Bradley, 2015, p. 63). In the early 2000's the popularity of content

management systems such as Drupal and WordPress for creating and managing large

amounts of web pages became popular. These systems still exist today. The creation of

dynamic URLs for webpages, which are created on the fly or have a designation such as

"`pgid="19"`" at the end of the URL, lack a human readable identity or meaning within

the URL often have scripts generating them, cause a problem for search engines and are

very difficult for them to read (George, 2005, p.26). Successful content management systems integrate stable URL paths for webpages in order to overcome this problem (Jones, 2013; Ledford, 2009).

Additional strategies include using localization with a geographic name in the building of domain names (Bradley, 2015; Lutze, 2009; Shenoy & Prabhu, 2016). If you are looking to advertise a physical service in an area such as Eugene, Oregon, one might use a domain like, "https://pizza-eugene-or.com." The advantage of this precise location serves not only to specifically promote where one's business is located, but will also help eliminate out of area hits. This could be very important, for example, if the business is receiving takeout and delivery orders from Seattle because of a similar pizza restaurant name. Having a geolocation in the URL does not overrule the strategy to have short and memorable URLs or ones that correspond with branding. Geolocation data can be added in different places within the rest of the webpage. Top-level domains can also be based on geographic regions, such as "https://pizza.uk." Application of top-level domains may also be chosen because they sync with a brand but are registered with ICANN as country-level top-domains. This can increase issues with search engines, social media platforms, and government blocking tools. For example, the website for the Open Graph Protocol specification uses "http://ogp.me," and ".me" is the geographic code for Montenegro. The ".me" top-level domain became popular during the 2010's as a way of personalizing sites (*6 Reasons Why We Like .ME Domain Names*, 2013).

Although all domain names are maintained through ICANN, a non-profit, the consumer level domain naming process goes through a registry and a reseller, often a

hosting platform or a company that specializes in selling domain names. See *Fig. 5.2* for

the relationship between entities involved in domain name processes.  Because of the
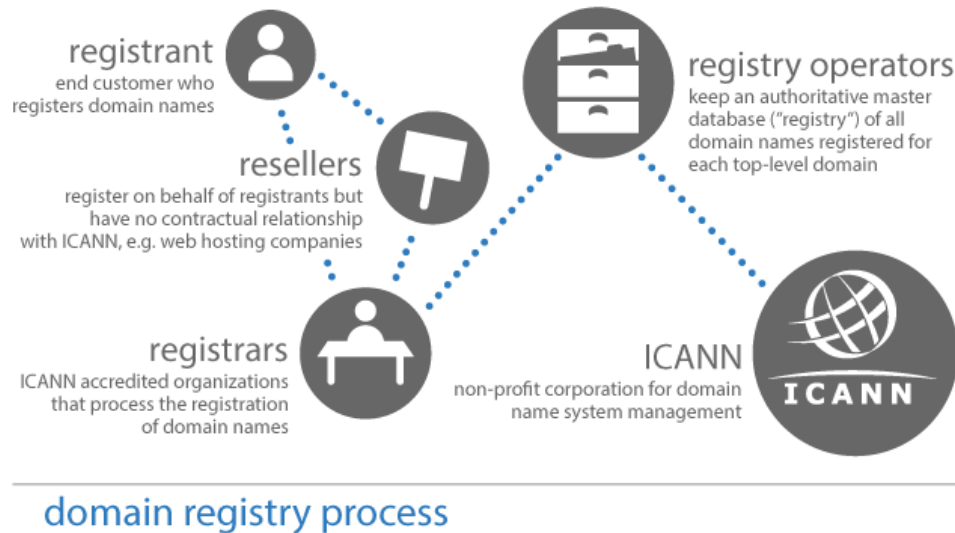


**Figure 5.2.** Domain registry process (*Domain Name Registration Process | ICANN WHOIS*, n.d.).

limited supply of domain names using particular words, resellers can request high prices

for the domains most in demand (George, 2005; Lindenthal, 2014). Search engines,

however, have also become aware of this problem, and over time have relied more

heavily on the domain, path, and page name as a whole to use for search matches

(Rowles, 2018, p. 87). In addition, the URL has become a less important factor in search

engine ranking (Bradley, 2015).  In order to use this strategy, a categorization of themes

and hierarchies should be used to construct the path of the webpages within your website

(Jones, 2013, p. 76). Simplicity wins over cleverness in the URL domain naming for

search engine and social media optimization.

**Strategies within the HTML Page's Header**

The structure of tags allowed in HTML code set the framework for search and social media optimization practices. Within the first part of an HTML file is the header, `<head>`, section of the HTML document. Most of the HTML tags and content are hidden to the viewer upon rendering, except for the `<title>` tag that may appear in a browser tab or window label. Within the `<head>` section of the HTML document, scripts, styles, and other coding that enumerate the style and display preferences can be specified for a browser and communicate with search engines and social media platforms. The basic structure of a webpage looks like the following, as illustrated in Figure 5.3.

For the most part, tags within the `<head>` are written primarily for machines. The code in this area is not rendered visible for the human reader on a browser, except for the `<title>` tag, which may appear in the tab or window at the top of the browser.

```
<!doctype html>
<html>
  <head>
    <title>Untitled Document</title>
    <meta name="description" content=
      "This page is about unique content">
  </head>

  <body>
    <h1>Main Heading</h1>
    <p>Content Here.</p>
  </body>
</html>
```

**Figure 5.3**. Basic HTML structure.

Additional code in the `<head>` may also include scripts to call particular functions and features for the webpage, as well as references to style instructions to the browser and

device for how to render the page on the screen. The `<meta>` tags available in HTML, which represent assigned metadata to the pages, include a fairly traditional set of metadata that can be applied (i.e., author, date, language, keywords, category, abstract, rating), as well as instructions on how to render the page on a screen (i.e., viewport). However, only a subset are considered important for search engine and social media optimization. Search engine optimization, in particular, uses the description and keyword tags with varying importance over time. (The "viewport" meta tag becomes important when mobile sites become popular.) Whereas, social media optimization uses a schema or protocol adopted by the platform that can be referenced through the HTML code structure in `<meta>` tags and is aimed at that particular platform. The content in these tags may also duplicate content in the standard HTML `<meta>` tags. In addition to the use of the `<title>` tag in the browser tab or window label, it is coupled with the `<meta name="description">` on the search engine results page and often in social media posts referencing a link. Because of the reuse of the data from these two tags within the search engine results pages, they have a particular importance in the user search engine experience, in addition to the search engine indices, and are the primary human-readable / user interactions with coding in the `<head>`.

### *Title*

The `<title>` tag has a consistent importance over time for SEO and SMO in the manuals examined. The earliest manual I examined points to the `<title>` as a primary area to focus on for SEO (George, 2005, p. 39), and the most recent notes "This [title] is the most important thing on the page, as it is generally given the greatest weighting by

the search engines…" (Rowles, 2018, p. 85), with notes in-between similar to "one of the most important [factors]" (Lincoln, 2009, p. 77). In line with the structural and coding advice in construction of the titles, the strategies suggested in the manuals follow two primary threads of advice: 1) use keywords to construct your page title, and 2) pay attention to title length. In the use of keywords within a title, strategies include putting the most relevant keywords to the beginning of the title, because the machines read it first, and order is important to the algorithms (Ledford, 2009). The keywords used in the `<title>` should also not be duplicative of other keywords used on the webpage, or the search engine will likely consider it spam (Shenoy & Prabhu, 2016, p. 83). Only a couple of manuals specifically noted the use of a human readable title that would encourage a user to click on the link in search results or a social media platform (Kelsey, 2016; Odden, 2012).

The recommendations for the length of the `<title>` tag, however, are directly related to the user view within a search engine result page or social media platform. The W3C, the body the oversees web standards, recommends 64 characters; whereas Google allows for 66 characters and Yahoo search results allow for 120 characters (Ledford, 2009, p.131). The limit on the number of characters isn't static, however, and another manual published in 2015, recommends limiting to 55 characters for Google search engine results pages (Bradley, 2015, p. 72). One of the most surprising aspects in reviewing these manuals is that while the `<title>` tag may have been listed as one of the most important factors for SEO and SMO, the manuals paid very little attention to strategies for title structure.

The lack of attention to the structure of the `<title>` becomes apparent in the following chapters, particularly with the relationship for the content of the webpage and the title of the website as a whole. The order and stacking of webpage vs. website title vary; for example, `<title>My web page — My web site</title>`, `<title>My website — My web page</title>`, or `<title>My web page</title>`. The `<title>` tag may be the most obvious in looking for topoi and practices that have been forwarded from previous media. The title does maintain an important function in discovery and describing the content overall, as the advice from the manuals recommend. This relationship between scaling of titles is more like the titles for articles and chapters. However, due to the disaggregated nature of the content and separation of context, the web authors often place the larger website title within the tag. This practice is not defined in any of the manuals.

### *Metadata Description*

The metadata tag for the page's description, `<meta name="description">`, is also a highly relevant tag for SEO strategies in the manuals. However, its relevance is a passing reference in the earlier manuals. In manuals post Google's Panda algorithm update (2011), the purpose of the SEO tag is stated as primarily for the user view in Search Engine Results Page (SERP) (Bradley, 2015; Kelsey, 2016; Moran & Hunt, 2015; Odden, 2012). Even with that primary function, it is still pointed to as the second most important factor for SEO in a 2015 manual (Bradley, 2015). In describing the human-readable strategies and functions for the description tag in SERPs, "The more compelling and relevant your meta description is, the more likely it will inspire a click to the web

page. More clicks mean more visitors, but also serve as a signal for potential influence on subsequent rankings. Pages that inspire more clicks may be rewarded with higher search visibility, because users are responding positively to them." (Odden, 2012, p.135). Although "compelling" is hardly a structural strategy, a couple of the manuals do provide further structural guidance.

Structural strategies for the meta description tag focus on two primary components. The first is the length of the tag so that the text is readable on a SERP; 150-155 characters is the limit for that function (Bradley, 2015, p.79; Ledford, 2009, p.137). The second strategy focuses on the content of the tag. It is unclear if the content of the description tag is used in ranking, and some propose that the content within it is treated much as the initial `<body>` text in ranking (Moran & Hunt, 2015, p. 72). The advice not to reproduce too many words of the `<title>` in the meta description serves both machine ranking and human-readable purposes (Ledford, 2009). Because search engines and social media platforms are focused on the uniqueness and page level strategies, it is also important that each webpage have a unique meta description that focuses on the content of that specific page (Bradley, 2015; Ledford, 2009). In some ways, it may seem odd to consider uniqueness a structural strategy. However, because of the parsing that is done by the algorithms for SEO, the placement and repetition of words becomes an important factor.

The role of the `<meta name="description">` on a SERP where the user is deciding whether the content is worth selecting mimics the earlier bibliographic systems and the role of the abstract, especially in article databases. Also similar to earlier information retrieval bibliographic systems, one of the most important factors for the

abstract or `<meta name="description">` is the length and how it fits within the medium. Although it can function for search and retrieval purposes, its primary role is for human-readable determinations.

### *Metadata Keywords*

"If eyes are the windows to the soul, the keyword search is the window to your customer's thinking process" (Lutze, 2009, p. 29). Perhaps one of the most contested and now obsolete SEO strategies is the use of the `<meta name="keywords">` tag. The purpose of metadata keywords was to anticipate the words that a user might type into a search engine and should be a combination of topics, geographic locations, personal names, and genre terms. Searchers typically use two to three keywords, and the webpage should be optimized around those words (George, 2005, p. 66). The value of keywords also depends on the term. A search term that is too general may not be helpful in searches unless you have earned the top spot for a general term, such as "computers," which is nearly impossible (Lutze, 2009, p.9). And with non-organic search, Google has an entire business around purchasing "AdWords," which support non-organic search results at the top of search result pages. All of the manuals except *Letting Go of the Words* have significant information on tools used to generate and test keywords, such as *Wordtracker*,[36] Google's *Keyword Planner*,[37] and *Free Keyword Tool*.[38] These tools are suggested in SEO beyond the `<meta name="keywords">` tag in other structural parts

[36] https://www.wordtracker.com/

[37] https://ads.google.com/home/tools/keyword-planner/

[38] https://www.keyword.io/

of the page where keywords assist in search placements, such as the `<title>`, headings, and paragraph text.

Early SEO manuals suggest that, in addition to the focused keywords of the page, common strategies to include common misspellings and errors should also be added in the `<meta name="keywords">` tag (George, 2005; Lutze, 2009; Michael & Salter, 2008). As algorithms increased in sophistication, automatically corrected searches through the "Did you mean…?" function nullified the need for this kind of keyword application. This is very important as the only place to manually enter common misspellings and not negatively affect the position of the rest of your webpage content was in the `<meta name="keywords">` tag.

One of the primary problems with the metadata keyword tag was trust, as it was one of the first SEO techniques that required major algorithm rewrites by search engines when web page authors employed the Black Hat strategy of "keyword stuffing." "Keyword stuffing" is when keywords and key phrases are "overused in content merely to attract the search engines" (Moran & Hunt, 2015, p. 459). Two main techniques have been identified as keyword stuffing: 1) keyword loading: disproportionate number of words and phrases for the content, and 2) keyword spam: words added that aren't relevant to the content on the page and may be directly targeted to attract traffic from a competitor (George, 2005; Bradley, 2015; Rowles, 2018). Keyword stuffing can occur in a variety of tags across HTML; however, it was most commonly found in tags in the `<head>` that were intended to be machine-readable and hidden from the user through the browser view of webpage (George, 2005, p. 69). With overt "keyword stuffing" prevalent, Google stopped using the `<meta name="keywords">` tag in 2009. By 2011,

with the Panda release, it was officially out of the SEO game. Still, it took time for the

SEO manuals to catch up, and the meta keywords tag is listed as the "single most

important" SEO factor in a text from 2013 (Jones, 2013, p. 40). The best assessment of

the role of the `<meta name="keywords">` tag comes from a recent manual: "These

[metadata keywords] used to be more important, but now they are less so" (Kelsey, 2016,

p. 113). Keyword usage outside of the `<meta name="keywords">` tag remains

important for SEO and is described later in this chapter.

The `<meta name="keywords">` tag is the closest tag functionally to traditional

indexing and classification of information. In traditional systems, this would be in a

controlled vocabulary of genre or subject terms. Even without the controlled source of

keywords and phrases, the purpose of the tag is similar in retrieving relevant information

based on anticipated user searches. In fact, the Library of Congress previously advised

that terms from the Library of Congress Subject Headings Classification[39] be added to the

`<meta name="keywords">` tag (Library of Congress, 2002). These controlled

classifications are also called "authorities" within information science and are still used in

traditional bibliographic systems for information retrieval and extracted or identified as

separate from the rest of the content of the materials. As part of this project, the subject

headings were instrumental in identifying the manuals and guidebooks. As the `<meta
name="keywords">` tag became the must untrustworthy assignment in SEO, it is

interesting that the extracted subject and other keyword definition loses its authority and

the guides recommend that keywords be embedded within the page content. It also raises

---

[39] https://id.loc.gov/authorities/subjects.html

some interesting questions about subject heading assignment in more traditional systems,

what is trustworthy and what communicates the content accurately for retrieval.

**Strategies within the HTML Page's Body**

As the how-to manuals and instruction guides move the bulk of their SEO and

SMO strategies, they point to the primary content of the HTML pages in the `<body>`

section of the page. Many of the manuals emphasize the best strategy is to provide unique

and interesting content (Bradley, 2015; Jones, 2013; Kelsey, 2016; Lutze, 2009; Moran &

Hunt, 2015; Odden, 2012; Redish, 2014; Shenoy & Prabhu, 2016). Perhaps one of the

biggest changes in content online is that the "user is in charge of the conversation" and

that the web creates a "pull" instead of "push" form of communication (Redish, 2014, p.

151).

> *For example, if you did nothing other than write high quality, compelling,*
> *relevant articles on topics related to your business and organization,*
> *Google will find them, and more importantly, people will find them. If*
> *they're relevant and interesting, meaningful or helpful, more people will*
> *share them with other people. If this happens, they will climb higher in*
> *rankings* (Kelsey, 2016, p. 5).

Audience analysis is a major component of the strategy, and a couple of the guides spend

significant time in audience analysis advice (Bradley, 2015; Moran & Hunt, 2015). Some

of the texts concentrate on searchers, as the audience, divided into categories of

navigational, transactional and informational searchers (Moran & Hunt, 2015, p. 35).

Another notes the significance of other web authors as audience and to aim for writing

content that others would want to link to or "remark on" in order to fuel the weight that

search engines and social media platforms give to interrelationships on the web and

popularity as authority (Redish, 2014, p. 74). Even with these distinctions, the general

consensus is to remind webpage authors that they should write content aimed at their

audience and follow best practices for communication. Write for people, not search

engines (Jones, 2013, p. 88); "too often newbies write for spiders alone" (Moran & Hunt,

2015, p. 96). How does this manifest in structural advice for creating HTML webpages

through the guides? None of the texts examined for this project were writing guides.

They provided structural advice both in terms of identifying content and how it is marked

up and structured within the HTML page that focus on a particular way of writing for the

web.

### *The Shape of Content*

SEO and SMO instruction manuals and how-to guides are explicit that the shape

of the content should follow two basic strategies: succinct and distinctive. "Any text not

directly relevant to the content should be removed" (George, 2005, p. 27).  The content

should be "bite-size" and "easy to digest" chunks (Redish, 2014, p. 149) with short

sentences (Shenoy & Prabhu, 2016, p. 83). This advice is similar to much

communications and writing advice; however, it is genre independent in the context of

webpages and SEO and SMO. The second strategy for distinctiveness revolves around

content within the webpages on a particular website. This is not about unique and

compelling content online, as much as a methodical analysis to ensure that duplicated

content is not re-used on multiple pages within a website and that each page is focused on

a single topic or focus (Bradley, 2015, p. 90; Odden, 2012, p. 133). With the webpage

114

composed of succinct and distinctive content, the instruction guides and how-to manuals then present strategies for SEO and SMO in structuring the content within the webpage.

### *Hierarchical Structure Tags*

HTML standards provide a set of heading tags for structural and hierarchical arrangement of text, `<h1>,` most important, through `<h6>`, least important.[40] Search engines check the content of the `<h1>` tag and subsequent children tags for relevancy (Bradley, 2015, p. 76). The instruction manuals and how-to guides consistently emphasize the importance and use of headings, especially the `<h1>` tag for search engine and social media optimization (Bradley, 2015; Jones, 2013; Ledford, 2009; Lincoln, 2009; Michael & Salter, 2008; Moran & Hunt, 2015; Redish, 2014; Shenoy & Prabhu, 2016; Shreves & Krasniak, 2015). The most common SEO and SMO strategy for headings is to use, format and stack them correctly (e.g., `<h1>` before `<h2>`), and to ensure they include important keywords for searchers. The length of the headings may also be important for optimization strategies, and eight words is the recommended maximum length for headings (Jones, 2013, p. 160). The strategy for using headings on webpages is not genre or content specific. The guides emphasize that the use of headings contributes to the user experience and the shape and quality of the content, as well as the SEO and SMO optimization.

In addition to the heading tags, strategies are also given to use the emphasis tags available in HTML (`<strong>` and `<em>`; formerly `<b>` and `<i>`) to especially

---

[40] https://www.w3.org/WAI/tutorials/page-structure/headings/

highlight keywords and signal to the search engine bots that these are significant and more important than other words in the `<body>` (Lincoln, 2009; Shenoy & Prabhu, 2016). The advantage of using both the hierarchical headings and emphasis tags is that keyword phrases may be singled out and encapsulated within the tags to queue the bots without significant parsing required. As search engines have advanced, the reliance on these tags may be reduced, yet the advice remains consistent to call out keywords in the `<body>` with these tags.

The headings and emphasis tags are present in prior media, especially government and technical documents (Gitelman, 2014). It is not surprising that because of the origin of the internet that these structural format elements would be important. Because of the reliance on SEO to access information, the relevance of these elements moves across genres of content and can influence the structure of communication in various formats, perhaps also creating a hierarchical structure to content when the rationale for such doesn't exist. On the other hand, headings make text much more browsable and easy to skim. Regardless of genre, internet users are skimmers first and then readers (Redish, 2014). How this practice is or is not extended for the content for newspaper articles and political campaigns will be examined in the next two chapters.

