Module 3: Social Media Text, Action & Hyperlink Analytics

Syllabus:

Lecture no	Content	Duration (Hr)	Self-Study (Hrs)
1	Social Media Text Analytics- Types of Social Media Text	1	1
2	Purpose of Text Analytics	1	1
3	Steps in Text Analytics	1	1
4	Social Media Text 8 Analysis Tools Social Media Action Analytics	1	1
5	What Is Actions Analytics?	1	1
6	Common Social Media Actions	1	1
7	Actions Analytics Tools Social Media Hyperlink Analytics	1	1
8	Types of Hyperlinks, Types of Hyperlink Analytics	1	1
9	Hyperlink Analytics Tools	1	1

Theoretical Background:

TEXT IS one of the fundamental elements of social media platforms. Textual elements of social media include comments, tweets, blog posts, product reviews, and status updates. Social media text analytics, also known as text mining, is a technique to extract, analyze, and interpret hidden business insights from textual elements of social media content.

ACTIONS ARE the cash cow of social media. It is what the users do on social media that matters most to social media marketers.

HYPERLINKS ARE the pathways of social media traffic. Hyperlinks are references to web resources (such as a website, document, and files) that users can access by clicking on it.

Key Definitions:

Social media network analytics is the process of collecting, analyzing, and interpreting data from social media platforms to gain insights into user behavior, engagement, and trends.

Lecture 1

TYPES OF SOCIAL MEDIA TEXT

Based on its nature, social media text can be broadly classified into two categories:
1) dynamic text and 2) static text.

DYNAMIC TEXT

Dynamic text is a real-time social media user-generated text or statement to express an opinion about content or information posted over social media. Dynamic text is mostly posed by social media users in response to social, political, economic, personal, cultural, or business issues to express their views and feelings related to it. Dynamic text is usually smaller in length (e.g., a couple of sentences), diverse in nature, and is updated or deleted more frequently.

Examples of dynamic social media text include tweets, Facebook comments, and product reviews. Below, we briefly explain the two most common dynamic social media texts: tweets and comments.

Tweet

A tweet is a one hundred forty-character message posted by a Twitter user. A tweet may include text, images, video, or links to other websites. A tweet may also include a hashtag (#). Hashtags are used to mark keywords or topics in a tweet, and are organically created by Twitter users as a method to categorize messages. A keyword marked by a hashtag can easily appear in Twitter search, and popular hashtags are often trending topics over Twitter. Tweets accumulate over time, carry a timestamp and user information, and mostly appear in descending order; that is, the most recent first. Tweet data provides a valuable source for mining value business insights, including exploring trending topics, measuring brand sentiment, and gathering feedback on new products and services.

Comments

Social media comments are written (usually short) statements that express opinions about content or information posted over social media. While most comments are text only, it can also include images, video, or links to other websites. The ability to post comments and participate in social media discourse is the underlying characteristic that distinguishes social media from traditional media (e.g., TV and print). Like tweets, social media comments are also a great source for mining valuable business insights from social media. Almost all social media platforms provide commenting features. Comments accumulate over time, carry a time stamp and user information, and mostly appear in descending order; that is, the most recent first.

Discussion

Discussion takes the form of textual or written conversation or debate about a certain topic, product, or service. Mostly, discussions among social media users happen through Internet forums. Internet discussion forums are treelike in structure; that is, a forum can contain a number of sub-forums focused on specific topics or threads. In these forums users can post questions and reply to questions posted by other users. Discussions accumulate over time, carry a time stamp and user information, and mostly appear in descending order; that is, the most recent first. Vault Network is an example of an Internet forum that focuses on online games.

Conversation

Social media textual conversation (also known as chatting) is an instant exchange of short written messages between two more people. Chatting usually takes a casual form and are carried out through dedicated messaging services/tools. A variety of messaging tools have been developed for textual conversation, including desktop-based (e.g., Skype); web-based (e.g., Google Hangouts and Facebook chat) and mobile-based (e.g., Viber). Note that these services are not only limited to textual conversation, but also support video and voice conversation. And now that all media are converging, most of the messaging services also come in desktop, mobile, and web forms. For example, Skype has both desktop and Smartphone versions. An important point to note here is that most of the social media textual conversation is private in nature and may not be subject to mining.