### Keywords in Context

Search engines process the entire content of the webpage and look for keywords in order to rank content in search results. Keywords usually consist of short phrases (two to three words) that are matched with the user search terms for search engines (Michael & Salter, 2008, p. 126). Keywords that consist of four or more words are usually very

specific and termed "long tail keywords." Even with the decline in the significance of the

`<meta name="keywords">` tag, keywords remain a primary consideration for SEO and

SMO and placement is woven throughout the content in the `<body>`.

> *Thanks to incredible leaps in processing technology and many years of crunching linguistics data in every language, Google's machines can now understand the content of your site and will either reward your website if deemed more relevant than other competing websites or penalize your website's pages and/or domain when they detect overoptimization*
> (Bradley, 2015, p. 60).

In order to support this strategy of identifying keywords, many of the guides provide

significant instructions on keyword research and are clear that keyword assignment is a

skill and both research and assessment techniques are needed to ensure proper keywords

in the meta tag or elsewhere in the page content. "Keyword research is the first step in

SEO and a fundamental best practice" (Kelsey, 2016, p. 43). Extending on the advice for

audience analysis, recommendations are also given to use different types of key words to

try to anticipate how different users may search for content (Moran & Hunt, 2015, p. 55).

Because keyword research is so important, most of the manuals spend significant time

including instructions on the tools from their time of their publication. Keyword research

tools can expose popularity of terms in specific regions, alternative words for terms, and

less used terms that could result in higher ranking because of less competition (Shenoy &

Prabhu, 2016). In *Optimize How to Attract and Engage More Customers by Integrating*

*SEO, Social Media, and Content Marketing,* the author also warns not to get too lost in

popularity and keyword research and risk losing track of relevancy for your audience

(Odden, 2012, p. 76).

Once keywords are identified, the guides provide additional strategies for placement of keywords within the `<body>` of the webpage. Consistently, the how-to guides and instruction manuals suggest that keywords need to be in the first few paragraphs and near the top of the `<body>`. Many of these sources report on "studies" that show that the first paragraph is most important for keywords. (Shenoy & Prabhu, 2016, p. 127). Webpage content "organized like a newspaper article (important words at the top, somewhat repeated throughout, and reinforced at the end) are sometimes said to have an advantage." However, these pages can also be flagged for using keyword stuffing and Black Hat SEO techniques (Moran & Hunt, 2015, p. 71).

This caution against keyword stuffing takes a concrete form in many of the manuals with specific criteria offered on density:

- Five to six times in the `<body>` or per 250 words (Michael & Salter, 2008, p. 75)
- No more than four to six times per 350 words in the `<body>` (Jones, 2013, p. 93)
- One to two percent of the `<body>` (Bradley, 2015, p. 90)
- Seven to ten percent of total words in the `<body>` (Ledford, 2009, p. 118)
- Two to three keywords in the `<body>` (Shenoy & Prabhu, 2016, p. 83)

The advice on keyword density is not consistent and appears to be based more on guesses than data or evidence. In addition to the density recommendations, the order of keywords within a keyword phrase is also noted as significant (e.g., "Hotel in Portland" "Portland hotel" (Moran & Hunt, 2015, p.61). This advice is consistent with traditional cataloging and indexing advice. Rather than being separated terms within a particular field, however, the recommendation is word order and density within a block of text.

*Links and Relationships*

Link building is one of the harder on-page optimizations to measure, especially initially. Links and relationships are relevant for how search engines rank webpages. The more links to your page from quality websites, the better. "Google believes that calculating links and taking into consideration such things as what those links say, along with the quality of the Web sites they come from, is an effective method of determining a Web site's authority" (Jones, 2013, p. 118).  To encourage links to your webpage and relationships with other content on the web, you should also link to other webpages and websites (Bradley, 2015, p. 83). Earlier manuals and guidebooks examined in this project note that there is often some hesitation in including links away from your content. "Include links. It's OK to distract people away from your writing. If you are good, they will come back" (Lutze, 2009, p. 112). Much like the keyword strategies in the text, strategies are also given for limiting the density of links to five or six within the `<body>` (Bradley, 2015, p. 74). The key in the link building strategy is to signal relationships to other sites and have them signal back to your webpage.[41]

In composing <body> text with links, it's important that the links serve a function and "move conversation ahead through links" (Redish, 2014, p. 108) or provide supplemental or contextual information to the text (Moran & Hunt, 2015). In the structure of these links, it is also very important that the text in the code of the link be descriptive of where the link takes you or what it provides (Bradley, 2015; George, 2005; Moran &

---

[41] There is a fairly large market in paying for social media influencer to link to and discuss content, and this originated in ling broker networks where you could pay for website owners and bloggers to link to your webpage (Jones, 2013).

Hunt, 2015). This strategy refers to the human readable text within a link (`<a>`, anchor

tag); for example:

```
<p><a href="https://myawesomewebsite.com">My awesome
website</a></p>
```

The reasons for the accurate human readable text are trust and accessibility but also user

experience and communicating with the user. This strategy is meant to address the

problem in early webpages where the text for a link may look like:

```
 <p><a href="https://myawesomewebsite.com">Click
here</a></p>
```

"That phrase ['click here'] is in no way related to the content…Think of your anchor text

as a chance to showcase the relationship you have with related companies" (Ledford,

2009, p. 104). This strategy within link building, to define the relationship to linked

content, is important for search engine optimization.

Another on-page optimization strategy for building relationships is adding the

social share widgets or buttons to the webpage. Even though this feature within the

structure of the web page does not guarantee linking to your page, it can significantly

increase both search engine and social media optimization by making it easy for others to

link to your content (Odden, 2012). Because these are widgets that are programmed

typically elsewhere or by another entity, how the page is shared and what information it

shares is determined by how that widget reads and re-shares the code on your webpage.

These widgets are usually programmed to pull the `<title>` tag or `<h1>` tag to

accompany the user share of the web page on a social media platform (Odden, 2012).

This social share feature also becomes more important for search engine results as search

engines examine the social media links as part of the search ranking criteria post

Google's 2009 implementation emphasizing social media as sites with authority.

Relationships and context have been an important component in previous

information retrieval models. Whereas previous relationships were defined by an editor

or cataloger, in the context of the Internet, relationships between webpages are defined by

the creator of the document through link building and the readers of the document

through social media sharing and linking on other webpages. This usage of the reader

identifying the relationships is like the vision model of the *Memex*, except that instead of

being defined by the relationships on an individual researcher, the relationships are

defined by the reading public as a whole. The substitution of popularity for authority in a

gatekeeping function comes from this practice of linking and relationship building.

**Linked Data and Semantic Markup**

The strategy of using advanced techniques such as linked data tags and semantic

markup for SEO and SMO were surprisingly absent from the manuals and guide books.

Because schema.org data was originally intended to help search engines, it is interesting

that it does not appear in the other manuals. The exception was *Social Media*

*Optimization for Dummies* by Shreves and Krasniak. This could be in part because it

requires more advanced technical knowledge than many of the manuals geared toward

beginners. Schema.org[42] tags include standards for labeling a large breadth of content,

---

[42] https://schema.org/docs/schemas.html

including: news articles, recipes, musical records, films, products, events, organization, people, and more.

*Social Media Optimization for Dummies* recommends the use of linked data from schema.org as microdata within the `<body>` of the HTML page, which describes elements within the webpage and Open Graph and Twitter card tags within the `<head>` of the HTML page, which describe the webpage as a whole for both search engine and social media optimization (Shreves & Krasniak, 2015, p. 123). These data elements were created in order to help search engines but can also aid social media optimization in Facebook and other tools (Shreves & Krasniak, 2015, p. 122). The key with using strategies such as linked data is to give the machines more information in order to categorize and rank the webpage correctly. These properties are closer to the subject heading and authorities from earlier models in defining controlled terms to be used. Figure 5.4 illustrates how microdata with schema.org standards may be encoded in an HTML page and includes elements that one might expect to find associated with a product listing.

```
…
  <body>
    <div itemscope itemtype="https://schema.org/Product">
    <span itemprop="image" src="hamilton-trent.jpg" />
    <h1 itemprop="name"><span itemprop="brand">Hamilton</span>
      <span itemprop="model">Trent</span></h1>
    <p itemprop="price">$225</p>
    </div>
  </body>
…
```

**Figure 5.4.** Schema.org example for a webpage with product information encoded in schema.org highlighted in purple text adapted from (Shreves & Krasniak, 2015, p. 122).

In addition to the microdata schema.org recommendation, Shreves and Krasniak recommend focusing a strategy for social media optimization using the page-level semantic markup standards with Open Graph tags to include in optimization strategies. The Open Graph protocol is used on Facebook, Twitter, Pinterest, and LinkedIn (Shreves & Krasniak, 2015, p. 123). Within the Open Graph protocol, four tags must be used at a minimum: title, type, image (URL to the image that accompanies content), and URL to the permanent ID of the object. Twitter also created its own standard in the form of Twitter cards, which are also placed in `<head>` section of the HTML page.  Figure 5.5 shows the recommended Twitter tags and the structural considerations for each tag, primarily dealing with length and size requirements to fit the Twitter design format.

- ✔ `twitter:card`: This tag holds the card type.
- ✔ `twitter:title`: Use this tag to contain a title for the page content. Keep the title shorter than 70 characters.
- ✔ `twitter:url`: This tag holds the URL the visitor will be taken to if he clicks the tweet.
- ✔ `twitter:description`: This tag holds a brief description of the page. Keep the description shorter than 200 characters.
- ✔ `twitter:image:src`: This tag holds the URL of the image you want to associate with the page. The image should be at least 280 × 150 pixels.

**Figure 5.5.** Minimum recommended Twitter card tags (Shreves & Krasniak, 2015, p. 127).

With many different ways to structurally call out the title, for example, on an HTML page, it is unclear whether search engines and social media platforms may be looking for uniformity and a true source of a title or content that is adapted for the

audience of the platform. Figure 5.6 shows the layering of titles using traditional HTML tags and linked data tags.

The advice for avoiding duplicative content is not applicable in this scenario. However, why so many titles are needed for these various services is not based on the HTML document and results from specific search engine and social media platform implementations and design parameters of that application such as title length or image format considerations to display within that application. In some ways, this can be akin to a title that is shortened on the spine of a book but has an extended proper title on the title page of the book. The difference here is that there are many spines for presenting the title of the webpage.

```html
<!doctype html>
<html>
  <head>
    <title>My Awesome Headline</title>
    <meta property="og:title" content="My Awesome Headline" >
    <meta name="twitter:title" content="My Awesome Headline">
    ...
 </head>
 <body>
    <div itemscope itemtype="https://schema.org/NewsArticle">
      <h1 itemprop="headline">My Awesome Headline</h1>
      ...
    </div>
  </body>
</html>
```

**Figure 5.6.** A layering of different structured and coded title tags in HTML for a supposed "My Awesome Headline."

**Summary**

Many of the SEO and SMO strategies remained static over the 12 years of the published manuals, despite multiple changes to the algorithms over time. The biggest exception is in the use of the `<meta name="keywords">` tag, where its misuse resulted in its disuse and directly affected how the manuals and guidebooks regarded its importance. The advancement of the search engines and social media platforms also eliminated the need for adding terms to the `<head>` or `<body>` tags for common misspellings and alternate stem endings of words. That keywords remain an important part of the search engine optimization outside of an extracted tag and within the text itself warrants further investigation. That the various titles can be extracted and defined for a variety of platforms presents a use of the title that existed within older media, such as the spine of books, and is extended to multiple manifestations of media where that title could appear to be helpful to the reader. As links and relationships are one of the defining characteristics of new media, the power of the gatekeeper makes the significant transition to the social media and search engine platforms based on community assessments. In the following two chapters, these strategies will be examined within the context of newspaper articles and political campaign webpages in order to see how and if these strategies were employed.

CHAPTER VI

NEWS STORIES USE OF SEO AND SMO STRATEGIES IN THE LA TIMES

This chapter analyzes the webpage structure from *Los Angeles Times*' articles in relation to the strategies for SEO and SMO recommended by the how-to guides and instruction manuals from the previous chapter. The webpages reviewed were published from 2000 to 2018 and are limited to this time period based on availability of archived webpages in the Internet Archive's Wayback Machine and also corresponds closely with the publication years of the manuals. During the publication years of 2003-2004, the *Los Angeles Times* required subscriber accounts to access content. As the web crawler tools used for harvesting from the Internet Archives cannot get past this block, no webpages were examined for 2003 and 2004. This examination of webpages also assumes that a content management system was used to generate the online content for the webpages and structure based on the size of the *Los Angeles Times,* requirements for frequent and quick web publishing, structure of the URL strings, and personal connections to reporters at the *LA Times*. Content management systems provide the basic page structure for publishing webpages and the user / reporter / article author would enter content in predefined boxes rather than code content. Because these content management systems require significant time to migrate to a new system (typically 6 months to two years), one article per year was examined with spot checking webpage structure at article webpages in-between years for verification. All of the articles examined were linked from the home page of the *latimes.com* website for the day of that archived issue.

The webpages in this chapter differ significantly from the webpages in the next chapter. The genre of the content is distinctive between the webpages: news articles vs

126

campaign materials. These webpages are all from a single publisher which is also a large organization, whereas the election pages are each published by their own campaign. As a large organization, it is able to set up resources for brand consistency and tools to help reporters and authors create content. This context both constrains and extends the ability for SEO and SMO strategies within *Los Angeles Times* news articles, as the following sections will reveal. This chapter is divided into three sections: 1) page structure analysis, 2) metadata analysis, and 3) relationships to other content on the web.

## Page Structure

In the structure of an HTML page, there is flexibility in the tags used, order of tags, application of tags, and types of content embedded (i.e., images, videos, audio), in addition to text. The news article webpages were examined for order, basic content elements, and structure of tags used against the recommendations from the SEO and SMO manuals. Because well-formed HTML and use of proper tagging has long been a strategy to support search engines, the pages were also examined for basic HTML compliance and accessibility.

Although HTML provides multiple levels of headings, the archived news articles did not implement multiple headings within the context of the article.  In the print versions, news articles rarely have a subheading, so this was not irregular. An `<h2>` tag was used in the webpage for ten of the 17 webpages examined. In all cases where an `<h2>` was used, an `<h1>` was used for the article title on the webpage.  Three of the webpages used the `<h2>` for a subtitle for the article, which is suspect but acceptable use since the subtitle does nest under the title and the text of the article nests under the

subtitle in some ways. In the other seven uses of the <h2> heading tag, it was applied

incorrectly: three times to call out a related article and four times for the section of the

newspaper. This use is curious because the section of the newspaper would be a proper

<h2> heading following an <h1> for the *Los Angeles Times*, but it is a backwards

implementation to be the <h2> of the <h1> for an article title. Because of the

inconsistent and miscoding of the heading tags, only the <h1> was analyzed for keyword

optimization in the next section. The misuse of the <h2> tag would be recognized by

search engines and could affect rankings.

One of the key content components for a social media optimized webpage is to

include an image or other visual media within the structure of the webpage (Bradley,

2015; Kelsey, 2016; Rowles, 2018; Shreves & Krasniak, 2015). Starting with the articles

examined in 2008, every article had an associated image or video. Unfortunately, the

archived webpages were unable to archive the images for many of the earlier webpages.

Because the technology for archiving webpages is not dissimilar from the crawlers for

webpages and scripts that read into a post format in a social media platform, these images

were likely difficult for search engine and social media platforms to extract as well.

Examples of these missing media formats from the pages include gif files and Adobe

Flash files. The manuals examined in the last chapter started warning against using Flash,

as it interferes with SEO, as early as 2009 (Lincoln, 2009) and limitations in its

accessibility for users (Rowles, 2018). "But there is a reason that Flash-designed websites

by Adobe ultimately failed. It isn't because they weren't effective. People loved

interacting with Flash websites. It wasn't because they weren't beautiful. Flash websites

failed because Google could not read them" (Bradley, 2015, p. 60). Flash objects were

identified in the *Los Angeles Times* webpages examined until 2009. Left in the archived

HTML code of these pages is a note to download the Adobe Flash player and reference to

a style id, which contains a script to call the appropriate Flash object:

```
<div id="divWNHeadline"><div id="help" style="text-
align:center;font-family:Arial, Verdana;font-
size:12px;color:#000000;font-weight:bold;background-
color:#999999;width:500px;height:25px;padding:12px;"><
div id="top"><a style="color:#333333;text-
decoration:none;"href="http://www.macromedia.com/go/ge
tflash/" target="_blank">You need to download the
latest version of flash player to use this
player</a></div><br><div id="bottom" style="vertical-
align:baseline"><a style="color:#333333;text-
decoration:none;"
href="http://www.latimes.com/news/local/undefined"
target="_blank">Need Help?</a></div> - (la08).
```

The discontinued use of Flash in the webpages corresponds with the timeframe when

Google and SEO strategies warned against using it.

With the addition of media in the structure of the webpage, accessibility and the

ability for machines to read more information about the objects is necessary for SEO and

SMO (Ledford, 2009; Rowles, 2018).  The image tag, `<img>`, has an attribute, `alt,`

which became required for compliance with HTML 4.01 released in 1999 (W3C Schools,

n.d.). The purpose of the `alt` attribute is to provide alternative text, description or

specification of function, and is essential for sight disabilities.  Even though images were

standardized as part of the format in 2008, the webpages do not employ a consistent use

of the accessibility attribute, `alt`, until 2012. Within the webpages examined, it doesn't

appear until 2013 with all the primary images associated with the article:

```
<img src="./bWhite House OKd spying on allies, U.S.
intelligence officials say - latimes.com_files/la-afp-
getty-u-s--embassy-at-focus-of-nsa-germany-20131028"
alt="U.S. Embassy in Berlin" border="0" width="600"
height="392" title="U.S. Embassy in Berlin"> - (la13).
```

Although, the sample is such that this doesn't mean that it was never applied to the main image prior to 2013. It illustrates that it was also not required for publication, however. Although the image tag is allowed with an empty value in the attribute, `alt=""`, those cases should be limited to design images only and preferably coded within an associated stylesheet instead of the HTML page (W3C, n.d.). For much of the time that the *Los Angeles Times* published webpages, they did not require complying with HTML 4.01 for images and media.

Most interestingly, any images in pages prior to 2012 which were part of the website frame and code would have been supplied through the content management system, do have an `alt` attribute, including functional images and logos for the site. For example, the image of the button next to search, in Figure 6.1, labeled "Go" is coded as:

```
<img src="./An unsettling portrait of
  &#39;America&#39;s Sheriff&#39; - Los Angeles
  Times_files/search-button-off.gif" alt="Go"
  width="62" height="19" border="0"
  onmouseover="this.src=&#39;/images/standard/search-
  button-on.gif&#39;;"
  onmouseout="this.src=&#39;/images/standard/search-
  button-off.gif&#39;;"> - (la06).
```
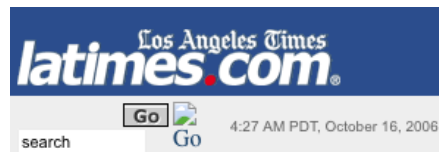


**Figure 6.1.** Screenshot of archived webpage published in 2006 with a "Go" search button.

Of the most striking changes in the page structure, for the news articles, is with

the most recent article examined from 2018 where the entire article text and metadata for

the article, author, and the *Los Angeles Times* as an organization is added to the `<head>`

in microdata format with schema.org tags.

```
<script data-schema="NewsArticle"
type="application/ld+json">
    {
      "@context": "http://schema.org",
      "@type": "NewsArticle",
      "mainEntityOfPage": {
        "@type": "WebPage",
        "@id": "http://www.latimes.com/nation/la-
na-trump-cruz-texas-20181022-story.html"
      },
      "headline": "No more 'Lyin' Ted' — Trump
heading to Houston to support Texas senator - Los
Angeles Times",
      "url": " http://www.latimes.com/nation/la-na-
trump-cruz-texas-20181022-story.html",
      "thumbnailUrl": "",
…
    "articleSection": "",
      "dateCreated": "2018-10-22T19:02:11Z",
      "datePublished": "2018-10-22T19:02:11Z",
      "dateModified": "2018-10-22T20:00:19.167Z",
      "articleBody": "When President Trump takes the
stage at Houston&#8217;s Toyota Center on Monday
night, it will be to deliver a message unthinkable two
years ago to many of his most ardent fans: Vote for
Ted Cruz.That&#8217;s the same Texas senator Trump
once dismissed as… - (la18).
```

Although only one of the manuals recommended semantic coding, its recommendation

was to code around the text as it appeared in the `<body>`. By having the full article text

and metadata duplicated in the `<head>`, the fully machine-readable version of the article

is available for both SEO and SMO without any of the advertisements, related links or

other content such as navigation menus. The remaining human-readable content related to

the article in the `<body>` is duplicated in the microdata and yet is less rich without

providing some of the elements about the author and publisher as organization that are

available in the microdata. As this full implementation of semantic web features is implemented in 2018, it will be interesting to examine future uses and effects of this strategy.