Reviews

Reviews are critical evaluations of a product or service performed by customer are mostly shorter when compared to formal reviews by experts. Reviews can include textual elements and ratings. ProductReview.com.au, for example, is a site devoted to product/service reviews and ratings submitted by customers. Product reviews can serve as an excellent source for mining customers' opinions and feelings about a product or service.

STATIC TEXT

Static social media text is usually large in length (e.g., several paragraphs) and is generated, updated, or deleted less frequently. Examples of static text include wiki content, a blog page, Word documents, corporate reports, electronic mail (e-mail), and news transcripts. At the highest level of abstraction, the purpose of static social media text is to inform, educate, and elaborate.

Lecture 2

PURPOSE OF TEXT ANALYTICS

Both dynamic and static text are subject to analytics. The following are some of
the objectives of social media text analytics for business intelligence purposes (Figure 4).

SENTIMENT ANALYSIS

Sentiment analysis analyzes and categorizes social media text as being positive, negative, or neutral. Social media sentiment analysis mostly focuses on dynamic text. The primary purpose of sentiment analysis is to determine how your customers feel about a particular product, service, or issue. For example, as a product manager, you might be interested to know how your customers on Twitter feel about a product/service that was recently launched. Analyzing your tweets or Facebook comments may provide an answer to your question. Using sentiment analysis, you may be able to extract the wordings of the comments and determine if they are positive, negative, or neutral. At the end of the chapter, several analytical tools are listed for semantic analysis. A **Semantria**.

Semantria is an example of a text sentiment analysis tool. It will go through the following steps to extract sentiments from a document:

Step 1: It breaks the document into its basic parts of speech, called POS tags, which identify the structural elements of a sentence (e.g. nouns, adjectives, later section in the chapter provides a step-by-step guide on analyzing social media text using verbs, and adverbs).

Step 2: Algorithms identify sentiment-bearing phrases like "terrible service" or "cool atmosphere."

Step 3: Each sentiment-bearing phrase earns a score based on a logarithmic scale ranging from negative ten to positive ten.

Step 4: Next, the scores are combined to determine the overall sentiment of the document or sentence. Document scores range between negative two and positive two. For example, to calculate the sentiment of a phrase such as "terrible service," Semantria uses search engine queries similar to the following:

"(Terrible service) near (good, wonderful, spectacular)"

"(Terrible service) near (bad, horrible, awful)"

Each result is added to a hit count; these are then combined using a mathematical operation called "log odds ratio" to determine the final score of a given phrase.

Semantria.

Semantria is an example of a text sentiment analysis tool. It will go through the following steps to extract sentiments from a document:

Step 1: It breaks the document into its basic parts of speech, called POS tags, which identify the structural elements of a sentence (e.g. nouns, adjectives,
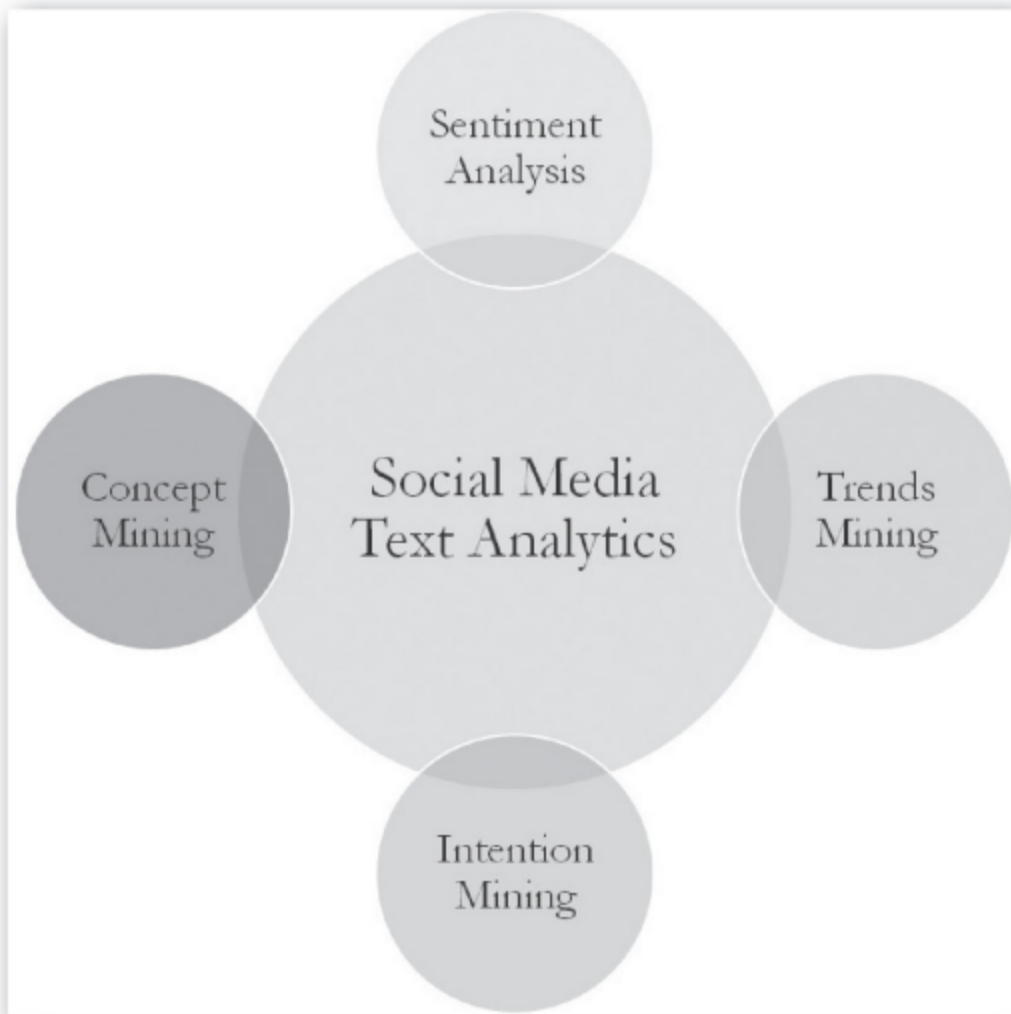
Figure 4. Purpose of social media text analytics

INTENTION MINING

Intention or intent mining (Chen, Lin et al. 2002) aims to discover users' intention (such as buy, sell, recommend, quit, desire, or wish) from natural language social media text such as user comments, product reviews, tweets, and blog posts. Social media as the integral part of our contemporary lives and is widely used by millions of customers to express desires, needs, and intention (Niven 2013). Companies may use intent mining to find new potential customers who intend to buy a product (or services) and service existing customers who have trouble with a product. For example, an analysis of company-related tweets may detect purchase intention based on the presence of the word "buy" or "purchase." Similarly, detecting the "quit" intention may identify and service the customers at risk of leaving the company. The Semantria analytical tool discussed later in this chapter, for example, can be used to mine intentions. customers at risk

of leaving the company. The Semantria analytical tool discussed later in this chapter, for example, can be used to mine intentions.

TRENDS MINING

Trends mining, also known as predictive analytics, uses huge amounts of historical and real-time social media data to predict future events. For example, a vast amount of social media data (e.g., comments and tweets) can be mined to identify patterns and trends for new product or service development or to improve customer satisfaction by anticipating their needs. Trend mining exploits patterns in large amounts of data by using sophisticated statistical techniques, including machine learning, data mining, and social network analysis. Predictive analysis using conventional business data has been used in a variety of domains, including marketing, banking, telecommunication, and healthcare. However, social media predictive analytics is still an emerging practice and may take some time for sophisticated tools and techniques to emerge.

CONCEPT MINING

Concept mining aims to extract ideas and concepts from documents. Unlike text mining, which is focused on extracting information, concept mining extracts ideas from large document sets. Thus, concept mining is useful in extracting ideas from large amounts of static social media text, such as wiki content, a web page, Word documents, and news transcripts. Concept mining can be employed to classify, cluster, and rank ideas.

Lecture 3

STEPS IN TEXT ANALYTICS

Text analytics, like any other form of social media analytics, is the art and science of getting the desired business intelligence from the text posted over social media (Figure 5). While the steps required for text analytics are largely dependent on the type of approach and tool employed, a typical social media text analysis includes the following cyclical steps.

IDENTIFICATION AND SEARCHING

The text analytics process starts with identifying the source of the text that will be analyzed. Text posted on social media is dynamic, huge, diverse, multilingual, and noisy. Thus, finding the right source for the purpose of text analytics is very crucial for gaining useful business insights. The genre of the source text also will determine the type of tool used to extract and analyze it. For example, extracting tweets requires different tools and approaches than analyzing a document or website text. Analyzing tweets, for example, requires API-based searching and extraction of data from the Twitter timeline based on criteria that you specify. You can choose to extract tweets that include specific keywords, such as your company name. The desired business question that needs to be answered with text analytics will serve as a good starting point.

TEXT PARSING AND FILTERING

The next step is to parse, clean, and filter the text, and create a dictionary of words using NPL, which is mostly based on machine learning techniques. In order for computer and algorithms to extract meanings from the text, the sentence structures and parts of speech are determined, named entities extracted (people, organizations, product/service names, etc.), stop words

removed, and spellings are checked. Most of these steps are automatic; however, in certain stages, human intervention is required. For example, in the filtering stage, manually cleaning (by humans with domain expertise) may be required to remove unwanted or irrelevant terms.

TEXT TRANSFORMATION

For analytical algorithms to be applied to the text, it should be transformed into a computer-readable format (e.g., 0s and 1s) for analysis. The cleaned text is thus transformed into numerical representations using linear algebra-based techniques, such as latent semantic analysis and vector space models.

TEXT MINING

At this step, the text is actually mined to extract the needed business insights. Varieties of text mining algorithms are applied to the text, such as clustering, association, classification, and predictive analysis, and sentiment analysis. Text analysis employs these sophisticated algorithms to extract sentiment and meanings from the text in a similar manner to the way human do; however, the process is thousands of times faster.

Association—Association or association mining is a data-mining technique used to determine the probability of the cooccurrence of items in a collection of documents. The relationships between cooccurring items are expressed as association rules. In text analytics, for example, social media text can be clustered together based on cooccurrence frequency. Or it can be used, for example, to find that a user who liked a social media content A and B is 90 percent likely to also like content C.

Clustering—Clustering or cluster analysis groups objects based on similarity in non-overlapping groups. Clustering is an important part of data mining and text analytics. Social media text (such as tweets or comments), for example, can be clustered into positive, negative, and natural categories. Or nodes in a social media network can be clustered based on importance.