## Basic Metadata and Keywords

As outlined in the strategy manuals, the metadata for a webpage is important for both SEO and SMO. In this section, metadata in both the `<head>` and `<body>` tags will be analyzed for the news article webpages from the *Los Angeles Times*. Metadata examined includes keywords, titles, descriptions, and any other coding that is tagged with data about the content on the webpage. Based on the SEO and SMO strategies from the manuals in the previous chapter, important keywords for the content are expected to be found in the URL, `<title>`, the `<meta name="keywords">` tag (prior to 2009), `<meta name="description">`, headings tags, emphasis tags, and first paragraph text. These elements are examined as they appeared with or without keywords in the news articles examined.

### *URLs*

Many of the early URLs for the news articles appear to be generated as part of the content management system and not a designed or purposeful URL. It wasn't until the *Los Angeles Times* started using more designed URLs that keywords appear in the URL strings, although words and phrases associated with the article may be in the URL. An example of this word but not necessarily keyword usage can be seen in the "Scarce Funds Imperil Bush Health Goals" article from 2001:

http://www.latimes.com:80/news/politics/la-082401tommy.story?coll=la-headlines-politics.[43] – (la01).

In this example, "tommy" in the URL doesn't appear in the headline; however, "Tommy G. Thompson" is in the first sentence and the caption on an image of him as the feature image of the story.



**Figure 6.2.** Screenshot of 2001 webpage with "Tommy" appearing in first sentence of article and photo caption.

Examining the news article from 2009, the structure of the page, layout, and design underwent significant changes.[44] See Figure 6.3. With these changes, more keywords were also added to the URL strings:

---

[43] The ":80" at the end of the primary domain in this example is for port 80, which is the default port for publicly accessible webservers. See: https://www.w3.org/Daemon/User/Installation/PrivilegedPorts.html

[44] The 2009 design change was not the only design change between 2001 and 2009; however, more than the others, it exhibited significant structural changes, such as with the ordering and coding of menus, placement of ads, and URL design.

http://www.latimes.com:80/news/nationworld/world/la-fg-obama-afghan27-2009oct27,0,7820767.story – (la09).

http://latimesblogs.latimes.com/technology/2011/10/obama-2012-campaign-starts-a-tumblog-tumblr.html - (la11).

The placement of keywords in the URLs continues through the latest versions of the webpages examined limited to two to four keywords in the URL string.



**Figure 6.3.** Screenshot of 2011 article where URL duplicates wording in article title (`<h1>`), "Obama 2012 campaign heads to Tumblr."

Although the 2011 URL is a duplicate of the `<h1>` article title, that pattern is not consistent throughout future pages and is not a transition to practice. This may signal an effort to not have too much duplicate text in the HTML and avoid keyword stuffing practices.

*Titles*

Because titles can exist in several places in HTML, in looking across the webpages for title analysis, four primary tags were examined across the `<head>` and `<body>`: `<title>`, `<meta property="og:title"…>`, `<meta name="twitter:title"…>`, and `<h1>`. Pages were also scanned for additional titles in other tags. This was especially important for the earlier webpages, in which the article titles were not properly coded in `<h1>` tags in the `<body>`. In examining the news article webpages, the titles were duplicated across tags. Similarities across the titles is not unexpected with a content management system, where a single entry for an author may be programmed to fill in several spaces within the HTML code. The article from 2010 was an outlier with using different titles. In this webpage, the `<title>` tag contains, "With midterm campaign in home stretch, conservatives struggle to unify - latimes.com," while the `<h1>` tag contains, "Conservatives struggle to unify for voter outreach." Between the two titles, the main phrase of the `<title>` tag starts the text of the `<h1>` tag, "conservatives struggle to unify." It is worth noting that this outlier with different titles, however, doesn't appear to be a widespread practice. In order to verify the anomaly, several other articles from that day's issue were examined, and their HTML used duplicate titles.

The tags associated with the titles evolved over time. Early renditions (2000-2002) had the headline title in the `<body>` coded incorrectly as `<span class="cHeadline1">` instead of `<h1>`. The function of the `<span>` tag provides style and formatting instructions for the text but does not declare it as a headline. The class of "cHeadline1" could be any other text and providing "headline" as part of the tag

clues in a human reader. However, a crawler wouldn't know to parse it for that

information. The 2002 article provides an illustration of the titles used in early 2000's

news articles for the *Los Angeles Times*:

```
<head>…
<title> Heat's on Senate After Campaign Reform
    Victory</title>
…</head>
<body>…
<span class="cHeadline1">Heat's on Senate After
    Campaign Reform Victory</span>
…</body> - (la02).
```

Webpages of the news articles from 2005 to 2010 fixed the `<h1>` issue and properly

applied the tags to titles/headlines and then followed the pattern of the earlier webpages

with two titles per page; one in the `<head><title>` and one in the `<body>`.

For webpages published between 2011 and 2017, Open Graph title tags are added

to the HTML code ("&#39;" is the HTML code for the single quotation symbol: ' ) :

```
<head>…
<meta name="fb_title" content="Biden on Romney Jeeps-
    to-China claim: &#39;Have they no shame?&#39;">
<meta property="og:title" content="Biden on Romney
    Jeeps-to-China claim: &#39;Have they no
    shame?&#39;">…
<title>Biden on Romney Jeeps-to-China claim: 'Have
    they no shame?' – latimes.com</title>
…</head>
<body>…
<h1>Biden on Romney Jeeps-to-China claim: 'Have they
    no shame?'</h1>
…</body> - (la12).
```

In this example from 2012, there is also an additional title "fb_title," which is not used by

Facebook but rather for the content management system to run a script in order for the

*Los Angeles Times* to post the article to Facebook. Open Graph tags are used by

Facebook when another user posts the HTML page to Facebook. As noted in the previous

chapter, the only manual that called out using Open Graph meta tags as a strategy was

*Social Media Optimization for Dummies* (2015). The inclusion of Open Graph tags in

2011 follows the 2010 Social Signals update that Google applied to increase relevancy

based on HTML pages shared in social media.

The webpage examined from 2018 article includes the patterns used in 2017 with

an additional schema.org title coded as microdata within the `<head>` of the HTML

document:

```
"headline": "No more 'Lyin' Ted' — Trump heading to
Houston to support Texas senator - Los Angeles
Times", - (la18).
```

The 2013 Hummingbird release of Google's search algorithm incorporated these

knowledge graph and semantic web tags into its relevancy calculations. This first

appearance of full schema.org tags in the *Los Angeles Times* is five years after that

change. It is somewhat surprising that this change took so long, and it will be interesting

to see how long this strategy persists and if it changes in the future.

One of the challenges of disaggregated web content is attribution and context of

the webpage. In the news articles, this issue of the context and larger entity was

addressed within the `<title>` tag for the news articles published after 2004. A suffix

was appended to the titles "- X":

```
<title> Bush Names Bernanke to Replace Greenspan as
Fed Chief – Los Angeles Times </title>. – (la05).
```

The suffix used to define the entity of the *Los Angeles Times* changed over the years with

the originally applied full suffix (2005-2008) "Los Angeles Times" reapplied in 2018.

Figure 6.4 illustrates the different suffixes used over time in the webpages examined.
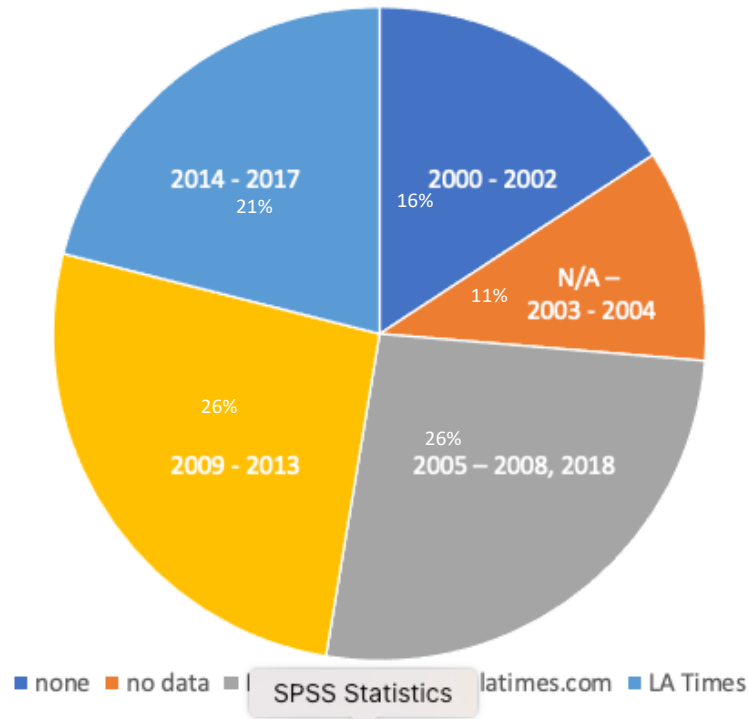
**Figure 6.4.** Suffixes applied in `<title>` tag for the *Los Angeles Times*.

The issue of a suffix title application for context and how it is used was not addressed in the SEO and SMO manuals as a strategy. It could be that the effect is simply irrelevant for search and social media and serves a role that is more for the human readable context where the `<title>` is produced in the tab or window bar of a browser, and placing it at the end of the title doesn't interfere with relevancy based on the title in the `<title>` tag.

*Descriptions*

The `<meta name="description">` tag for all articles was a duplication of the first paragraph of the body text. There was a slight exception to this from the 2017 article, where the phrasing is similar, yet slightly different:

```
<meta name="Description" content="The ACLU asks a federal
    court to reenter the case of a pregnant 17 year old
    immigrant held in federal detention who is seeking an
    abortion."> - (la17).
```

The ACLU asked a federal appeals court Sunday night to reenter the case of a 17-year-old pregnant immigrant in detention whose request for an abortion has been blocked by federal officials. – (la17).

The length of the descriptions was also not capped at the recommended 150-160 characters. The longest `<meta name="description">` tag in the webpages examined was 321 characters and ended with a "…" mid-sentence. None of the recommendations from the manuals or the W3C were followed in the application of this tag. Because, as some of the manual authors asserted, it is mainly used for a human-readable selection on a search engine results page, this may be sufficiently like the typical practice in the genre of news articles that no changes were made for SEO and SMO.

### Meta Keywords

With the warnings against keyword stuffing in the early SEO manuals (George, 2005) and changes in Keyword Trust to Google's search algorithms in 2009, which devalued the keywords in the `<meta name="keywords">` tag, it is somewhat surprising that the `<meta name="keywords">` tag endures in the webpages for the news articles. The article from 2006 has the largest number of key phrases with eleven, including many duplicated words in key phrases. The `<meta name="keywords">` tag disappears in 2010 and returns in some form with up to five key phrases in the most recent webpages examined. See Table 6.1 for the varied data in the `<meta name="keywords">` tag.

**Table 6.1.** `<meta name="keywords">` tag in the news articles examined from the *Los Angeles Times.*

| ID | year | `<meta name="keywords">` |
|---|---|---|
| la00 | 2000 | Los Angeles Times Voters Guide 2000 |
| la01 | 2001 | tommy thompson, budget |
| la02 | 2002 | campaign finance reform, finance |
| la05 | 2005 | null |
| la06 | 2006 | foreign policy, armed forces, foreign policy iraq armed forces security legislat, the conflict in iraq, legislation, lead story, security, iraq, advisory committees, bush  george w |
| la07 | 2007 | news |
| la08 | 2008 | absentee voting, voting, trends, news, california, elections 2008 |
| la09 | 2009 | world, news, afghanistan, asia, armed forces, military deployment, united states |
| la10 | 2010 | null |
| la11 | 2011 | null |
| la12 | 2012 | null |
| la13 | 2013 | not_live_web,world,news |
| la14 | 2014 | Iran, nuclear, negotiation, weapon, U.S., centrifuge, uranium |
| la15 | 2015 | Humboldt County, Hoopa Valley reservation, law enforcement, tribal police |
| la16 | 2016 | null |
| la17 | 2017 | Jane Doe, abortion, immigrant, Trump, E. Scott Lloyd |
| la18 | 2018 | Trump, Ted Cruz, Beto O&#39;Rourke, Texas, Houston |

The application of `<meta name="keywords">` tags corresponds with some of the recommendation in SEO and changes in Google's algorithms, such as the reduction of keyword stuffing like the 2006 example. However, there isn't a discernable pattern that connects the application of the `<meta name="keywords">` tag with SEO and SMO practices.

### *Keywords in the `<body>`*

The SEO manuals suggest three locations for keywords within the `<body>` text, headings, emphasis tags, and the first paragraph. The news article webpages do not make

use of headings, as described previously. They also do not use emphasized text within the copy in bold or italics tags.  The first paragraphs for all the articles are composed of key words and phrases that both signal to the reader and search engines what the article will be about. Because the keywords, in this context, are integrated into the text, the analysis of which words are keywords post publication is more like a traditional cataloging analysis and is subject to the bias and judgment of the evaluator and/or tools used. To identify the keywords in each first paragraph of the article, traditional cataloging practices were used in Figure 6.5.[45]

> The **White House** and State Department signed off on **surveillance** targeting **phone conversations** of friendly **foreign leaders**, current and former U.S. intelligence officials said Monday, pushing back against assertions that **President Obama** and his aides were unaware of the high-level **eavesdropping**. – (la13).

**Figure 6.5.** First paragraph of from 2013 article, "White House OKd spying on allies, U.S. intelligence officials say" with example keywords highlighted in bold.

This practice of good keyword and key phrase placement in the first paragraph may be attributed to the genre of the news article as one of the SEO manual authors (Moran & Hunt, 2015) alluded to. It may be that the lack of keywords in other places, other than the title, is to avoid the keyword stuffing label and Black Hat practices.

---

[45] I experimented with topic modeling and other text analysis tools to identify keywords in the context of the main content of the articles; however, I was unable to confirm any more validity that my assessment as a former cataloger using traditional techniques.

**Relationships with Other Web Content and Social Media**

Relationships and links are central factors in the structure of new media online and also relevant for SEO and SMO techniques. The news articles webpages for outbound links and link building to create relationships among other sites were surprisingly sparse. See Table 6.2 for categories of links on the webpages. Consistently, the webpages link to external websites for advertisements. For the purposes of SEO, these links are somewhat irrelevant, as part of the analysis of the search engines involves topical relevancy of the external links. For example, advertisements for a dish soap likely don't have any relevancy to the content of the news article on the state of national politics. Starting in 2009, the pages also included links to newspapers owned by the same parent company as the *Los Angeles Times*: *Baltimore Sun*, *Chicago Tribune*, etc. These links are part of the page frame design structure. The only article that included a link to an external website that was neither paid for by advertisers or part of the same company was the 2011 article, which links to Obama's Tumblr page. This is also interesting, because in this case, Obama's Tumblr page is the subject of the article. The strategy to build links to other websites in order to establish relationships and credibility is not illustrated in the tactics used by the *Los Angeles Times*.

In terms of providing links to external websites for SMO, the webpages include links to the social media accounts (Facebook and Twitter) for the article author(s), as well as the accounts for the *Los Angeles Times* starting with the webpage examined from 2011. In 2010, Google applied the Social Signals update to increase relevancy and

**Table 6.2.** Prescence of links from news article webpages by category off of the webpage; * outbound links to an external website.

| ID | Advertisers* | Site structure and navigation links | Related *LA Times* Articles, Photo Gallery, Media | Topics | Newspapers from same parent company* | External Websites* | Social Media accounts for author(s)* | Social Media accounts for *LA Times* |
|---|---|---|---|---|---|---|---|---|
| la00 | X | | | | | | | |
| la01 | X | X | X | | | | | |
| la02 | X | X | X | | | | | |
| la05 | X | X | X | | | | | |
| la06 | X | X | X | | | | | |
| la07 | X | X | X | | | | | |
| la08 | X | X | X | | | | | |
| la09 | X | X | X | | X | | | |
| la10 | X | X | X | X | X | | | |
| la11 | X | X | X | X | X | X | X | X |
| la12 | X | X | X | X | X | | X | X |
| la13 | X | X | X | X | X | | X | X |
| la14 | X | X | X | X | X | | X | X |
| la15 | X | X | X | X | X | | X | X |
| la16 | X | X | X | X | X | | X | X |
| la17 | X | X | X | X | X | | X | X |
| la18 | X | X | X | X | X | | X | X |

importance of webpages shared in social media. The strategy of applying these external links for social media accounts, supports both SEO and SMO and increases the access to the content.

The news article webpages also include some expected off-page and in-site links, such as to menu and navigation categories, as well as related articles on the site as part of a feature or separate element on the webpage. In addition, with the page published in 2010, the article content includes links to terms that take the reader to a summary topic directory page hosted by the *LA Times*:

```
But the push to get the nation's conservative voters
to the polls is fractured and untested, with some <a
class="taxInlineTagLink" id="ORCIG000068" title="Tea
Party Movement" href="
http://www.latimes.com/topic/politics/tea-party-
movement-ORCIG000068.topic">"tea party"</a> activists
refusing to cooperate with more mainstream… - (la10).
```

The effect of this addition of a directory of terms, which are used frequently in *LA Times*

articles, both provides the user with information and keeps them on the *LA Times*

website.

As the links become more prevalent within the text of an article, links to related

articles were also added directly to phrases within the text and not relegated solely to a

related articles box or associated feature on the page:

```
That's the same Texas senator Trump once dismissed as
"Lyin' Ted," whose father Trump suggested <a href="
http://www.latimes.com/politics/la-na-trump-cruz-
oswald-20160503-story.html" target="_blank">played a
role in the assassination</a> of President Kennedy… –
(la18)
```

This new structure of including the related articles within the text, in addition to the

topics, demonstrates a maturity in the use of links and the online document for news

articles. Despite these changes, it is interesting that webpages do not take advantage of

link building beyond sites paying for the links, as with advertisements, or are part of the

larger organization's web presence.


## Summary

The articles from the *Los Angeles Times* webpages used several SEO and SMO

strategies and incorporated more strategies as time progressed or changed strategies, such

as removing the large number of keywords/ key phrases in the `<meta`

`name="keywords">` tag. Some of the strategies, such as good titles and keywords in the

first paragraph, are not unique to SEO and SMO and may be more based on the genre of

content for news articles. Likewise, the reuse of the first paragraph in the `<meta`

`name="description">` tag is not particular to an SEO or SMO strategy. Strategies

that were not used included the proper use of hierarchical headings, emphasized text, and

link building. Additionally, it was surprising that the use of the `alt` attribute in the

`<img>` tag and accessibility features were not common until the most recent webpages.

The evolution of linking within the article text, first to locally held terms, and then

to related articles marks an important transition to using the features of HTML and the

web environment that wasn't possible in print. The additions of both links to social media

accounts and metadata in the `<head>` to support social media posting began around 2009

and continued to be a prominent feature of the page structure with more and more tags

added to support social media platform integrations.  In addition, one of the more

advanced strategies was the addition of the semantic web micro data into the `<head>` of

the 2018 article. This fully formatted and tagged machine readable version of the text in

addition to the text in the body is worth further exploration of how long this persists, how

it is used, and what effect it has on access to content.

CHAPTER VII

U.S. SENATE ELECTION POLITICAL CANDIDATE WEB PAGES USE OF SEO

AND SMO STRATEGIES

This chapter analyzes the webpage structure from political candidate campaign

websites in relation to the SEO and SMO strategies recommended by the how-to guides

and instruction manuals from Chapter V. The webpages reviewed were published by U.S.

Senate campaigns from 2002 to 2016 and are limited to this time period based on

availability of archived webpages in the Library of Congress' Elections Web Archive.

The publication dates also correspond closely with the publication years of the manuals

examined in Chapter V. The technology used to harvest for the Elections Web Archive is

the same technology used by the Internet Archive for the *Los Angeles Times* webpages in

the previous chapter. The forty webpages of political websites examined were selected

for particularly close election races based on the election results with five webpages per

each election year. (See appendix B for margin of victory.) Many webpages within the

Elections Web Archive were eliminated due to lack of code and content successfully

harvested. This often occurred when a pop-up blocked the content to the site, the site was

encoded all in Adobe Flash, or the content was limited to an image and less than three

sentences of text. (See appendix B for quality evaluation of archived political webpage

content.) Three candidates had websites evaluated twice between 2002 and 2016 (Mark

Begich: 2008 & 2014; Jon Tester 2006 & 2012; Pat Toomey 2010 & 2016). In each of

these cases, the structure of the website was completely revised and not an updated

version of the former website and was captured in a separate object in the Library of

Congress record. Two of the sites were designed by the same company, Wide Eye Creative (pc14e and pc16a). If any of the other sites were composed by the same company or author, it was not clear from the code, and there were not enough similarities between the sites to assume the same creator.