Classification—From the text analytics perspective, classification or categorization is used to find similarities in the document and groups them with predefined labels based on the themes contained in the document (Chakraborty, Pagolu et al. 2013). For example, an e-mail can be classified as spam based on its contents.
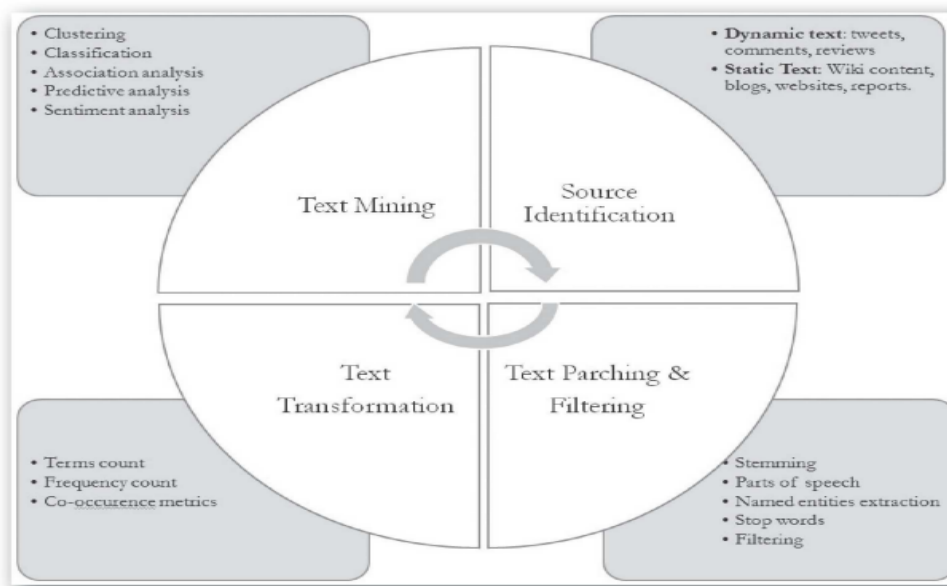
Figure 5. Steps in text analytics

Lecture 4

SOCIAL MEDIA TEXT ANALYSIS TOOLS

A variety of social media text analysis tools are available on the market. Some are free and others are paid. Below we list some popular text analysis tools. Discovertext: Discovertext (http://discovertext.com/) is a powerful platform for collecting, cleaning, and analyzing text and social media data streams.

Lexalytics: Lexalytics (http://www.lexalytics.com/) is a social media text and semantic analysis tool for social media platforms, including Twitter, Facebook, blogs, etc.

Tweet Archivist: Tweet Archivist (https://www.tweetarchivist.com/) is focused on searching, archiving, analyzing, and visualizing tweets based on a search term or hashtag (#).

Twitonomy: Twitonomy (https://www.twitonomy.com/) is a Twitter analytics tool for getting detailed and visual analytics on tweets, retweets, replies, mentions, hashtags, followers, etc.

Netlytic: Netlytic (https://netlytic.org) is a cloud-based text and social network analytics platform for social media text that discovers social networks from online conversations on social media sites.

LIWC: Linguistic Inquiry and Word Count (LIWC) is a text analysis tool for analyzing emotional, cognitive, structural, and process components present in individuals' verbal and written speech samples: http://www.liwc.net/

Voyant: Voyant (http://voyant-tools.org/) is a web-based text reading and analysis. With Voyant, a body of text can be read from a file or directly exported from a website.

SOCIAL MEDIA ACTIONS ANALYTICS

SOCIAL MEDIA ACTIONS ANALYTICS ACTIONS
ARE the cash cow of social media. It is what the users do on social media that matters most to social media marketers. Typical actions performed by social media users include likes, dislikes, shares, views, clicks, tags, mentions, recommendations, and endorsements. Actions are ways to express symbolic reactions to social media content. Symbolic actions are an easy and fast way to express feelings, unlike written reactions in the form of textual comments. Actions are not just symbolic reactions; they carry emotions and behaviors that can be harnessed. More importantly, social media actions are social expressions; that is, a user who performs an action (e.g., liking certain content) is visible to (or shared with) other social media users, in particular with their friends. This shareable nature of social media actions makes it very attractive to social media marketers and businesses. Take as an example Moviefone (an American-based movie listing and information service company), which enabled logins with Facebook and Twitter credentials. Enabling such login services not only allow users to use the Moviefone service conveniently, but also let them connect with their social media friends and share content over the Moviefone site. Enabling social logins led to a 300 percent increase in site traffic, a 40,000 to 250,000 increase in referrals per month, and a 40 percent increase in click-through rate

Lecture 5

WHAT IS ACTIONS ANALYTICS?
Social media actions analytics deals with extraction, analysis, and interpretation of the insights contained in the actions performed by social media users. Social media actions are of great value to social media marketers because of their role in increasing revenue, brand value, and loyalty. Organizations can employ actions analytics to measure popularity and influence of a product, service, or idea over social media. For example, a brand marketer might be interested to know how popular their new product is among social media users. Analyzing your Facebook likes and Twitter mentions, for example, may provide an answer to your questions.