Within the websites for the political candidate campaigns, the webpages analyzed were topical or issue-based. These types of pages were selected in order to examine how SEO and SMO strategies were implemented in the persuasive industry of political campaigns on topics of interest to the public and that couldn't easily be found with a specific or long-tail search. Campaign topics for these pages included, for example: economics, agriculture, jobs, and women's rights. These search phrases alone in search engines would retrieve a large number of pages and have a high keyword competitive index.[46] Each topic page was available through a direct link on the candidate's home page as a top-level navigation link. Some of the topic pages described multiple topics. Five out of the 40 pages had multiple topics ranging from nine topics per page with Webb 2006 (pc06c) to 29 topics on Tester 2006 (pc06a). The majority of the webpages, however, are devoted to a single topic or issue.

The webpages examined in this chapter differ significantly from the news articles in the previous chapter, not only in format and genre, but also in infrastructure and resources. Each website is produced by its own organization or entity and often may have hired a firm to design and develop the content for the site. As these websites are

---

[46] The keyword competitiveness is directly related to the number of times a word is used across all pages in the search engine index (George, 2005, p. 67).

examined, there is some overlap in firms and content management, which led to similar

SEO and SMO techniques applied on those websites. The first instance of WordPress, a

content management platform with no HTML coding knowledge necessary, in the pages

examined was with Begich 2008 (pc08a).[47] By 2012, WordPress was used with the Yoast

SEO plug-in to automate some SEO features in Tester 2012 (pc12d).[48] Three out of the

five websites for the 2014 election (pc14a; pc14b; pc14e), and four out of five websites

for the 2016 election also used WordPress with the Yoast SEO plug-in (pc16a; pc16b;

pc16d; pc16e).  The importance for making this content available through search engines

and social media platforms is essential to provide access to the content.


## Page Structure & Content

The analysis of political campaign webpages followed the same methodology as

the news article webpages in the previous chapter and concentrated on examining page

structure overall, the headings tags, application of emphasis tags, types of embedded

content (i.e., images, videos, audio), and text in the HTML code. The webpages were also

examined for well-formed HTML, proper tagging for basic HTML compliance, and

---

[47] 11 of the 15 webpages examined in the 2002, 2004, and 2006 campaigns used a table design structure for content on the page. This strategy well organizes content; however, it causes issues for screen readers and mobile versions of the pages. With the onset of tools like WordPress, much of the work for the design layout is baked into the tool, and table layouts are not used as much. Because the table layout doesn't have a known effect on SEO and SMO, this structural feature was not looked into further for this project.

[48] The Yoast SEO plug-in has both free and premium versions, and one of its selling features is that it updates frequently to keep pace with search engine algorithm changes. https://yoast.com/wordpress/plugins/seo/.

accessibility. The analysis was based on the recommendations from the SEO and SMO

manuals in Chapter V.

Overall, the political candidate issue pages made use of hierarchies through

structured headings, styles, and emphasized text and used more of the HTML tags

available than the news articles utilized. The application of the HTML tags was

inconsistent and not always applied according to HTML specifications, however. In the

pages examined, the first instance of an `<h1>` heading tag applied correctly for the page's

content title was found in 2008 candidate campaign webpages (pc08b; pc08c), e.g.:

  `<h1>Growing Rural Oregon</h1>` (pc08c).

Incorrect level heading tags were applied to page titles with an `<h2>` (pc06a; pc08a;

pc10b; pc12e; pc14b; pc14d; pc16d) or `<h3>` (pc06c; pc10a; pc10c) as the coding for the

primary title of the page. The application of HTML headings tags was common, although

not always correctly applied in 22 of the 25 pages examined for campaigns between 2008

and 2016. In 2002 and 2004 candidate campaign pages, the primary page title was coded

a third of the time in emphasize bold `<b>` tags, a third in image tags, and a third in a

styled class. The image tag coding of the title is especially problematic to be machine-

read as a title without a parent heading tag designation. It is especially egregious when

the title is represented by an image without an `alt` class in the `<img>` tag and not

providing machine-readable text for a title (pc02e; pc04d; pc06e). The title of the page is

essentially hidden within the image. (See Figure 7.1).
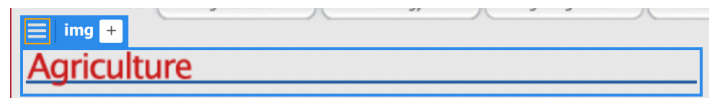


**Figure 7.1.** Page title only visible through image for "Agriculture" (pc06e).

This is likely a result of carelessness or lack of awareness of the impact of an image without an alt tag and no other heading. Whether correctly coded or styled to highlight, the campaign issue pages used HTML coding to set the page title apart from the rest of the page with the exception of one page during the 2016 campaign, which used the `<h1>` tag for a call in the footer instead of a title in the page `<body>`: `<h1>Join <b>TEAM TOOMEY</b></h1>` (pc16c).

Many of the pages also made use of subheadings and coding for subtopics within the issues. Emphasis tags of bold (`<b>` or `<strong>`) and italics (`<i>` or `<em>`) were applied to subheadings through the span of 2002 to 2016 election pages. Consistent with the proper application of the `<h1>` tag for primary headings in 2006 campaigns, the proper cascade of headings starts to appear. However, in the first case identified in the pages examined, the flow cascades appropriately from an `<h2>` to an `<h3>` as a subcategory of the first (pc06a), yet the `<h2>` should have been coded as an `<h1>`. Although the headings tags were not always applied correctly, the use of the subheadings were also prominent within the issue webpages either with appropriate `<hX>` heading tags, emphasis tags, or stylized to set apart from other text on the page. Figure 7.2 illustrates the different types of structures applied around headings tags in the webpages. In many cases where the HTML headings tags were not used, the primary heading / page title was "styled" through a stylesheet class using Cascading Style Sheets (CSS). By applying this type of coding, the page titles can be styled for a specific appearance in font, color, size. The same style coding could also have been applied to the `<h1>` tag through CSS, so it's unclear why a made-up class was used instead. Examples of styled headings were coded as:

- `<span class="header">` - (pc02a)
- `<div class="body_title">` - (pc02d)
- `<p class="headline">` - (pc04b, pc06d)
- `<div class="redheader1">` - (pc08e)
- `<section id="pagetitle">` - (pc12c)
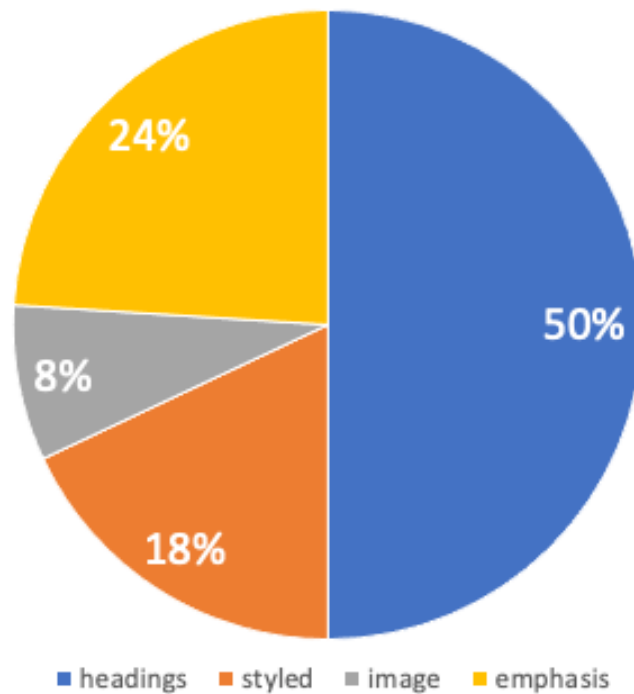


**Figure 7.2.** Application of structured tags in the page `<body>`. N=50; ten of the webpages employed two techniques for hierarchical structure within the page `<body>`.

The real disadvantage of this coding through stylesheets without the HTML headings tags is that the classes do not match the standard and are therefore not readily machine readable or identifiable as titles. One might read "`title`" or "`headline`" in the class attribute as illustrated above, but a machine, such as search engines and social media platforms, wouldn't know to call that class for the title information. The example of "`redheader1`" is a great example of "red" read into the human-readable class. However,

because the color is defined in the CSS, the color could be green, blue, purple, or any other color. There is nothing inherent in the code that makes the "red" true.

Building on the structure and content in HTML that also helps make content more appealing to readers, search engines, and social media platforms, the political candidate issue pages integrated images and media into the main page content. Two of the pages examined also offered alternative text only versions of their webpages, which would be extremely useful for people on slow internet connections (pc02c; pc04b). Similarly to the news article pages, there is lost content from Adobe Flash objects (pc02a; pc08d), as well as Flickr image galleries (pc10a) and linked YouTube videos (pc10b), as external linked content that was originally embedded within the pages. The other use of images within the structure of the pages is as a background image to the page itself. In cases where the background image is purely stylistic for aesthetic purposes, it doesn't necessarily break accessibility standards (see Figure 7.3). A background image can cause additional problems, however, if content is not exposed until the image loads because of contrast or style applications (see Figure 7.4). In instances where information or content is part of the image, it should be described for the vision impaired. Using a background image in this manner breaks accessibility standards (see Figure 7.5). In the background image for Figure 7.5, the image content is part of the message of the page and used to convey a mood to match the message of the page (See: Redish, 2014, p. 279). When an image is applied to convey part of the message or story, an `alt` attribute should be used to comply with the W3 HTML standards (W3C, n.d.). Without the `alt` attribute, the content is not easily machine-readable, cannot be communicated to most search engines or social media platforms, and is hidden from vision impaired users.

**Figure 7.3.** Screenshot of Jon Tester's 2012 campaign website with antiqued textured image background (pc12d).



**Figure 7.4.** Screenshot of Jon Tester's 2012 campaign website before background images load, resulting in some text, logos, and menu options rendering faint and/or invisible (pc12d).
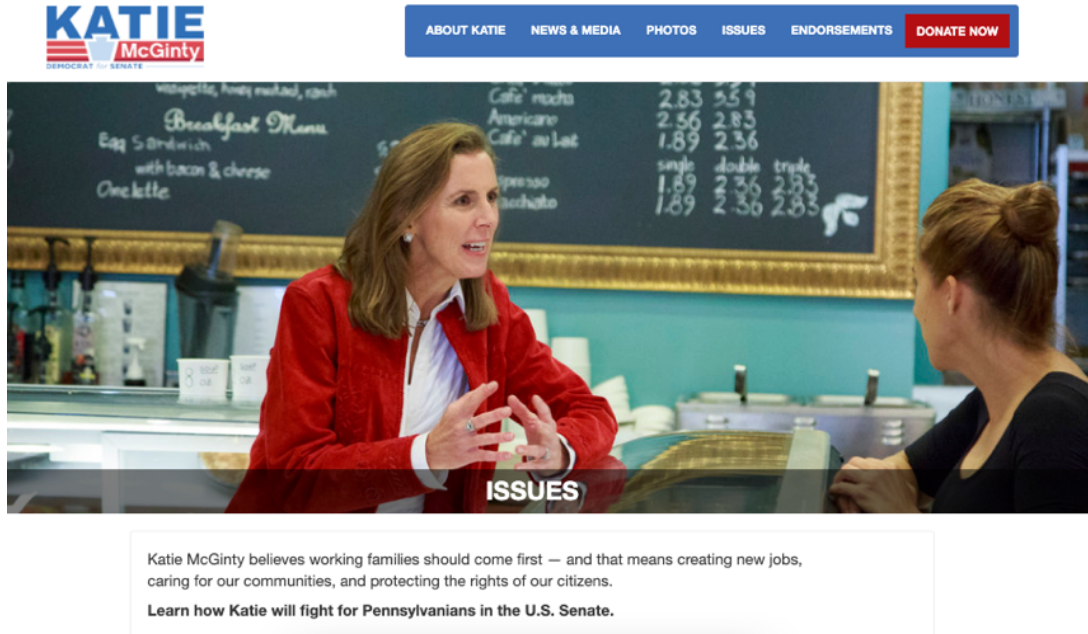
**Figure 7.5.** Screenshot of Katie McGinty's 2016 campaign website with background image of McGinty in a café talking with assumed proprietor or staff (pc16d).

In examining the accessibility of the media and particularly the application of the `alt` class for the `<img>` tag, the usage is sporadic through the 2002 and 2016 campaign issue pages. Interestingly, the `alt` class is used in the earliest pages examined (pc02a; pc02b; pc02c; pc02d) but is missing in some of the latest (pc14d; pc16c). There is not an overwhelming increase in the practice of using an `alt` for an `<img>` tag over time. This is unlike the *Los Angeles Times* news articles, where `alt` became more standard over time. The `alt` attribute has another function in webpages, separate from SEO and sight-impaired accessibility, where the `alt` attribute text may be rendered by a browser while the image is being loaded. This usage may have been very useful in early 2000's webpages and with slower internet connections. The content of the attribute may vary in utility from a one-word description that could describe the image or the location of the image; it's unclear:

```
 <img src="./Social Security _ Buck For
Colorado_files/header_buck.png" alt="head"
id="buck_head"> - (pc10b)
```

to a descriptive alternative text that provides information that the image is carrying:

```
<img src="./Pete Coors for U.S. Senate – On The
Issues_Jobs and the Economy_files/hd_coors_right.gif"
width="151" height="178" alt="American Flag flowing
over a Colorado Mountain Range" border="0"> - (pc04e)
```

Overall, the `alt` attribute was used properly in 50% of the webpages examined. In cases where it was not used correctly, it was either missing (null) on 12 pages, was an empty attribute, `alt=" "`, on 15 pages when the image was used for more than style, or a combination of both. The frequency of the correct usage, after viewing the lack of application in the *Los Angeles Times* was surprising. These sites did not have the infrastructure that the newspaper organization provided, and yet, had more accessible webpages in the early and mid 2000's.

The final important structural element that appeared in a few of the political campaign issue pages examined is the use of multiple languages for the page content. Three of the webpages included Spanish translations of the content (pc04a; pc04c; pc06c) with one that also included a Vietnamese version (pc06c). For the purposes of this project, multiple translations of the same content can cause problems for search engines unless properly coded and indicated to search engines that there are multiple versions for your pages. This is important for two reasons: 1) the page should notify the search engine of different versions of the page in which language, in order to avoid duplicate content flags, and 2) the keywords used in different languages are likely not a direct translation and need to be researched and designed per language version (Ledford, 2009, p. 188,

387). If the tagging and structure is applied correctly, then Google can also treat each version separately and optimize for relevancy within that language or region.[49] Each webpage with multiple languages was examined for the presence of a `<link rel="alternate" hreflang="lang_code" href="`*`url_of_page`*`" />` tag in the page's `<head>` code. None of the pages included this HTML element to let search engines know there were multiple versions of the page. There are two additional ways that sites can alert machine-readable queues that there are different languages available for the content: 1) HTTP headers (used primarily for PDF and attachment files), and 2) sitemaps. It is possible that these sites had sitemaps that were not captured by the web archiving tools and not available. An interesting future analysis could include looking at the pages with multiple languages and how SEO and SMO strategies are applied similarly or differently in the versions, in addition to the indications of the languages for the search engines.

The semantic web and micro data structure that was found in the `<head>` of the 2018 news article examined was not present in any of the political candidate issue webpages. However, the Yoast plug-in, which is frequently used in the sites examined, added features for schema.org and the semantic web with micro "data blocks" in its premium version in 2020.[50]  It will be interesting to examine future campaign issue webpages for the application of schema.org or other semantic web structures. Will this

---

type of microdata become a common way of connecting content through context and standardized references so that the page is not lost from the context of the whole and the authorship or organizational home is clearly defined?

## Basic Metadata & Keywords

The metadata for a webpage, which the how-to manuals and instruction guides suggested as important for both SEO and SMO are examined in this section. Metadata, in both the `<head>` and `<body>`, will be analyzed for the political campaign issue webpages including keywords, titles, descriptions, and any other coding that is tagged with data about the content on the webpage. Based on the SEO and SMO strategies from the manuals in the previous chapter, important keywords for the content are expected to be found in the URL, `<title>`, the `<meta name="keywords">` tag (prior to 2009), `<meta name="description">`, headings tags, emphasis tags, and first paragraph text.

### *URLs*

The political candidate issues webpages, as a whole, took advantage of designed URL strings. This strategy was found in the earliest webpages examined from the 2002 campaign. 31 out of the 40 pages examined had designed URLs, which were human-readable, while the remaining nine had auto-generated URL strings for the issue page:

> *Example Designed URL string:*
> http://www.timjohnsonforsd.com/workinghard/agriculture.php – (pc02b).

> *Example Auto-generated URL string:*
> http://www.johnthune.com/issues.asp?formmode=issue&id=3 – (pc02c).

In the pages examined, the last auto-generated URL was found for a 2010 campaign:

http://www.pattymurray.com/issues?id=0005 –  (pc10e).

Even with the designed human-readable URLs, the application of keywords was minimal and usually limited to the primary theme of the issue. A long-tail keyword or specific keyword may be a more strategic approach for SEO and SMO. For example, "economy" is very broad topic and many webpages on the Internet are vying for a search engine retrieval with that term alone.

http://joesestak.com/Economy.html –  (pc10d).

The combination of the candidate's name and the "agriculture" may argue toward multiple keywords applied. In contrast, a well applied keyword / key phrase designed URL looks like the following two examples:

http://www.bennetforcolorado.com/issues/details/2010-09-building-a-21st-century-economy –  (pc10a).

http://www.lautenbergfornj.com/issues-homeland-security-and-combating-terrorism.php –  (pc08e).

URL strings that had fully developed keyword and key phrases in the URL were limited to five of the examined webpages. Overall, the human-readable strings provide a much better strategy for SEO and SMO than auto-generated URLs. However, it appears that most of the webpages did not extend to applying important keywords in the URL strings. The use of the human-readable strings in the early 2002 campaign pages is evidence that the structure of a designed URL was adopted early on for these types of pages.

*Titles*

Using the same process as with the news articles, four primary tags were examined across the `<head>` and `<body>`: `<title>`, `<meta property="og:title"…>`, `<meta name="twitter:title"…>`, and `<h1>`. Pages were also scanned for additional titles in other tags. Looking for titles in tags other than these tags was especially important, as noted in the previous section, in which the issue titles were not properly coded in `<h1>` tags in the `<body>` and may have been styled as the only signifier that they were the title for the page content.

The contents of the `<title>` tag and relationship and order to the `page` and site title varied among the webpages. Like the *Los Angeles Times* article pages, many of the titles included the site title in the title tag, as well as the page title. The placement and order of the titles differed among the pages. The following examples illustrate the basic structures identified; (see Figure. 7.6 for distribution):

- Site title only:
  - `<title>Mel Martinez for Senate</title>` - (pc04a)
  - `<title>Joe Heck for U.S. Senate</title>` - (pc16e)
- Site title as prefix:
  - `<title>Pete Coors for U.S. Senate – On The Issues/Jobs and the Economy</title>` - (pc04e)
  - `<title>Rick Berg for Senate    » Jobs and the Economy</title>` - (pc12b)
- Site title as suffix:
  - `<title>Growing Rural Oregon – Jeff Merkley for U.S. Senate, Oregon.</title>` - (pc08c)
  - `<title>Issues | Katie McGinty: Democrat for Senate, Pennsylvania</title>` - (pc16d)
- Parent section as <title>:
  - `<title>John Thune :: U.S. Senate [Issues]</title>` - (pc02c)

- o `<title>Michael Bennet for U.S. Senate | Issues</title>` - (pc10a)
- Other title (distinct from phrases in the `<body>`):
  - o `<title>Kelly Ayotte's Record on Student Loans &amp; College Affordability</title>` - (pc16b)
    - ▪ `<h1 class="page-title">Kelly is working to make <strong>college more affordable</strong></h1>`
  - o `<title>Pat Toomey On Iran &amp; ISIS</title>` - (pc16c)
    - ▪ `<h1>Join <b>TEAM TOOMEY</b></h1>`
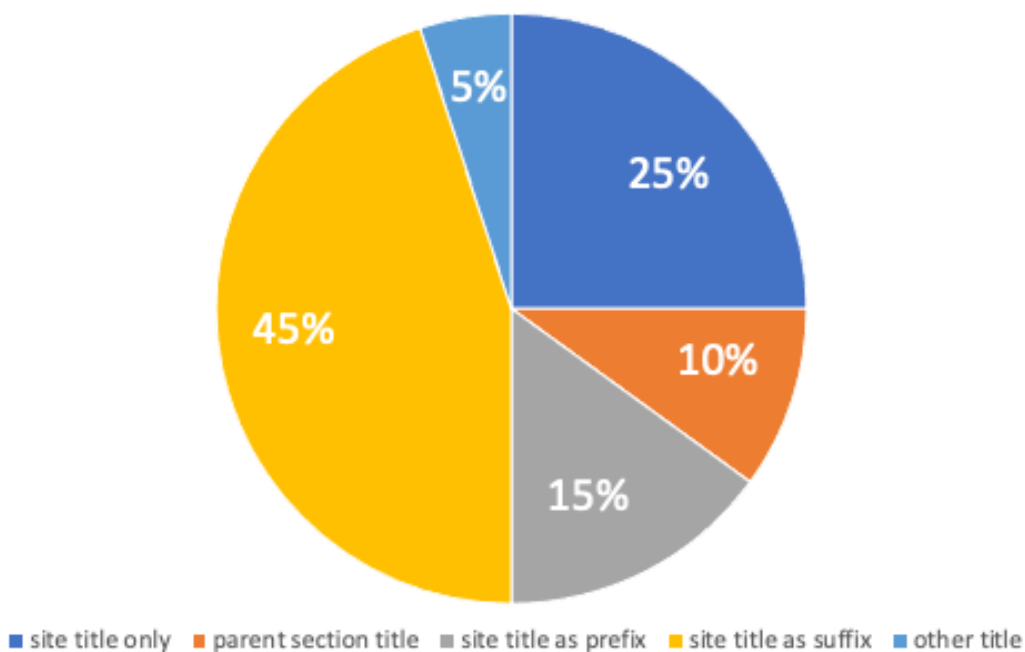


**Figure 7.6.** Title components and order in political campaign issue webpages.

Punctuation and special characters separating the level of titles in the `<title>` tag were not standard across the pages. Separators included: |, -, ::, :, », with some `<title>` tags including multiple punctuation marks. These special characters can cause problems for indexing with search engines or may simply be ignored, depending on the script for the search engine (Bradley, 2015). The lack of consistency is not surprising, as the manuals

provided little advice on how to reference the context and the whole in the `<title>` (or

not to do so). There is a slight trend toward more site titles as a suffix. This would be

consistent with the pattern observed in the *Los Angeles Times*, which despite the various

formations of the site title, always applied it as a suffix. The site title in the suffix also

corresponds to the general SEO and SMO practice of putting the most important (most

topical) terms up front. With the two outlier `<title>` tags for the 2016 campaign that

represent content not found elsewhere in the `<body>` of the page, it is difficult to discern

any clear pattern for the structure of these tags over time in the campaign issue webpages.

   The application of Open Graph and twitter card data doesn't become more

standard in the pages examined until 2012 campaigns, which follows the 2010 Social

Signals Google algorithm change. The exception in the pages examined was found as the

first Open Graph data for a title appeared in a page for a 2008 campaign, which was a

direct duplicate of the `<title>` tag (pc08a). In examining the social media metadata tags

from the 2012, 2014, and 2016 campaigns, eight of the 15 pages used the Yoast SEO

plug-in for WordPress. An early version of the Yoast plug-in resulted in minimal tags,

such as in the following example, which did not include a separate title tag:

```
<!-- This site is optimized with the Yoast WordPress
  SEO plugin v1.2.8.7 - http://yoast.com/wordpress/seo/
  -->
<meta name="description" content="Jon knows that
  Montana's businesses need low taxes, reliable
  infrastructure, and common sense regulations in order
  to grow and create jobs.">
<link rel="canonical"
  href="http://www.jontester.com/issues/creating-
  jobs/">
<!-- / Yoast WordPress SEO plugin. —>. — (pc12d)
```

Later versions included more robust social media metadata, including both Open Graph

and twitter card titles:

```
<!-- This site is optimized with the Yoast SEO plugin
   v3.6.1 - https://yoast.com/wordpress/plugins/seo/ -
   ->
<link rel="canonical"
   href="http://katiemcginty.com/issues/">
<meta property="og:locale" content="en_US">
<meta property="og:type" content="article">
<meta property="og:title" content="Issues | Katie
   McGinty: Democrat for Senate, Pennsylvania">
<meta property="og:description" content="Katie McGinty
   believes working families should come first — and
   that means creating new jobs, caring for our
   communities, and protecting the rights of our
   citizens. Learn how Katie will fight for
   Pennsylvanians in the U.S. Senate. Creating Jobs
   and Growing the Economy &gt;&gt; Our middle-class
   and working families have gotten the short end of
   Read More">
<meta property="og:url"
   content="http://katiemcginty.com/issues/">
<meta property="og:site_name" content="Katie McGinty:
   Democrat for Senate, Pennsylvania">
<meta property="og:image"
   content="http://katiemcginty.com/wp-
   content/uploads/2016/01/facebook-share.jpg">
<meta name="twitter:card" content="summary">
<meta name="twitter:description" content="Katie
   McGinty believes working families should come first
   — and that means creating new jobs, caring for our
   communities, and protecting the rights of our
   citizens. Learn how Katie will fight for
   Pennsylvanians in the U.S. Senate. Creating Jobs
   and Growing the Economy &gt;&gt; Our middle-class
   and working families have gotten the short end of
   Read More">
<meta name="twitter:title" content="Issues | Katie
   McGinty: Democrat for Senate, Pennsylvania">
<meta name="twitter:image"
   content="http://katiemcginty.com/wp-
   content/uploads/2016/01/facebook-share.jpg">
<!-- / Yoast SEO plugin. --> - (pc16d).
```

In this example, the `<title>` tag, Open Graph metadata title, and twitter card title all

have the same content. This is not always the case, even when using the Yoast SEO Plug-

in. The Yoast tool provides data entries that automatically copy the data from the

`<title>` but are also editable. In one of the pages examined, the Open Graph is edited to

only the main site title, whereas the <title> tag contains, "Issues | Scott Brown":

> `<meta property="og:title" content="Scott Brown for`
> `U.S. Senate">` - (pc12c).

The Yoast plug-in also isn't used for all the pages that implemented SMO metadata. One

of the pages aptly calls out the data in an HTML comment as "`<!-- Facebook`

`OpenGraph Protocol stuff -->`" - (pc14b). This is interesting because the web

page creator chose to specifically, and likely manually, define the Open Group Tags,

even though they were using the Yoast plug-in, which has Open Group features. It may

be indicative that the automated Yoast translation is not satisfactory, or it could be a quirk

of that creator. Despite the different implementations of SMO metadata tags, the presence

of the tags increases over time, subsequently increasing the readability and access

potential of the page.

 The campaign issue pages had more distinction between the `<title>` and other

title tags and the content in the `<h1>` or other coding for the main heading of the page in

the `<body>`. Half of the titles represented duplicate titles between the tags with either the

exact title or the page title with the site title as a prefix or suffix, such as:

"`<title>Growing Rural Oregon - Jeff Merkley for U.S. Senate,`

`Oregon.</title>`" with "`<h1>Growing Rural Oregon</h1>`" (pc08c). The

variety between the titles presented a different finding than in the news articles examined.

An example of very different titles includes a page from a 2004 campaign:

```
<title>Tom Coburn for U.S. Senate 2004</title>…
<b>Dr. Coburn's Five Point Prescription for Better and
More Affordable Health Care</b> - (pc04c).
```

Typically, every field that is not a duplicate also has to have been overridden or particularly authored that way; automated fields are good at copying. The wider the use of the `<h1>` tag for the title, the more likely it is to be a duplicate form of the `<title>` tag.

None of the campaign webpages used schema.org tags for metadata on the page. Because several of the pages use the Yoast SEO plug-in for WordPress, and it provides schema.org optimization in the premium version as of 2020, it will be worth tracking to see if the campaign pages start to use schema.org tags. Another controlled schema for terms was found in a unique example of additional metadata using the of Dublin Core[51] metadata schema in the `<head>` of one of the pages:

```
<meta name="DC.Title" content="">
<meta name="DC.Description" content="The official Web
site of James H. &#39;Jim&#39; Webb for U.S. Senate">
<meta name="DC.Creator" content="Kevin Druff, Webb for
Senate">
<meta name="DC.Subject" content="Webb, James Webb,
Senate, Virginia, Born Fighting, Scots Irish, Jim
Webb, George Allen">
<meta name="DC.Publisher" content="">
<meta name="DC.Type" scheme="DCMIType" content="Text">
<meta name="DC.Format" content="text/html">
<meta name="DC.Language" scheme="RFC1766"
content="en">
```

---

[51] https://www.dublincore.org/specifications/dublin-core/dcmi-terms/

```
<meta name="DC.Rights" content=""> - (pc06c)
```

In this example, the title tag for Dublin Core is empty, and the creator tag has content for the page author. This is interesting for a couple of reasons: the items would have to be defined for the Dublin Core tags (they aren't copies of other page text), and the page creator chose to leave the title empty.

### *Descriptions*

Like the titles, multiple fields were examined for metadata descriptions. In looking at the metadata in the `<head>`, descriptions were identified for `<meta name="description">`, "`og:description`", and "`twitter:card`" description tags, and the one instance of the Dublin Core description. Table 7.1 illustrates the source of the content for the various description tags and when and where content was duplicated or unique.

Examples of unique descriptions include an assertion that it is the official website, "The official Web site of James H. &#39;Jim&#39; Webb for U.S. Senate" (pc06c) and "Dr. Heck has spent more than 35 years in public service as a physician, member of the Army Reserve, and community volunteer" (pc16e). Neither of these descriptions are actually descriptions of the page content. In the applications of the `<meta name="description">` tag that isn't the first paragraph, only one of the tag contents describes the page content: "Jon knows that Montana's businesses need low taxes, reliable infrastructure, and common sense regulations in order to grow and create jobs" (pc12d). The contents of the other tags contain slogans or biography information about

**Table 7.1.** Content of description metadata tags for campaign issue pages with descriptions.

| ID | Year | meta description | Open Graph description | Twitter Card description | DC.Description |
|---|---|---|---|---|---|
| **pc04b** | 2004 | | | | |
| **pc04c** | 2004 | unique | | | |
| **pc06c** | 2006 | unique | | | meta description |
| **pc06d** | 2006 | | | | |
| **pc06e** | 2006 | site title | | | |
| **pc08a** | 2008 | | *empty* | | |
| **pc08d** | 2008 | unique | | | |
| **pc08e** | 2008 | | | | |
| **pc10e** | 2010 | site title | | | |
| **pc12c** | 2012 | | unique | | |
| **pc12d** | 2012 | unique | | | |
| **pc12e** | 2012 | unique | meta description | | |
| **pc14a** | 2014 | | unique | unique | |
| **pc14b** | 2014 | | *empty* | | |
| **pc14c** | 2014 | | | " Read more › " | |
| **pc14e** | 2014 | | unique | | |
| **pc16a** | 2016 | | first paragraph | first paragraph | |
| **pc16b** | 2016 | unique | meta description | meta description | |
| **pc16c** | 2016 | first paragraph | meta description | | |
| **pc16d** | 2016 | | first paragraph | first paragraph | |
| **pc16e** | 2016 | unique | meta description | meta description | |

the candidate, which likely applies to the broader website rather than the particular issue

page. For example:

> Senator Casey has been an independent voice who puts Pennsylvania
> families first. He has a record of working with Democrats and Republicans
> to work out fair solutions to problems facing Pennsylvania families
> (pc12e).

Only one other webpage provided an actual description of the page content for the Open

Graph description of a page on "Women's Rights":

> Her leadership gained passage of history-making legislation, known as the
> Shaheen Amendment, which provides health insurance coverage for
> abortion for women serving in the military who are victims of rape or
> incest. She has been outspoken in the fight to stop sexual assault in the
> military. Jeanne was a leader in the effort to reauthorize … (pc14e).

In the instance with a different Open Graph description from the Twitter Card description, it appears to be an error, "Read more ›" (pc14c). The candidate issue pages did not make good use of the `<meta name="description">` tag either for human-readable filtering when viewing search engine results pages or for providing text to support SEO and SMO.

### Meta Keyword

The use of the `<meta name="keywords">` tag was also lacking in the candidate issue webpages. Interestingly, the 2009 keyword trust transition for Google and other search engines did not seem to have an effect on how or when the `<meta name="keywords">` tag was applied. In table 7.2, all of the instances of the `<meta name="keywords">` tag are given. Three of the ten applications had an empty metadata keywords tag (pc04b; pc06b; pc08e). The Black Hat technique of keyword stuffing is also found in many of the pages.

In addition to the obvious fields filled with massive amounts of keywords and key phrases as a keyword stuffing technique, a couple of the pages have added terms that do not appear on the page or are not relevant to the content. For example, in pc06c, these keywords appear in the metadata, "Born Fighting, Scots Irish," that have nothing to do with the page content. In this same page, the keyword is used for the opponent, "George Allen." In this case, it is relevant, as much of the `<body>` text mentions George Allen

**Table 7.2.** `<meta name="keywords">` data used in the campaign issue pages.

| ID | Year | `<meta name="keywords">` |
|---|---|---|
| **pc04b** | 2004 | *empty* |
| **pc04c** | 2004 | www.coburnforsenate.com, Coburn For Senate, Tom, Coburn, 2004, Oklahoma, U.S. Senate, Campaign, GOP, Republican, Republican Party, Republican National Committee, politics, Senate, House, Congress, Conservative, Political activism, 2004 Election, Taxes, Tax Relief, Cuts, Economy, Education, Defense, Judicial Nominees, Protecting Social Security, Prescription, Drugs, Rx Drugs, 2nd Amendment, Homeland Security, Republican party, republican national committee, RNC, gop, republican, republican Party Platform |
| **pc06c** | 2006 | Webb, James Webb, Jim Webb, Senate, Virginia, Born Fighting, Scots Irish, George Allen |
| **pc06d** | 2006 | *empty* |
| **pc06e** | 2006 | Jim Talent, Talent, Senator, Missouri Sen. Talent, Sen. Jim Talent, Senate, MO, Republican, GOP, James Talent, J. Talent, election, campaign |
| **pc08d** | 2008 | smith, Gordon, Gordon Smith, senator, Oregon, OR, election, 2008, campaign, politics, Sen, elect, 08 |
| **pc08e** | 2008 | *empty* |
| **pc10a** | 2010 | &quot;Michael Bennet&quot; Colorado Democrat Veterans Seniors &quot;Reproductive Choice&quot; &quot;Economic Recovery&quot; Economy &quot;Fiscal Responsibility&quot; &quot;Health Care&quot; &quot;Health Care Reform&quot; &quot;Public Option&quot; &quot;National Security&quot; <br>&quot;New Energy&quot; &quot;Clean Energy&quot; Agriculture Rural &quot;Colorado Agriculture and Rural Communities&quot; Senator United States &quot;Bennet for Colorado&quot; &quot;Senator Michael Bennet&quot; &quot;Michael Bennet campaign&quot; &quot;Michael Bennet campaign 2010&quot; &quot;Sen. Michael Bennet&quot; &quot;Sen. Bennet&quot; &quot;Senator Bennet&quot; &quot;Michael Bennet&#39;s campaign&quot; &quot;Michael Bennett&quot; &quot;Bennett for Colorado&quot; &quot;Senator Michael Bennett&quot; &quot;Michael Bennett campaign&quot; &quot;Michael Bennett campaign 2010&quot; &quot;Sen. Michael Bennett&quot; &quot;Sen. Bennett&quot; &quot;Michael Bennett&#39;s campaign&quot;" |
| **pc10d** | 2010 | Joe Sestak Senate Pennsylvania Primary Arlen Specter Congressman Admiral Democrat Democratic Election 2010 Campaign Pat Toomey economy recession jobs |
| **pc16b** | 2016 | Kelly Ayotte for Senate, Ayotte for Senate, Kelly Ayotte, Senator Kelly Ayotte, Sen. Kelly Ayotte, Senator Ayotte, Sen. Ayotte, Kelly Ayotte for New Hampshire, Kelly Ayotte New Hampshire, Ayotte New Hampshire, Kelly Ayotte New Hampshire Senate, Kelly Ayotte Senate New Hampshire, Ayotte New Hampshire Senate, Ayotte Senate NH, Kelly Ayotte for Senate NH, Ayotte for NH Senate, Kelly Ayotte for NH, Kelly Ayotte NH, Ayotte NH, Kelly Ayotte NHSenate, Kelly Ayotte Senate NH, Ayotte NH Senate, Ayotte Senate NH, Kelly Ayotte for Senate NH, Ayotte for NH Senate |

and his policies and how Webb would do things differently. In contrast to this, the page

for pc10d, lists names of two people in the keywords, "Arlen Specter" and "Pat

Toomey," who do not appear in any context on the page itself. This is an adversarial SEO

strategy, because the keywords are not related to the page. Overall, the `<meta`

`name="keywords">` tag is rarely used, but when it is, it is used poorly.

### Keywords in `<body>`

The quality, length, and structure of the primary text in the `<body>` differs

among the issue pages. In most of the pages, however, the keywords and key phrases for

the page can be found in the first paragraph on the page, as in Figure 7.7.

> Agriculture is the lifeblood of rural Oklahoma, and I believe that when
> **farmers thrive**, Oklahoma's entire **economy prospers**. For generations,
> our farmers have worked to feed America and the world, and they have
> been the backbone of **rural communities** that we cannot afford to lose. I
> realize that decisions made in Washington can have a huge impact on
> these farmers and communities: that's why I will always stand up for
> **Oklahoma's family farmers** and promote their interests at home and
> abroad. That means making sure Oklahomans can sell their crops for a fair
> price in the global marketplace, as well as fighting to preserve **commodity
> payments** and conservation programs that help farmers keep our
> environment healthy.

**Figure 7.7.** Keywords in <body> for pc04c with keywords identified in analysis
highlighted in bold.

In one exception, the keywords are in an unordered list in the middle of the page.

(See Figure 7.8.)  This breaks the SEO and SMO advice of putting the more

relevant content toward the beginning of the page. Because the text is coded

specifically as a list, however, search engines will treat the text more like

emphasized text in the analysis for relevancy.

- Passing legislation, now law, to add at least 7.5 billion gallons of ethanol and biodiesel to the **nation's fuel supply** by 2012
- Opening **new markets** for **farmers** including U.S. rice sales to Iraq
- Supporting legislation to help producers add value to their products
- Passing legislation to protect **Missouri's Farm Service Agencies**
- Co-sponsoring legislation to help **lower energy costs** for farmers
- Sponsoring a plan to help **lower health care costs** for Missouri producers

**Figure 7.8.** Keywords in <body> for pc06e with keywords identified in analysis highlighted in bold.

The addition of the unordered list provides another point for the robots to check for content, accuracy, and provide more information to support relevancy. In reviewing the <body> of the pages, most of them are full of buzz and / or keywords to the point where keyword stuffing could also be problematic in the main text.

**Relationships with Other Web Content and Social Media**

The practice of link building and linking to external content was absent from all of the webpages examined with the exception of SMO links to particular social media accounts. It may be a strategy to try to keep the reader on one's site, but it does not increase search engine optimization or access of content because the site isn't defining itself in relation and authority to other pages on the web. As advised from one of the manuals from Chapter V, "[i]nclude links. It's OK to distract people away from your writing. if you are good, they will come back" (Lutze, 2009, p. 112). One of the

webpages had a paragraph with endorsements from other organizations and could signal a good opportunity to link to external websites and build credibility. However, the organizations were not linked and only emphasized in bold text:

> Sen. Talent has been endorsed by the **Missouri Farm Bureau**, the **Missouri Corn Growers Association**, the **Missouri Soybean Association**, the **Missouri Cattlemen's Association**, the **Missouri Dairy Association** and the **Missouri Pork Association** (pc06e).