Lecture 6

COMMON SOCIAL MEDIA ACTIONS
Below, we briefly discuss some of the most prevalent social media actions. All these actions are performed by social media users and can become your social media metrics. Metrics, in simple words, are anything you want to measure. Social media users can come in many forms, including followers, fans, and subscribers.

LIKE
Like or "Like" buttons or like options are a feature of social media sites (e.g., social networks, blogs, and websites) that allow users to express their feelings of liking certain products, services,

people, ideas, information, places, or content. They are actions performed by social media users to express symbolic positive reaction to social media content. Facebook's "Like" button enables users to easily express their feelings and give your product or service a virtual thumbs up. Incorporating a "Like" button in social media platforms and websites is becoming a norm. Social media platforms display accumulated likes received by content over time. Facebook's "Like" button is the most famous one. Google+ social networking platform uses a "+1" symbol to express liking. Companies use Google+ and Facebook fan pages to receive likes from customers, but the "Like" button can also be incorporated into a company website or blog. The "Like" button can be easily incorporated into a website as follows: 1. Visit Facebook's developers' page: https://developers.facebook.com/docs/plugins/like-button 2. Customize and generate a code. 3. Paste the code into your website after the tag.

DISLIKE

"Dislike" buttons are included in some social media platforms (e.g., YouTube) and allow users to express their negative feelings of disliking certain content (e.g., products, services, people, ideas, information, or places) posted over social media. Similar to the "Like" feature, it is visible to others and accumulated over time. The "Dislike" button is not as prevalent as the "Like" button.

SHARE

Share or "Share" button or sharing is a feature that allows social media users to distribute the content posted over social media to other users. For example, the Facebook "Share" button lets users add a personal message and customize who they share the content with. The WordPress (a blogging platform) "Share" button, for example, allows users to share their blog content across a range of social media platforms. Companies incorporate share buttons into website to boost their website traffic by channeling visitors from social media sites.

VISITORS, VISITS, REVISITS

A visitor is a person who visits your website or blog. A single visitor may visit a page or content one or more times (revisits). Visits are also known as sessions. Other related concepts are: Unique visitor—A person who arrives at your page first time. Average bounce rate—the percentage of visitors who visit a website and leave the site quickly without viewing other pages. Session duration—The average duration of a visit or session. VIEW

Views are the number of times social media content (a post, video, graphic, etc.) is viewed by users. A slightly different but related concept is page views, which is each time a visitor views a page on your company website or blog.

CLICKS

Clicks are the actions performed by users by pressing or clicking on the hyperlinked content of your website or blog. Through clicks, users navigate the web. Click data can be harvested for business intelligence purposes, such as, to reduce bounce rate and improve website traffic. A technique called clickstream analysis is used by business managers for a variety of business intelligence purposes, including website activity, website design analysis, path optimization, market research, and finding ways to improve visitor experience on the website. The clickstream

is the semistructured data trail/log (such as date and time stamp, IP address, and the URLs of the pages visited) a user leaves while visiting a website.

TAGGING

Tagging is the act of assigning or linking extra pieces of information to social media content (such as photographs and bookmarks) for identification, classification, and search purposes. Tagging lets user classify social media content the way they see it. Tagging may take a variety of forms. For example, bloggers can attach descriptive keywords (tags) to their posts to facilitate classification and searching of content, and Facebook users can add tags to anything they post on their status, including photos and comments. Social bookmarking services (such as del.icio.us) let users organize their bookmarks flexibly by adding descriptive tags. This practice of collaborative tagging is commonly known as folksonomy—a term coined by Thomas Vander Wal (Wal 2005). These days, almost all prominent companies (e.g., Facebook and Flickr) provide tagging services to their users. Because the contents are tagged with useful keywords, social tagging expedites the process of searching and finding relevant content.