Since these are official organizations, they also likely carry weight in quality and authority, and the relationships to these sites would increase the quality that search engines assign to the candidate's page.[52] The first instance of social media linking to social media sites for the candidates was for the 2008 campaign (pc08e), and this feature appeared on some level in 75% of the webpages examined from the 2010 campaign through 2016. The range of the social media platforms was greater than observed in the news article webpages and included: Facebook, Twitter, YouTube, Google+, Flickr, Blogspot, Instagram, and LinkedIn with Facebook profile links in all of the occurrences and Twitter account links in all but the 2008 instance. Social media links were clearly important for most of the campaign pages. It is interesting, especially on issue and topic pages, that they all chose not to link to any external sites. This includes citations, congressional records, endorsements, and links for further information or context, which were provided in later years among the news articles. An opportunity lost to both inform

---

[52] The cautionary Black Hat technique here is called link farming, where one links to a bunch of sites to try to increase relevancy and relationships. However, the search engines can tell that they are either not relevant to the content based on text analysis or an over-use, much like keyword stuffing (George, 2005; Ledford, 2009; Shenoy & Prabhu, 2016).

the reader and increase SEO or a solid strategy in keeping the focus on a perspective of the issue?

## Summary

The political candidate issue pages provided a different view of how SEO and SMO strategies may have been applied to these pages in order to increase access to the content. Surprisingly, the pages provided `alt` attributes to `<img>` tags throughout the timeframe of the pages examined and included both more structured HTML tags and more accessibility than found in the earlier *Los Angeles Times* news articles. The page titles were also more distinct across the page itself in the various fields than found in the news articles.

The political candidate issue pages had two structural issues that interfere with SEO and SMO: hidden information and bad metadata. With the first problem, information is buried in the form of images which contained important information of either the page title or a main page image conveying part of the story of the candidate. In these cases, the images were supplied either without an `alt` attribute or used as a background image. The information in the images is not accessible to sight impaired users, search engines, or social media platforms. Fortunately, this problem was fairly rare in the pages examined. In contrast, the metadata applied both for meta tags, `<meta name="description">` and `<meta name="keywords">`, for the HTML `<head>` and social media metadata in the `<head>` was missing, inconsistent, not useful, or misleading. The first paragraph was the best place for identifying relevant keywords

across the pages, and the more helpful descriptions duplicated the content from the first paragraph in the description tag.

Much like the *Los Angeles Times* news article pages, the political candidate issue pages examined had no link building strategies but did include links to social media accounts in the more recent campaigns. The absence of link building may be strategic as part of this genre of pages to limit views and information on issues or it may be an oversight. Regardless, it is interesting, that this essential SEO technique which provides relevancy, credibility, and authority for webpages is missing from these pages.

CHAPTER VIII

CONCLUSION

Through this project, I have explored how search engine optimization (SEO) and social media optimization (SMO) strategies have evolved to keep pace with changes in search engine and social media platform algorithms and requirements, and how the structure of HTML and pages on the Internet has shaped content online. This dissertation demonstrated how communication technologies have been designed from theoretical and practical applications of communication theory, how the new media aspect of web pages is inherently networked and yet decontextualized, how politics of information and information organization become necessary when information quantities are vast, yet also a reflection of the political institutions that construct those systems, and how gatekeeping in communications has moved in the online environment beyond the editor and publisher to search engines and social media platforms.

This project used a media archaeological analysis through the examination of instructional manuals and how-to-guides on SEO and SMO strategies and subsequent analysis of the HTML code for webpages from major persuasive industries, newspapers and political campaigns, to identify strategies and implementation of those strategies in order to provide access to content. A media archaeological approach was used to explore the invisible rules and structures that serve the discourses in online communications through a historical qualitative approach that emphasized the functionality of the technical architecture, operations, and processes that exist within the norms of HTML documents.

The following sections of the conclusion will review the summary of findings to the research questions posed in this study, discuss contributions and limitations of the study, and finally, suggest future directions for study.

## Summary of Findings

This project proposed three primary research questions with each question building off the previous question. First, to establish a historical baseline of SEO and SMO strategies and relationships to prior communication technologies and forms. Secondly, to explore actual historic uses of SEO and SMO strategies through archived web pages. Lastly, to explore how these strategies may have shaped communications online.

### Research Question One

Research question one asked what the historical development of SEO and SMO strategies was over time, including the interplay with changes in proprietary algorithms over time and what topoi were reflected in these SEO and SMO strategies.

In order to begin this analysis and to properly identify topoi in context, this dissertation reviewed pre-existing communication and media systems in the context of information retrieval and access to information. An increase of information repository size necessitates new or modified strategies for retrieving information, in which the role of norms, rules, and gatekeeping determines access to the content across all models. In print catalogs and early databases, taxonomies and vocabularies were necessarily transparent to the user as a reference point and a dual search process, first to find the

terms used in the system, then to find where those terms were applied. With search engines and social media platforms, the rules of HTML and best practices released by the platforms provide some guidance. However, the primary retrieval mechanism is hidden in proprietary algorithms. Within these algorithms, the criteria used to retrieve documents is expanded from the more traditional subject, author, place, time retrieval to include both creator and user relationships, as well as additional unknown characteristics defined by the search engine.

Even with the publicly available versions of major search engine algorithms over time, including twelve in Google alone,[53] much of the SEO and SMO strategies remained static over the 12 years of the published manuals examined. This is primarily because many of the features added to Google's search algorithm were enforcements of HTML and W3C standards and best practices, such as: 1) 2005's Jagger, which focused on good quality and format of links, 2) 2012's Penguin, which further penalized sites not following Google's Webmaster quality guidelines,[54] and 3) 2015's mobile friendliness advantage. The advice to follow basic HTML standards and web best practices around links, well-formatted HTML and Google's Webmaster guidelines was present in all the manuals examined.

Interestingly, the advice on mobile design was sparse in the manuals with only two devoting significant time to mobile design published a decade apart (Rowles, 2018;

---

[53] See ("Google Algorithm Change History," 2015, "Timeline of Google Search," n.d.)

[54] https://developers.google.com/search/docs/advanced/guidelines/webmaster-guidelines. Google's Webmaster Guidelines became "Google Search Console" in 2015. The site contains Google webmaster guidelines starting in 2005 via the Google "Search Central" blog.

Michael & Salter, 2008). This is significant as it demonstrates that the signals and strategies used for SEO and SMO are largely reliant and defined by the structure and standards of HTML. The guidelines applied in configurations of Google's search algorithm, such as Hummingbird (2013) that utilized the semantic web and knowledge graphs, are also reliant on the structure of the semantic web and code allowed on the World Wide Web. For the most part, the changes introduced by search engines and social media platforms are extensions of compliance to those standards. Another important finding from this trend is that if SEO and SMO are necessary skills for communicators in order to make information accessible and to get past gatekeepers that the skills are consistent enough to build expertise without rapid changes and could be further integrated into communications curriculum beyond marketing.

Three significant changes were attended to in the manuals: 1) the elimination of the importance of the `<meta name="keywords">` tag due to its misuse from webpage creators; 2) reduction of need for multiple variants of word forms due to the advancements in the linguistics capabilities of the search engine indices; and 3) the addition of social media platform-specific code and links to social media accounts due to the increased importance of social media in online communications as defined by the major information and communication technology companies.  These suggestions followed Google algorithm changes: 1) 2009's keyword trust, 2) 2013's Hummingbird, which included major semantic and linguistic changes in the search index, and 3) 2010's social signals, which increased both relevancy and individual search results based on social media links. All three of these changes are due to the increased linguistic processing power of machines. The changes around the decreased relevance of the `<meta`

177

`name="keywords">` tag and word variants can be seen and are reflective of how the search engines and platforms transferred trust and expertise to the algorithms.

On the other hand, the latter change with the relevance of social media links increased is designed to set authority and interest based on the community of users, which is then automated through the algorithms. In this instance, the gatekeeper that was the newspaper editor or publisher is replaced by the algorithm. However, the trust in the community of users may be a missed decision, and indeed many scholars argue against popularity as a form of trust. What's interesting from the findings of this project is that with the reliance on the structure of the code, Google is limited in the ways it can establish authority without forming specific partnerships, such as the relationship with Wikipedia to fill the Knowledge Graph panels on the SERPs. Using links and relationships as sources of authority is one of the only entity-agnostic ways of that kind of attribution or trust.

In the review of topoi between SEO and SMO practices, previous structures within information retrieval metadata such as titles, keywords (whether creator or cataloger assigned), and descriptions remain important access points to content, demonstrating their longevity as useful access points to finding and accessing information. With the various titles allowed in HTML and those composed specifically for various platforms, these alternate titles function much like a title on a book spine that is specifically designed for placement on that medium. Even as the `<meta name="keywords">` tag drops from importance, keywords remain an important feature in the structure and rules for information retrieval. As the explicit tag is replaced by keywords in context, extraction is determined by the search engines in URLs, `<title>`

tags, headings, emphasized text, and content on the page, particularly the first paragraph of text.  The description ceases to be a search entry point. However, it remains an important factor in determining access when a user is directly looking at options to select content, such as on a search engine results page or text accompanying a link in a social media platform. The identification of these topoi is important when considering the novelty of the medium and understanding how cultural and structural features are passed from one medium to the next in new media formats.

The analysis of SEO and SMO strategies over time as reflected in instruction manuals and how-to guides demonstrated the continued reliance on traditional metadata points for access: title, keywords, and descriptions. The utility of these metadata is likely to remain as important access points to information as media and information carriers again transform, just as they once did from the analog to the digital. The SEO and SMO strategies also included little variation and remained largely static over time with major changes in search algorithms primarily focusing on stricter compliance with HTML standards and thus not necessitating new strategies for SEO and SMO. It also demonstrates how W3C web compliance standards control and inform the gatekeeping function of search engine and social media platforms.

The changes which occurred around keyword tags and social media links are examples of trust and authority transferred to computing algorithms. This is concurrent with the published efforts by Google in the yet to be released Phantom algorithm changes predicting Google's ability to rank webpages based on "truthfulness" (Dong et al., 2015). SEO and SMO strategies are examples of a place of agency where a webpage creator has an opportunity to jump the gates of the gatekeepers online and promote access to their

179

content. If Google and other platforms move toward more algorithmic trust and define authority and truth, will these opportunities to jump the gates be eliminated? What and whose truth will be represented in this gatekeeper to accessing information and control? Also, does retaining the relationships as authority provide some sort of cover to a sense of neutrality?

**Research Question Two**

Research question two asked how has the development of SEO and SMO strategies been actualized in HTML practices for major persuasive information industries, newspapers and political campaigns. Through this project, HTML code was analyzed from a selection of news articles from the *Los Angeles Times* archived by the Internet Archive and political candidate campaign issue webpages archived by the Library of Congress.

The articles from the *Los Angeles Times* news article webpages (published between 2000 and 2018) used several SEO and SMO strategies, which also changed over time. All of the pages examined in this project were articles linked from the home page of the LA Times website. Due to the use of a content management system, the page structure was consistent, and one article per year was examined. The pages were archived by the Internet Archive in the Wayback Machine. The web archiving tools were unable to capture content for webpages from 2003 and 2004, as during that time content was only available through a subscriber account.

The news articles incorporated well-supplied metadata. The use of `<title>` tags with clear keywords, article headline, keywords and relevant phrases in the first

paragraph of text, and `<meta name="description">` tags that reproduced the first

paragraph was consistent throughout the pages examined. The `<title>` tags in the

`<head>` were consistent with the titles applied on the article page in the `<body>` properly

structured in a `<h1>` tag from 2005 -2018, and a stylized class in the pages examined

from 2000-2002, `<span class="cHeadline1">`. With the change in importance of

the `<meta name="keywords">` tag after 2009 for search engine ranking relevancy, the

news article webpages either dropped the tag or greatly reduced the number of keywords

or key phrases to less than ten. Over time, the news article pages also incorporated

additional SEO and SMO strategies. URL strings were formatted with keywords, e.g.,

http://latimesblogs.latimes.com/technology/2011/10/obama-2012-campaign-starts-a-

tumblog-tumblr.html - (la11). The `alt` attribute in `<img>` tags was regularly added to

make the main page content images visible to both sight impaired viewers and the scripts

of search engines and social media platforms starting in 2013:

```
<img src="./bWhite House OKd spying on allies, U.S.
intelligence officials say - latimes.com_files/la-afp-
getty-u-s--embassy-at-focus-of-nsa-germany-20131028"
alt="U.S. Embassy in Berlin" border="0" width="600"
height="392" title="U.S. Embassy in Berlin"> - (la13).
```

Previous iterations of the news article pages included the `alt` attribute in the frame of the

website, such as for logos and the occasional main page image. The lack of the `alt` as a

required attribute in the earlier pages was surprising, as the effect, intentional or not, is to

hide information that's part of the news story.

Starting in 2009, the news article webpages began to take advantage of the

linking and relationship structures provided by HTML and webpage format through

adding links in the main textual content to definition / topic pages hosted by the *Los*

*Angeles Times* and related articles, as well as links to social media accounts for the

authors and the *LA Times*. The only outbound external sites linked from the webpages

included paid advertisers and other news publications related through the parent media

company. The webpage published in 2018 takes further advantage of the structural

elements defining relationships and metadata by including schema.org tags as micro data

in the `<head>` of the page. The application of schema.org tags should be further

investigated for structural influence and connections with semantic web relationships for

impact on communications and accessibility.

      The political candidate campaign issue webpages examined from closely

contested U.S. Senate campaigns included five sites from each election period between

2002 and 2016 for a total of 40 pages analyzed. The political candidate issue pages used

several structures in HTML that increase SEO and SMO throughout the period examined.

The webpages, on a whole, provided structured tags to headings whether using the

HTML `<hX>` headings tags, emphasized text, or stylized text. These structural elements

are important for the relevancy of search engine ranking, as the keywords within these

highlighted elements have greater weight. The webpages examined did not include a

proper stacking of headings tags until 2008 with:

      `<h1>Growing Rural Oregon</h1>` (pc08c).

The first `<hX>` headings tags were found in 2006, and ten of the webpages examined

between 2006 and 2016 had the page title in an `<h2>` or an `<h3>`, often with the overall

site title tags in the `<h1>`, where the page title should be applied. Although the

application of the headings tags was not compliant with HTML standards, the breadth of

application throughout the period examined demonstrates the importance of headings in these pages.

The political candidate campaign issue webpages also utilized designed URLs strings, which are also SEO and SMO strategies, and can increase relevancy of the webpage with keywords or key phrases and human readable string.  31 of the 40 webpages examined had designed URLs. Here's an example of a well-designed URL from a 2002 campaign:

http://www.timjohnsonforsd.com/workinghard/agriculture.php – (pc02b). Many of the URL strings, although designed and human-readable, did not include keywords beyond a very broad category, such as agriculture, economy, etc., and may have benefited from increased SEO and SMO with more long-tail keywords. Long-tail keywords and key phrases include more specific terms, such as "agriculture-south-dakota" or "agriculture-appropriations" in the string above. The advantage of long-tail keywords in URL strings is less competition for search results pages. Despite not including long-tail keywords, the design of this URL string uses primary keywords that are helpful for search engine ranking and findability.

Another important structural element identified in the political candidate issue webpages was the proper application of the `alt` attribute for images. There was not universal compliance with using the `alt` attribute when describing image content. However, proper usage was found in the earliest campaign cycle examined in 2002. The most useful example was found as early as 2004:

```
<img src="./Pete Coors for U.S. Senate – On The
Issues_Jobs and the Economy_files/hd_coors_right.gif"
width="151" height="178" alt="American Flag flowing
over a Colorado Mountain Range" border="0"> - (pc04e)
```

The alt attribute was applied correctly in 50% of the political candidate webpages examined. In cases where it was not applied correctly, the content in those images was hidden both to sight impaired users and to scripts from search engines and social media platforms. Both search engines and social media platforms are less likely to surface or render pages correctly with images that do not also include machine-readable text in the `alt` attribute. In the case of Google, the absence of the alt attribute will automatically lessen the webpage's ranking in search results.

The political candidate issue pages struggled with providing good metadata for SEO and SMO. In many cases, it was either missing, irrelevant, or misleading, particularly in the `<meta name="description">` and `<meta name="keywords">` tags in the HTML `<head>`, which are hidden from the typical user in a browser view. Good metadata would have included content in the `<meta>` tags that was related to the content in the main `<body>` of the page and aligned with researched user search terms. When the `<meta name="keywords">` tag was supplied (10 out of 40 pages), in 30% of the occurrences it was empty, and in 40% keyword stuffing was used with either an overabundance of words and phrases or words and phrases that did not reflect relevant content on the page, e.g., "Born Fighting, Scots Irish," on pc06c). Tags for social media metadata in the `<head>`, such as Open Graph tags for Facebook and Twitter Cards for Twitter, were also largely missing except in the case of webpages that were built on WordPress with the Yoast SEO plug-in, and in one case where Open Graph tags for

184

Facebook were manually added (pc14b). Most of the content in those tags repeated content in other places, were empty, or used metadata about the larger site instead of about the page. Most of the pages followed the recommendations of placing keywords toward the beginning of the page, and the first paragraph was the best place for identifying relevant keywords across the pages. Pages with a helpful `<meta name="description">` duplicated the content from the first paragraph in this tag. In general, the metadata did not provide additional access points or information to help support SEO and SMO strategies.

The use of outbound links in the political candidate issue pages was limited to social media accounts for the candidates in more recent campaigns. The SEO strategy to build links and create relationships, authority, and verification through linking among websites was not part of the structure of any of the webpages. Even as webpages cited endorsements from various interest groups and organizations, they withheld from linking to the sites of those groups (pc06e).

The findings for how the *Los Angeles Times* utilized SEO and SMO strategies and increased compliance over time demonstrates the increased reliance on search engines and social media platforms to access content. This sometimes resulted in an advantage toward more inclusivity of the content as with the use of the `alt` attribute. With the adoption of semantic web coding for the news articles, it will be interesting to watch if this format and structure becomes common for traditional news media and also if it is adopted by more nefarious "news" sites. As the question of finding accurate information and discerning fake news becomes a greater hurdle in the user's pursuit of accessing

information, the further implementation of SEO and SMO strategies may be necessary to break through the noise and provide content from news organizations.

The findings from the political candidate campaign issues pages and use of SEO and SMO, particularly for metadata, may have a significant impact on the ability of political campaigns to effectively fundraise from individual donors.[55] These pages used the structures that help support SEO and SMO, such as the metadata tags, headings, `alt` attributes (consistently over time if not over all webpages). The content in these tags, however, provided little information that was useful to search engine and social media platforms. Many of these candidate campaign webpages are central hubs for fundraising for the campaigns. If individuals contributing small amounts to political campaigns continue to have increasing influence on elections, then the impact of providing campaign webpage content that conforms to SEO and SMO practices may be essential to capture those donors and thus affect the success of political campaigns. This is particularly important for the tightly contested campaigns examined in this project where more undecided voters may look toward issues instead of candidates and choose their votes and/or donations based on the candidate's online materials.

**Research Question Three**

The third question asked how the SEO and SMO strategies have shaped communications online by looking at the evidence within the archived HTML code of

---

[55] See: The Campaign Finance Institute. (2012, February 8.) *48% of President Obama's 2011 Money Came from Small Donors – Better than Doubling 2007. Romney's Small Donors: 9%.* http://cfinst.org/Press/PReleases/12-02-08/Small_Donors_in_2011_Obama_s_Were_Big_Romney_s_Not.aspx

selected webpages from the *Los Angeles Times* news articles and political candidate campaign issue pages. Four areas of note were identified while looking at the HTML webpages: 1) the application of headings, 2) the construction of `<title>` tags, 3) the application of the `alt` attribute to expose content, and 4) the construction of links within the pages. In the application headings and use of headings tags, both the news articles and the political candidate webpages progressed to more structured formats over time that could be more easily machine-readable and signal relevant text to search engines and social media platforms. Whereas the news articles largely did not use subheadings, the transition to correctly using the `<h1>` tag for the headline is still notable.

In order to push webpages higher in the search result rankings and work with the search engine platforms, do news content producers need to reconsider composition and structure to include more headings and emphasis tags? The political candidate webpages take advantage of more of the structures in HTML available and frequently use headings and emphasis tags in body content.  If news articles adapted the use of more headings and emphasis tags in the article text, would it increase the findability and ranking of the articles with search engines? Or does the quality writing in these news articles supersede the need for these web formatting and hierarchies within the text body? Will these structural elements in HTML have a long-term effect on writing for the web as genres of content within the Internet begin to merge in structure and form, in order to ensure content is available through the virtual gatekeepers?

In the construction of the `<title>` tags for the pages, the guidance for SEO and SMO was fairly limited and the structure and application of the tags resulted in varied applications among the webpages. The *Los Angeles Times* pages consistently used a

suffix in the `<title>` tag in order to link the page title to the newspaper as a whole, although the form of the title of the whole changed. Conversely, the political candidate web pages had titles that referenced the main site at times but were not connected to the content on the page at all, such as with a tagline. The HTML structure with SEO and SMO strategies allowed for a nuanced format for metadata titles that was not applied in the pages examined.

The third example of SEO and SMO strategies in influencing HTML code and communications online is the increased usage of the `alt` attribute for images. It is likely that because both search engines and social media applications promote pages with proper `alt` attributes and the images can be more easily machine readable that this also prompted increase usage and exposed / communicated content in multiple ways that may have only been available through viewing the images previously.

The final example of communications practices is significant for the lack of application, which is the addition of outbound external links. One of the new media features of webpages is the ability to link and remix content from various sites, which search engines and social media platforms use as an indicator of authority. The primary outbound links found on the pages were to social media platform accounts directly associated with the HTML pages and their creators and organizations or paid / related financial interest sites. The more recent news articles in the *Los Angeles Times* started to take advantage of linking relationships within the text of the article body by linking to other online properties they owned, such as glossaries, or related news articles directly from the text. This usage was closer to what was expected among all the pages, yet was still limited to related entities.

188

At this time, it does not appear that the reliance on linking and community-defined relationships as valued by search engines and social media platforms has translated into changes for the *Los Angeles Times* news article and political candidate webpages. This is important for discussion of authority and credibility in online content and, perhaps, which industries may not have bought into that concept. If that is the case, what kind of impact is it having on making their content accessible? This project was limited in its examination of news articles from the *Los Angeles Times*, and through a cursory look at news articles from the *Washington Post* and the *NY Times,* they are more likely to link to external sites within article text. Does this external linking and relationship building between external sites grant them more authority for the algorithms? If news media does not use this strategy, are users less likely to find their content? Is their relevancy to a broader audience limited in some way because they do not use it? Are they limiting the authority of their content for search engine and social media platforms by restricting readers to links within their own created content, advertisements, and sister newspapers? The question to be explored further is whether the benefits of keeping users on one's site outweigh the benefits of building relationships and authority on the web through external linking and thus increasing ranking within SERPs.

The political candidate campaign issues pages also refrained from external linking. In the cases of these pages, is the desire to keep them onsite worth the trade of participating in the game of external links? Will this practice have a long-term effect on the findability of these pages and thus the ability to raise funding from individual donors? Could candidates who take advantage of the impact of external linking on authority and ranking within search engines be better positioned to reach wider audiences? Will the use

and strategies of the search engines and the social platforms in coding, structure, and metadata content impact elections?

## Contributions of the Study

This media archaeological analysis of SEO and SMO strategies through instruction manuals and how-to guides and then verified for actualization of practices through persuasive industries of newspapers and political campaigns exposes portions of the hidden mechanisms and rules that determine what content is accessible online. This project is an important contribution and complementary study to the communications field where research on gatekeeping of search engines and social media platforms has focused on the algorithms and advocates for exposing those hidden mechanisms. Until those systems are disrupted or exposed, what strategies can be used to make content more findable? This study provides a complementary view of the online communications environment by looking at how webpage authors communicate within sociotechnological and cultural structures via the hidden and chosen few exposed rules from search engines and social media platforms in order to make content accessible.

This dissertation is important for communication studies to develop an understanding of how we enable and influence discussions in our current digital cultural moment and to provide strategies for how communications are accessed. The practical implications of this study include opportunities to further implement SEO and SMO strategies in order to increase access to content, which in both the news articles and political campaigns may include allowing outbound links and relationships between internal and external content. At the same time as playing the game, by further

implementing SEO and SMO strategies, this project also exposes through examination of the actualization of the techniques a process by which HTML web content can be interrogated to provide evidence as to why some content may be more viral based on the structure and not content alone.

## Limitations of the Study

There were three primary limitations to this study. The first was the availability and selection of content from published manuals. Because a lot of outdated manuals are withdrawn from library collections or no longer published and available for sale, the instruction manuals and how-to guides were limited by availability.  The second limitation related to selection and analysis of the HTML archived webpages. This second limitation had three components: 1) the pages selected to be archived the archives; 2) the content that was able to be harvested by the web archiving tools; and 3) the chunkiness of tools available for exploring web archives. The Internet Archive has recently released tools to explore data within the Wayback Machine archive through an Application Programming Interface (API).[56] A project using the API may be able to compare broader sets of data and return and analyze results more efficiently. The Library of Congress is also currently developing tools to make the United States Election Archive and their other archived web collections easier to analyze for researchers.[57] Finally, the media

---

[56] Early editions of the API allowed for querying of the existence of pages but did not provide more complex queries. https://archive.readme.io/docs/overview

[57] https://www.loc.gov/apis/

archaeological approach is focused on structure and form and is both useful to explore content structures within norms and influences on communications capabilities. A more traditional content discourse analysis may provide a complementary set of exploration by analyzing the content of the text for meaning, order, and context of the webpages within the websites.

An additional limitation of this study is the focus on U.S. politics and English language webpages. Three of the political candidate issue pages examined included multiple language variations of the page. How are SEO and SMO implemented within these additional languages and how does it compare with the English language pages? Is there a trend to provide more or less multi-lingual versions of the candidate content pages and what may be the motivation to do either in terms of the algorithms and systems that identify and provide relevant content through search engines and social media platforms?

## Future Directions

One of the goals of this project was to lay some groundwork on the usage of SEO and SMO in HTML pages in the context of prior media as a complementary analysis to critical studies of search engine and social media platform algorithms. An interesting future study should examine the structural content of HTML pages and online media that has had a demonstrated effect on bias and relevance in search engine and social media platforms. Better understanding of the structural elements that affect the relevancy of online content may help as a side project to combatting current contemporary communications issues such as identifying fake news.

With the increase in tools that provide schema.org and support the semantic web and structured data, it will be interesting to examine how and when schema.org tags are used in content online. Since this information is within the code of the page and not viewable in the rendered browser view, will it produce the same issues as the `<meta name="keywords">` tag and be misused? Will the more controlled structure reduce the likelihood of Black Hat techniques? Will future platforms create new standards for each of their platforms, like Twitter, and if so, how will this affect the use of such a schema and content on that platform?

This project provided a framework for reviewing SEO and SMO strategies actualized in HTML pages and an analysis of how those structures reinforce and merge prior communication and information access strategies. Future projects may extend the scope, content, and focus in order to more closely examine particular genres of content online.

**Example Data Collection Sheet for Manuals and How-To Guides**

1. **ID.** Identification number assigned by researcher to track data related to texts.

2. **Author(s) biography and expertise.** What information is provided about the author(s), typically located on back cover, front matter, or in introduction. How does this information situate the author as an expert in SEO and/or SMO? What is emphasized about the authors for the source of their expertise in this subject matter?

3. **Goals.** Since ostensibly, manuals and how-to guides are to support some sort of skill development, what are the goals our outcomes expected for the reader after finishing the text? How does the author(s) identify for address possibly conflicting goals?

4. **Audience.** Does the author identify a specific audience by discipline or ability for the text? What is the technical expertise expected of the reader?

5. **SEO strategies.** What specific SEO strategies does the author(s) recommend? What context is provided to support these strategies? Are there references to growth or change in techniques over time?

6. **SMO strategies.** What specific SMO strategies does the author(s) recommend for HTML on-page strategies? What context is provided to support these strategies? Are there references to growth or change in techniques over time? Are specific social media platforms addressed?

7. **Recommended tools.** Does the author(s) recommend any specific tools to either produce or test for SEO and SMO? What kind of technical expertise is needed to use these tools?

8. **Specific algorithms or search version called out.** Are any specific search engine algorithms addressed specifically in the text? If so, are they all related to Google? How does the author(s) situate the algorithms within recommended strategies?

9. **Writing and communication tips.** What tips or strategies are recommended for text composition? How do these strategies relate to the SEO and SMO strategies in the text?

10. **Design tips.** What tips or strategies are recommended for visual design or placement of media and text on web pages? How do these strategies relate to the SEO and SMO strategies in the text?

11. **Black Hat.** Are Black Hat or subversive SEO and SMO strategies addressed in the text? If so, how are they framed? Does the author(s) provide as information or is it enveloped with recommendations to either do or not do, as well?

### Example Webpages Data Collection Sheet

1. **ID.** Identification number assigned by researcher to track data and files.

2. **Harvest Date.** What date was the webpage harvested?

3. **URL design.**

    a. URL content

    b. How is the URL designed? What strategies are used?

4. **Title tag.**

   a. Content of tag

   b. Is the title tag the same as the article title? What differs? How long is it? Does it have any distinguishing features?

5. **<meta name="description">.**

   a. Content of tag

   b. Does the descriptive content differ or is it a copy from the <body> text? What differs? How long is it? Does it have any distinguishing features?

6. **<meta name="keywords">.**

   a. Content of tag

   b. Does the page provide <meta> keywords? How many? Are there any distinguishing features?

7. **<meta>.** Additional <meta> tags to note.

8. **Schema.org.** Which schema.org tags are used to classify page content?

9. **Open graph.** Which Open Graph elements are used?

10. **Twitter card.** Which Twitter card metadata are provided?

11. **Publication Date.** What publication date and time is listed with the article?

12. **Article Author.** Who is the author listed of the article if applicable?

13. **Article Title.**

    a. What is the title of article as rendered on the HTML page?

    b. What tag or structure holds the title?

14. **Structured and emphasized text.** Does the body have well-structured headings? If so, do they contain keywords? Are bold and italics tags used to highlight keywords within the body?

15. **Keywords in body.** Where do the `<meta name="keywords">` repeat within the `<body>` text? **NEI.** Not easily identified.

16. **Links**. How many outbound links are present? How many within the site?

17. **Social Share**. Does the page have social share buttons? If so, which ones?

18. **Accessibility.** Are `alt` attributes uses with `<img>` tag? What other accessibility features are coded in the page?

19. **Other**. Are there other distinguishing structural features of the content, page, design, or code?

20. **Screenshots**. Screenshots of features of code or the page worth noting.

21. **Number of articles or issues on single web page**. *Political candidate pages only*.

APPENDIX B:

DATA SELECTION OF POLITICAL CANDIDATE WEBSITES

**Data Harvest Condition Collected**

1. **State**

2. **Election year**

3. **Candidate Winner**

   a. **Harvest Condition**

      i. **H.** Harvested.

      ii. **HPO.** Home page only.

      iii. **MIPP.** Multiple issues per page / on a single webpage.

      iv. **NUP.** No useful pages. Typically, a site like this is a menu / frame only with broken content, pages within site are produce errors or weren't harvested, has a click-through ad that blocks getting to any content, or all content on site is a list of news stories from news media publications.

      v. **NH.** No archived copies of the website available in the Library of Congress Collection.

4. **Second Place Candidate**

   a. **Harvest Condition**

5. **Margin of Victory.** Percent of the vote for winning candidate over the second-place candidate's percent of the vote.

# Data Collected for Political Candidate Condition

**Table B.1.** U.S. Senate closest races and available condition of harvested webpages with issue content at the Library of Congress.*

| State | Year | Winner | Harvest Condition | Second-place | Harvest Condition | Margin of victory % |
|---|---|---|---|---|---|---|
| SD | 2002 | Tim Johnson | H | John Thune | H | 0.1 |
| MS | 2002 | Jim Talent | MIPP | Jean Carnahan | H | 1.1 |
| MN | 2002 | Norm Coleman | HPO | Walter Mondale | NH | 2.2 |
| LA | 2002 | Mary Landrieu | H | Suzanne Haik Terrell | H | 3.4 |
| SD | 2004 | John Thune | HPO | Tom Daschle | NH | 1.1 |
| FL | 2004 | Mel Martinez | H | Betty Castor | H | 1.2 |
| OK | 2004 | Tom Coburn | H | Brad Carson | H | 1.6 |
| KY | 2004 | Jim Bunning | H | Daniel Mongiardo | H | 2 |
| AK | 2004 | Lisa Murkowski | H̶ | Tony Knowles | H | 3.1 |
| MS | 2004 | Kit Bond | HPO | Nancy Farmer | NH | 3.2 |
| CO | 2004 | Ken Salazar | H | Pete Coors | H | 4.8 |
| VA | 2006 | Jim Webb | MIPP | George Allen | H | 0.4 |
| MT | 2006 | Jon Tester | MIPP | Conrad Burns | H | 0.9 |
| MS | 2006 | Claire McCaskill | NUP | Jim Talent | H | 2.3 |
| TN | 2006 | Bob Corker | H | Harold Ford Jr | HPO | 2.7 |
| AK | 2008 | Mark Begich | H | Ted Stevens | H | 1.2 |
| GA | 2008 | Saxby Chambliss | NUP | Jim Martin | NUP | 3 |
| OR | 2008 | Jeff Merkley | H | Gordon H Smith | H | 3.3 |
| NJ | 2008 | Frank Lautenberg | H | Dick Zimmer | NH | 6 |
| CO | 2010 | Michael Bennet | H | Ken Buck | H | 1.8 |
| IL | 2010 | Mark Kirk | H | Alexi Giannoulias | HPO | 1.9 |
| PA | 2010 | Pat Toomey | H | Joe Sestak | H | 2.02 |
| WA | 2010 | Patty Murray | H | Dino Rossi | NUP | 4.8 |
| ND | 2012 | Heidi Heitkamp | NUP | Rick Berg | H | 0.9 |

**Table B.1. (continued).***

| State | Year | Winner | Harvest Condition | Second-place | Harvest Condition | Margin of victory % |
|---|---|---|---|---|---|---|
| NV | 2012 | Dean Heller | H | Shelley Berkley | NUP | 1.2 |
| AZ | 2012 | Jeff Flake | NUP | Richard Carmona | H | 3 |
| MT | 2012 | Jon Tester | H | Denny Rehberg | H | 3.7 |
| PA | 2012 | Bob Casey Jr | H | Tom Smith | NUP | 9.1 |
| VA | 2014 | Mark Warner | H | Ed Gillespie | H | 0.8 |
| NC | 2014 | Thom Tillis | NUP | Kay Hagan | NUP | 1.5 |
| CO | 2014 | Cory Gardner | HPO | Mark Udall | H | 1.9 |
| AK | 2014 | Dan Sullivan | H | Mark Begich | H | 2.2 |
| NH | 2014 | Jeanne Shaheen | H | Scott Brown | MIPP | 3.2 |
| NH | 2016 | Maggie Hassan | H | Kelly Ayotte | H | 0.1 |
| PA | 2016 | Pat Toomey | H | Katie McGinty | H | 1.4 |
| NV | 2016 | Catherine Cortez Masto | HPO | Joe Heck | H | 2.4 |
| MS | 2016 | Roy Blunt | NUP | Jason Kander | NUP | 2.8 |

*In Table B.1, the shadowed cells represent content that was not selected for analysis due to the lack of useful or presence of content.*

REFERENCES CITED

*6 Reasons Why We Like .ME Domain Names*. (2013, January 11).
https://www.name.com/blog/domains/2013/01/6-reasons-why-we-like-me-
domain-names/

Allen, R. (2016). *The Rise of the Bots – What marketers need to know about chatbots*.
http://www.smartinsights.com/managing-digital-marketing/managing-marketing-
technology/the-rise-of-the-bots/

*American Feel Better Informed Thanks to the Internet*. (2014). Pew Research Center.

Assistant Secretary for Public Affairs. (2016, December 7). *Writing for the Web*.
Department of Health and Human Services. http://writing-for-the-web.html

Bar-Ilan, J. (2007). Google bombing from a time perspective. *Journal of Computer-
Mediated Communication*, *12*, 910–938. https://doi.org/10.1111/j.1083-
6101.2007.00356.x

Bates, M. (2002). Toward and Integrated Model of Information Seeking and Searching.
*New Review of Information Behaviour Research*, *3*, 1–15.

Baym, N. K. (2010). *Personal Connections in the Digital Age*. Polity Press.

Beer, D. (2017). The Social Power of Algorithms. *Information, Communication &
Society, 20*(1), 1-13.

Berman, R., & Katona, Z. (2013). The Role of Search Engine Optimization in Search
Marketing. *Marketing Science*, *32*(4), 644–651.
https://doi.org/10.1287/mksc.2013.0783

Berners-Lee, T. (1989). *Information Management: A Proposal*. CERN.
https://www.w3.org/History/1989/proposal.html

Berry, D. M. (2011). *The Philosophy of Software: Code and Mediation in the Digital
Age*. Palgrave Macmillan.

Bhargava, R. (2006, August 10). 5 Rules of Social Media Optimization (SMO).
*Influential Marketing Group Blog*.

Boutet, C.-V., & Quoniam, L. (2012). Towards Active SEO (Search Engine
Optimization) 2.0. *Journal of Information Systems and Technology Management*,
*9*(3), 443–458. https://doi.org/10.4301/S1807-17752012000300001

Bradley, S. V. (2015). *Win the Game of Googleopoly: Unlocking the Secret Strategy of
Search Engines*. Skillsoft/ John Wiley & Sons.

Brin, S., & Page, L. (1998). The anatomy of a large-scale hypertextual Web search engine. *Computer Networks and ISDN Systems*, *30*(1–7), 107–117. https://doi.org/10.1016/S0169-7552(98)00110-X

Brown, R. H., & Davis-brown, B. (1998). *The making of memory: The politics of archives , libraries and museums in the consciousness*. *11*(4), 17–32.

Brügger, N. (2012). Web historiography and Internet Studies: Challenges and perspectives. *New Media & Society*, *15*, 752–764. https://doi.org/10.1177/1461444812462852

Bush, V. (1945). As we may think. *The Atlantic*. https://www.theatlantic.com/magazine/archive/1945/07/as-we-may-think/303881/

The Campaign Finance Institute. (2012, February 8.) *48% of President Obama's 2011 Money Came from Small Donors – Better than Doubling 2007. Romney's Small Donors: 9%.* http://cfinst.org/Press/PReleases/12-02-08/Small_Donors_in_2011_Obama_s_Were_Big_Romney_s_Not.aspx

Carey, J. (1989). *Communication as Culture: Essays on Media and Society*. Unwin Hyman.

Castells, M. (2000). *The rise of the network society* (2nd ed.). Blackwell Publishers, Ltd.

Castillo, C. (2010). Adversarial Web Search. *Foundations and Trends® in Information Retrieval*, *4*(May 2011), 377–486. https://doi.org/10.1561/1500000021

Cerf, V. G., & Kahn, R. E. (1974). A Protocol for Packet Network Intercommunication. *IEEE Transactions on Communications*, *Com-22*(5), 637–648.

Chun, W. H. K. (2006). *Control and Freedom: Power and Paranoia in the Age of Fiber Optics*. MIT Press.

Coddington, M. (2012). Building Frames Link by Link: the Linking Practices of Blogs and News Sites. *International Journal of Communication*, 6, 2007-2026.

Conway, F., & Siegelman, J. (2009). *Dark Hero of the Information Age: In Search of Norbert Wiener, the Father of Cybernetics*. Basic Books.

Costa, M., Gomes, D., & Silva, M. J. (2017). The Evolution of Web Archiving. *International Journal on Digital Libraries*, *18*, 191–205.

Cui, X., & Liu, Y. (2017). How Does Online News Curate Linked Sources? A Content Analysis of Three Online News Media. *Journalism, 18*(7), 852-870.

De Maeyer, J. (2012). The Journalistic Hyperlink: Prescriptive Discourses about Linking in Online News. *Journalism Practice*, 6(5-6), 692-701.

De Maeyer, J., & Holton, A. E. (2016). Why Linking Matters: A Metajournalistic Discourse Analysis. *Journalism*, *17*(6), 776–794. https://doi.org/10.1177/1464884915579330.

Dean, J. (2009). *Democracy and Other Neoliberal Fantasies: Communicative Capitalism and Left Politics*. Duke University Press.

Deuze, M. (2003.) The Web and Its Journalisms: Considering the Consequences of Different Types of NewsMedia Online. *New Media & Society*, 5(2), 203-230.

Deuze, M. (2006). Participation, Remediation, Bricolage: Considering Principal Components of a Digital Culture. *The Information Society*, *22*, 63–75. https://doi.org/10.1080/01972240600567170

Dewey, J. (1946). *The Public and Its Problems: An Essay in Political Inquiry*. Gateway Books.

Dimmick, J. (1974). *The Gatekeeper: An Uncertainty Theory* (Vol. 37). The Association for Education in Journalism.

*Domain Name Registration Process | ICANN WHOIS*. (n.d.). Retrieved December 28, 2020, from https://whois.icann.org/en/domain-name-registration-process

Dong, X. L., Gabrilovich, E., Murphy, K., Dang, V., & Watts, I. (2015). Knowledge-Based Trust: Estimating the Trustworthiness of Web Sources. *Arxiv Preprint*, *Section 3*.