MENTIONS

Mentions or social mentions are the occurrence of a person, place, or thing over social media by name. For example, a brand name maybe mentioned in a Facebook comment, blog post, YouTube video, or tweet. Mentions are important and can indicate popularity of person, place, or thing. For example, a social marketer may be able to gauge the popularity of a product/service/campaign by mining Twitter mentions data. A Twitter mention is the inclusion of a "@username" in a tweet.

HOVERING

Hovering is the act of moving a cursor over social media content. Capturing users' cursor movement data can help you understand user behavior on a social media site. Cursor movement/hovering over an ad, for example, can be considered as a proxy for attention. Most people who view an ad do not necessary click on it, thus if we are relying on clicks analytics only, we may lose a vital piece of information (i.e., attention). Studies have even suggested a strong correlation between hover time and purchases. Traditionally, hovering data has been used in website design and for improvement of user experience.

CHECK-IN

Check-in is a social media feature that allows users to announce and share their arrival at a location, such as a hotel, airport, city, or store. Many social media services, including Facebook and Google+, provide check-in features. The location of the user is determined using GPS (global positioning system) technology. Check-in data can, for example, be mined to offer location-based services/products.

PINNING

Pinning is an action performed by social media users to pin and share interesting content (such as ideas, products, services, and information) using a virtual pinboard platform. Some famous pinning platforms include Pinterest, Tumblr, StumbleUpon, or Digg. Business can use these virtual pin boards to share information and connect with and inspire their customers. Four

Seasons Hotels and Resorts, for example, use Pinterest to curate travel, food, and luxury lifestyle content to inspire customers.

EMBEDS

Embedding is the act of incorporating social media content (e.g., a link, video, or presentation) into a website or blog. An embed feature lets users embed interesting content into their personal social media outlets.

ENDORSEMENT

Endorsement is a features of social media that lets people endorse and approve other people, products, and services. For example, LinkedIn lets user endorse the skills and qualifications of other people in their network.

UPLOADING AND DOWNLOADING

In simple words, uploading is the act of adding new content (e.g., texts, photos, and videos) to a social media platform. The opposite of uploading is downloading; that is, the act of receiving data from a social media platform. All most all social media content is created and uploaded by users, which is better known as user-generated content. For some companies, uploading and downloading is the single most important action to measure. For Instagram and Flickr, which are both photo-sharing platforms, the number of photos uploaded daily matters more than anything else.

Lecture 7

ACTIONS ANALYTICS TOOLS

Currently, there is no single platform that can capture all the actions discussed in this chapter. Certain platforms can be employed to measure social media actions across platforms. Below we list some popular actions analytics tools.

Hootsuite: Hootsuite is an easy-to-use online platform that enables you to manage your social media presence across the most popular social networks. Hootsuite offers different plans depending on your business needs and budget: free, pro, or enterprise. In this tutorial, we will employ the free version, which supports up to five social media profiles and has limited analytics information.

SocialMediaMineR: SocialMediaMineR is a social media analytics tool that takes one or multiple URLs and returns the information about the popularity and reach of the URL(s) on social media, including the number of shares, likes, tweets, pins, and hits on Facebook, Twitter, Pinterest, StumbleUpon, LinkedIn, and Reddit. The tool can accessed from here: http://cran.rproject.org/web/packages/SocialMediaMineR/index.html

Lithium: Lithium (http://www.lithium.com/) is social media management tool that provides a variety of products and services, including social media analytics, marketing, crowd-sourcing, and social media marketing.

Google Analytics: Google Analytics (http://www.google.com/analytics/) is an analytical tool offered by Google to track and analyze website traffic. It can also be used to for blogs and wiki analytics.

Facebook Insights: Facebook Insights (https://www.facebook.com/insights/) helps Facebook page owners understand and analyze trends within user growth and demographics.

Klout: Klout (https://klout.com/) measures your influence across a range of social media channels based on how many people interact with your posts. Your Klout score measures your influence on a scale from one to one hundred.

Topsy: Topsy (http://topsy.com/) is similar to Icerocket and Social Mention, with its main focus around social media, especially multimedia sites and blogs.