Ernst, W. (2005). Let There Be Irony: Cultural History and Media Archaeology in Parallel Lines. *Art History*, *28*(5), 582–603.

Ernst, W. (2013). *Digital Memory and the Archive* (Jussi Parikka, Ed.). University of Minnesota Press.

European Commission. (2014). Factsheet on the "Right to be Forgotten" ruling (c-131/12). *European Commission*.

Fall, K. R., & Stevens, W. R. (2011). *TCP/IP Illustrated, Volume 1* (2nd edition, Vol. 15). Addison-Wesley Professional. https://books.google.com/books?id=X-l9NX3iemAC&pgis=1

Febvre, L., & Martin, H.-J. (1976). *The Coming of the Book: The Impact of Printing 1450-1800*. Verso.

Fink, K., & Schudson, M. (2014). The Rise of Contextual Journalism, 1950's-2000's. *Journalism, 15*(1), 3-20.

Fishkin, R. (2008, June 30). White Hat Cloaking: It Exists. It's Permitted. It's Useful. *Moz*.

Fishkin, R. (2015a). *The Beginner's Guide to SEO*. https://moz.com/beginners-guide-to-seo

Fishkin, R. (2015b, January 1). The Beginner's Guide to SEO. *Moz*.

Foster, M. (2015, February 27). What is Social Media Optimization? *Payment Week*.

Foucault, M. (1972). *The Archaeology of Knowledge & the Discourse on Language* (A. M. S. Smith, Trans.). Pantheon Books.

Franklin, M. (2013). *Digital Dilemmas: Power, Resistance, and the Internet*. Oxford University Press.

Fuchs, C. (2008). *Internet and society: Social theory in the information age*. Routledge.

Fuchs, C. (2010). Labor in Informational Capitalism and on the Internet. *The Information Society*, *26*(3), 179–196. https://doi.org/10.1080/01972241003712215

Fuchs, C. (2012a). Dallas Smythe Today—The Audience Commodity, the Digital Labour Debate, Marxist Political Economy and Critical Theory. Prolegomena to a Digital Labour Theory of Value. *Triple C: Cognition, Communication, Co-Operation*, *10*(2), 692–740.

Fuchs, C. (2012b). The Political Economy of Privacy on Facebook. *Television & New Media*, *13*(2), 139–159. https://doi.org/10.1177/1527476411415699

Galloway, A. R., & Thacker, E. (2007). *The Exploit: A Theory of Networks*. University of Minnesota Press.

Gee, J. P. (1999). *An Introduction to Discourse Analysis: Theory and Method*. Routledge.

George, D. (2005). *The ABC of SEO : search engine optimization strategies*. Lulu Press.

Gerlitz, C., & Helmond, A. (2013). The like economy: Social buttons and the data-intensive web. *New Media & Society*, *15*(8), 1348–1365. https://doi.org/10.1177/1461444812472322

Gitelman, L. (2006). *Always Already New: Media, History, and the Data of Culture*. MIT Press.

Gitelman, L. (2014). *Paper Knowledge: Toward a Media History of Documents*. Duke University Press.

Goldsmith, J., & Wu, T. (2006). *Who Controls the Internet?: Illusions of a Borderless World*. Oxford University Press.

*Google Algorithm Change History*. (2015). Moz. https://moz.com/google-algorithm-change

Google does not use the keywords meta tag in web ranking. (2009, September 21). *Google Webmaster Central*.

Granka, L. A. (2010). The Politics of Search: A Decade Retrospective. *The Information Society*, *26*, 364–374. https://doi.org/10.1080/01972243.2010.511560

Guilbaud, G. T. (1959). *What is Cybernetics?* Heinemann.

Hall, S. (2006). Encoding / Decoding. In M. G. Durham & D. Kellner (Eds.), *Media and Cultural Studies: Key Works* (Revised Ed, pp. 163–173). Blackwell.

Haraway, D. (1987). A manifesto for Cyborgs: Science, technology, and socialist feminism in the 1980s. *Australian Feminist Studies*, *2*(4), 1–42. https://doi.org/10.1080/08164649.1987.9961538

Hayles, N. K. (2004). Print Is Flat, Code Is Deep: The Importance of Media-Specific Analysis. *Poetics Today*, *25*(Spring 2004), 67–90. https://doi.org/10.1215/03335372-25-1-67

Helton, L. E. (2019). On Decimals, Catalogs, and Racial Imaginaries of Reading. *PMLA*, *134*(1), 99–120. https://doi.org/10.1632/pmla.2019.134.1.99

Hermida, A., Fletcher, F., Korell, D., & Logan, D. (2012). SHARE, LIKE, RECOMMEND. *Journalism Studies*, *13*(5–6), 815–824. https://doi.org/10.1080/1461670X.2012.664430

Hirsch, P. M. (1972). Processing Fads and Fashions: An Organization-Set Analysis of Cultural Industry Systems. *American Journal of Sociology*, *77*(4), 639–659. https://doi.org/10.1086/225192

Houston, R. D., & Harmon, G. (2007). Vannevar Bush and Memex. *Annual Review of Information Science and Technology*, *41*(1), 55–92. https://doi.org/10.1002/aris.2007.1440410109

Huhtamo, E., & Parikka, H. (Eds.). (2011). *Media Archaeology: Approaches, Applications, and Implications* (Kindle Edi). University of California Press.

*Internet Use Over Time*. (2014).

Introna, L., & Nissenbaum, H. (2007). *Shaping the Web: Why the politics of search engines matters.* 169–185. https://doi.org/10.1080/01972240050133634

Israel, T., Collins-thompson, K., & Kurland, O. (2013). Shame to be Sham: Addressing Content-Based Grey Hat Search Engine Optimization. *Proceedings of the 36th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '13)*, 1013–1016. https://doi.org/10.1145/2484028.2484135

Jenkins, H. (2004). The Cultural Logic of Media Convergence. *International Journal of Cultural Studies*, *7*(1), 33–43. https://doi.org/10.1177/1367877904040603

Jones, K. B. (2013). *Search engine optimization: Your visual blueprint for effective Internet marketing*. John Wiley & Sons.

Katz, E. & P.F. Lazarsfeld. (1955). *Personal Influence.* The Free Press.

Kelsey, T. (2016). *Introduction to Search Engine Optimization A Guide for Absolute Beginners*. A Press.

Khang, H., Ki, E.-J., & Ye, L. (2012). Social Media Research in Advertising, Communication, Marketing, and Public Relations, 1997-2010. *Journalism & Mass Communication Quarterly*, *89*(2), 279–298. https://doi.org/10.1177/1077699012439853

Killoran, J. B. (2013). How to use search engine optimization techniques to increase website visibility. *IEEE Transactions on Professional Communication*, *56*(1), 50–66. https://doi.org/10.1109/TPC.2012.2237255

Kinsella, S., Passant, A., & Breslin, J. G. (2011). *Topic Classification in Social Media using Metadata from Hyperlinked Objects*. *1380*(January).

Kitchin, R., & Dodge, M. (2011). *Code/Space: Software and Everyday Life*. MIT Press.

Kittler, F. (1995). There is No Software. *CTHEORY*, *October 18, 1995*.

Krajewski, M. (2011). *Paper Machines: About Cards & Catalogs, 1548-1929*. MIT Press.

Krapp, P. (2006). Hypertext Avant La Lettre. In W. H. K. Chun & T. Kennan (Eds.), *New Media Old Media: A History and Theory Reader* (pp. 359–373). Taylor & Francis.

Krogue, K. (2012, July 20). The Death Of SEO: The Rise Of Social, PR, And Real Content. *Forbes*.

Landow, G. P. (2006). *Hypertext 3.0: Critical theory and new media in an era of globalization*. JHU Press. http://books.google.com/books?hl=en&lr=&id=exzQDHI8rpQC&pgis=1

Ledford, J. L. (2009). *Search Engine Optimization Bible*. John Wiley & Sons.

Lessig, L. (2006). *Code* (Version 2.). Basic Books.
http://codev2.cc/download+remix/Lessig-Codev2.pdf

Levin, S. (2017, May 16). Facebook promised to tackle fake news. But the evidence shows it's not working. *The Guardian*.

Library of Congress. (2002, November 25). *ONIX TOC*. Bibliographic Enrichment Advisory Team. https://www.loc.gov/catdir/beat/onix.toc.html

Light, B., & McGrath, K. (2010). Ethics and social networking sites: A disclosive analysis of Facebook. *Information Technology & People*, *23*(4), 290–311. https://doi.org/10.1108/09593841011087770

Lincoln, S. R. (2009). *Mastering Web 2.0: Transform your business using key website and social media tools*. Kogan Page.

Lindenthal, T. (2014). Valuable words: The price dynamics of internet domain names. *Journal of the Association for Information Science and Technology*, *65*(5), 869–881. https://doi.org/10.1002/asi.23012

Lipsman, A., Mudd, G., Rich, M., & Bruich, S. (2012). The Power of Like: How Brands Reach (and Influence) Fans through Social-Media Marketing. *Journal of Advertising Research*, *51*(1), 40–52.

*Los Angeles Times | History, Ownership, & Facts. (2019, Apr. 5).* Encyclopedia Britannica, from https://www.britannica.com/topic/Los-Angeles-Times.

Lovejoy, K., & Saxton, G. D. (2012). Information, Community, and Action: How Nonprofit Organizations Use Social Media. *Journal of Computer-Mediated Communication*, *17*(3), 337–353. https://doi.org/10.1111/j.1083-6101.2012.01576.x

Lu, L., & Lee, W. (2011). SURF : Detecting and Measuring Search Poisoning Categories and Subject Descriptors. *Proceedings of the 18th ACM Conference on Computer and Communications Security (CCS '11)*, 467–476. https://doi.org/10.1145/2046707.2046762

Lutze, H. (2009). *The findability formula: The easy, non-technical approach to search engine marketing*. John Wiley & Sons.

Mackenzie, A. (2006). *Cutting Code: Software and Sociality*. Peter Lang.

MacKenzie, D., & Wajcman, J. (Eds.). (1999). *The Social Shaping of Technology* (2nd ed.). Open University Press.

Mager, A. (2012). Search Engines Matter: From Educating Users Towards Engaging with Online Health Information Practices Search Engines Matter: From Educating Users Towards Engaging with Online Health Information Practices. *Policy & Internet*, *4*(2), Art. 7. https://doi.org/10.1515/1944-2866..1166

Mager, A. (2013). *In search of ideology: Socio-cultural dimensions of Google and alternative search engines*. http://epub.oeaw.ac.at/ita/ita-manuscript/ita_13_02.pdf

Mager, A. (2014). Defining algorithmic ideology: Using ideology critique to scrutinize corporate search engines. *TripleC*, *12*(1), 28–39.

Malaga, R. A. (2008). Worst practices in search engine optimization. *Communications of the ACM*, *51*(12), 147–150. https://doi.org/10.1145/1409360.1409388

Manovich, L. (2005). Understanding Meta-Media. *CTHEORY*, *td020*(October 25). http://www.ctheory.net/articles.aspx?id=493

Marino, M. C. (2006, December). Critical Code Studies. *Electronic Book Review*.

McCombs, M. E., & Shaw, D. L. (1972). The Agenda-Setting Function of Mass Medi. *The Public Opinion Quarterly*, *36*(2), 176–187.

Messing, S., & Westwood, S. J. (2012). Selective Exposure in the Age of Social Media: Endorsements Trump Partisan Source Affiliation When Selecting News Online. *Communication Research*, 0093650212466406-. https://doi.org/10.1177/0093650212466406

Michael, A. & Salter, B. (2008). *Marketing through search optimization: How people search and how to be found on the web*. Elsevier.

Moran, M., & Hunt, B. (2015). *Search engine marketing, Inc. : Driving search traffic to your company's website*. IBM Press.

Mordecai, A. (2014, July 17). What tools does Upworthy employ to test its headlines ? *Quora*.

Morley, D. (2007). *Media, Modernity, and Technology: The Geography of the New*. Routledge.

Nahin, P. J. (2013). *The Logician and the Engineer: How George Boole and Claude Shannon Created the Information Age*. Princeton University Press.

Napoli, P. M. (2014). Automated media: An institutional theory perspective on algorithmic media production and consumption. *Communication Theory*, *24*, 340–360. https://doi.org/10.1111/comt.12039

Noble, S. (2013). *Google Search: Hyper-visibility as a Means of Rendering Black Women and Girls Invisible | InVisible Culture*. http://ivc.lib.rochester.edu/portfolio/google-search-hyper-visibility-as-a-means-of-rendering-black-women-and-girls-invisible/

Odden, L. (2012). *Optimize: How to attract and engage more customers by integrating SEO, social media, and content marketing*. John Wiley & Sons.

Owens, T. (2015). *Designing Online Communities*. Peter Lang.

Pan, B., Hembrooke, H., Joachims, T., Lorigo, L., Gay, G., & Granka, L. (2007). In Google we trust: Users' decisions on rank, position, and relevance. *Journal of Computer-Mediated Communication*, *12*, 801–823. https://doi.org/10.1111/j.1083-6101.2007.00351.x

Parikka, J. (2011). Operative Media Archaeology: Wolfgang Ernst's Materialist Media Diagrammatics. *Theory, Culture & Society*, *28*(5), 52–74. https://doi.org/10.1177/0263276411411496

Parikka, Jussi. (2012). *What is Media Archaeology*. Polity Press.

Park, D. W., Jankowski, N., & Jones, S. (2011). *The Long History of New Media: Technology, Historiography, and Contextualizing Newness*. Peter Lang.

Prior, L. (2003). *Using Documents in Social Research*. Sage Publications.

Purcell, K., Brenner, J., & Rainie, L. (2012). Search Engine Use 2012. *PEW Research Center*, 42.

Ranganathan, S. R. (1973). *Philosophy of Library Classification*. Sarada Ranganathan Endowment for Library Science.

Rayson, S. (2013, August 17). The Social Media Optimization (SMO) of SEO: 7 Key Steps. *Social Media Today*.

Redish, J. (2014). *Letting go of the words*. Morgan Kaufmann.

Rieder, B. (2012). What is in PageRank? A Historical and Conceptual Investigation of a Recursive Status Index. : Computational Culture. *Computational Culture: A Journal of Software Studies*, *2*. http://computationalculture.net/article/what_is_in_pagerank

Rogers, E. M. (1997). *History Of Communication Study*. Free Press.

Rowles, D. (2018). *Digital Branding: A Complete Step-by-step Guide to Strategies, Tactics, Tools, and Measurement* (2nd ed.). Kogan Page.

Savetz, K. (1993). Life Before (And After) Archie. *Internet Business Journal*.

Schröter, J. (2012). The internet and "frictionless capitalism." *TripleC*, *10*(2), 302–312.

Scott, J. (1990). *A Matter of Record*. Polity Press.

Shannon, C. E. (1948). A mathematical theory of communication. *The Bell System Technical Journal*, *27*(July 1928), 379–423. https://doi.org/10.1145/584091.584093

Shaw, A. (2017). Encoding and Decoding Affordances: Stuart Hall and Interactive Media Technologies. *Media, Culture & Society,* 39(4), 592-602.

Shenoy, A., & Prabhu, A. (2016). *Introducing SEO : Your quick-start guide to effective SEO practices*. A Press.

Sherrod, J. (2010, September 8). SEO for Bing—Google and Bing Indexing Differences. *Search Discovery*.

Shreves, R., & Krasniak, M. (2015). *Social Media Optimization for Dummies*. John Wiley & Sons, Incorporated. http://ebookcentral.proquest.com/lib/csu/detail.action?docID=1895251

Silver, D. (2004). Internet/Cyberculture/ Digital Culture/New Media/ Fill-in-the-Blank Studies. *New Media & Society*, *6*(1), 55–64. https://doi.org/10.1177/1461444804039915

Sizov, S. (2010). Geofolk: Latent spatial semantics in web 2.0 social media. *WSDM '10 Proceedings of the Third ACM International Conference on Web Search and Data Mining*, 281–290. http://dx.doi.org/10.1145/1718487.1718522

Smarty, A. (2009, August 25). SEO Differences Between Google and Bing. *Search Engine Journal*.

Smith, M. Y. (1981). The method of history. In G. H. Stempel & B. H. Westley (Eds.), *Research Methods in Mass Communication* (pp. 305–319). Prentice-Hall.

Startt, J., & Sloan, Wm. D. (1989). *Historical Methods in Mass Communication*. Lawrence Erlbaum Associates Publishers.

*State of the News Media 2015*. (2015).

Sterne, J. (2005). Digital Media and Disciplinarity. *The Information Society*, *21*, 249–256. https://doi.org/10.1080/01972240591007562

Sterne, J. (2012). *The Meaning of a Format: MP3*. Duke University Press.

Sullivan, D. (2004, June 14). Who Invented the Term "Search Engine Optimization"? *Search Engine Watch Forums*.

Sutherland, J. W. (1975). System Theoretic Limits on the Cybernetic Paradigm. *Behavioral Science*, *20*(3), 191–200.

Taigman, Y., Yang, M., Ranzato, M., & Wolf, L. (2014). DeepFace: Closing the Gap to Human-Level Performance in Face Verification. *Conference on Computer Vision and Pattern Recognition (CVPR)*. https://doi.org/10.1109/CVPR.2014.220

Tanguay, D. (2009, April 8). Search the Rainbow. *Official Google Blog*.

Tatum, C. (2005). A breach of symbolic power or just a goofy prank? In *First Monday* (Vol. 10, Issue 10). Ghosh, Rishab Aiyer.

*Technology for Librarians 101: Anatomy of a Web Address*. (2014). Nebraska Broadband and Planning Initiative. https://broadband.nebraska.gov/documents/library/1%20AnatomyURL.pdf

Terranova, T. (2004). Communication Beyond Meaning: On the Cultural Politics of Information. *Social Text*, *22*(3), 51–73.

*The Differences Between Google & Bing SEO Algorithms*. (2014). Orange Patch. http://patch.com/connecticut/orange/the-differences-between-google--bing-seo-algorithms

*The size of the World Wide Web (The Internet)*. (n.d.). WorldWideWebSize.Com. Retrieved January 26, 2015, from http://www.worldwidewebsize.com/

*Timeline of Google Search*. (n.d.). Wikipedia. Retrieved January 27, 2015, from http://en.wikipedia.org/wiki/Timeline_of_Google_Search

Tofel, B. (2007). Wayback for accessing web archives. *Proceedings of the 7th International Web Archiving Workshop*, 27–37.

Vismann, C. (2008). *Files: Law and media technology* (G. (trans.) Winthrop-Young, Ed.). Stanford University Press.

W3C Schools. (n.d.). *HTML History*. https://www.w3schools.in/html-tutorial/history/

Wang, D. Y., Mohammad, M. Der, Saul, L., Mccoy, D., Savage, S., & Voelker, G. M. (2014). Search + Seizure: The Effectiveness of Interventions on SEO Campaigns. *Proceedings of the 2014 Conference on Internet Measurement Conference (IMC '14)*, 359–372. https://doi.org/10.1145/2663716.2663738

Wang, D. Y., Savage, S., & Voelker, G. M. (2011). Cloak and Dagger: Dynamics of Web Search Cloaking. *Proceedings of the 18th ACM Conference on Computer and Communications Security (CCS '11)*, 477–489. https://doi.org/10.1145/2046707.2046763

West, A. W. (2012). Search Engine Optimization. In *Practical HTML5 Projects*.

White, D. M. (1950). The "Gate Keeper": A Case Study in the Selection of News. *Journalism Quarterly*, *27*, 383–391.

Wiener, N. (1961). *Cybernetics, Or, Control and Communication in the Animal and the Machine*. M.I.T. Press.

William, R. (1961). *The Long Revolution.* Columbia University Press.

Williams, R. (1975). The Technology and the Society. In *Television: Technology and Cultural Form*. Schocken Books.

Winthrop-Young, Geoffrey, & Wutz, M. (1999). Translators' Introduction: Friedrich Kittler and Media Discourse Analysis. In *Gramophone, Film, Typewriter* (pp. xi–xxxviii). Stanford University Press.

Yalçın, N., & Köse, U. (2010). What is search engine optimization: SEO? *Procedia - Social and Behavioral Sciences*, *9*(July 2009), 487–493. https://doi.org/10.1016/j.sbspro.2010.12.185

Zeckman, A. (2014, July 11). Organic Search Accounts for Up to 64 % of Website Traffic. *Search Engine Watch*.