Tweetreach: This tool helps you measure the number of impressions and reach of hashtags. The tool can be accessed here: https://tweetreach.com

Kred: Kred helps you measure the influence of a Twitter account: www.kred.com Hashtagify: This tool measures the influence of hashtags: http://hashtagify.me

Twtrland: Twtrland is a social intelligence research tool (http://twtrland.com/) for analyzing and visualizing your social footprints.

Tweetstats: using your Twitter user name, Tweetstats graphs Twitter stats including tweets per hour, tweets per month, tweet timelines, and reply statistics (http://www.tweetstats.com).

Lecture 8

SOCIAL MEDIA HYPERLINK ANALYTICS HYPERLINKS

ARE the pathways of social media traffic. Hyperlinks are references to web resources (such as a website, document, and files) that users can access by clicking on it. Hyperlinks can link resources within a document (inter-linking) and among documents (intralinking). For example, clicking on a hyperlink in a tweet can link you to other resources (e.g., websites) available over the Internet. Hyperlinks are not merely technical links between two websites, but serve a more symbolic means (Park 2003; Kim and Nam 2012). As a website is an official and unique entity representing an organization itself (Garrido 2003); therefore, embedding hyperlinks in an organization's website can be considered an official act of communication between two organizations. Hyperlinks among websites represent not only a reasonable approximation of a social relationship (Jackson 1997), but also serve as a symbolic meaning of validating or endorsing the linked organization (Vreelnad 2000). In conjunction with this, these hyperlinks that exist between two organizational websites reflect a sense of validation, trust, bonding, authority, and legitimacy (Vreelnad 2000; Park 2003; Nam, Barnett et al. 2014). Websites mostly connect or link to other websites of similar nature, so hyperlinks can also serve as indicators of content similarity (Chakrabarti, Joshi et al. 2002).

TYPES OF HYPERLINKS

From a hyperlink analytics point of view, mainly there are three types of hyperlinks, 1) in-links, 2) out-links, and 3) co-links.

IN-LINKS

In-links are the incoming hyperlinks or links directed toward a website or originated in other websites (Björneborn and Ingwersen 2004). For example, consider the top left image in the

Figure 9, page A is receiving two in-links coming from pages B and C. In-links are of great interest to social markers, because they bring traffic to a particular website. Thus, harvesting them can help us understand where the traffic to a corporate website is coming from. In-links also play an important role in website analytics, as both the quality and number of in-links can impact the search engine ranking of the website (more details on this are provided in the search analytics chapter). (Thelwall 2001) In-links can also impact the popularity of social media contents. A study on YouTube viral videos, for instance, found that among other things, in-links play crucial roles in the viral phenomenon, particularly in increasing views of videos posted on YouTube (Khan and Vong 2014). Studies have also shown that in-link counts strongly correlate with measures describing business performance (Vaughan 2004).

OUT-LINKS

Out-links are hyperlinks generated out of a website (Bjorneborn 2001). As shown in the top-right image in the Figure 9, page A is sending two out-links: one to page B and one to page C.

CO-LINKS

Co-links have two dimensions. First, if two websites receive a link from a third website, they are considered to be connected indirectly. For example, page A links to both pages B and C, therefore B and C are considered to be co-linking, or connected indirectly (bottom-left image in the Figure 9). Second, if two pages link to a third page, they are also considered to be colinking. As shown in the bottom-right corner of the Figure 9, Pages B and C are linking to page A; therefore, B and C are connected indirectly. Co-links have been used to compare and map competitive similarity among companies (Vaughan and You 2006). Figure 9. Different types of hyperlinks

HYPERLINK ANALYTICS

Hyperlink analytics deals with extracting, analyzing, and interpreting hyperlinks (e.g., in-links, out-links, and co-links). The basic assumption pertaining to hyperlink analytics is that the number and quality of hyperlinks to a website equates to its importance or value (Thelwall 2014). Hyperlink analytics can also reveal, for example, Internet traffic patterns and sources of the incoming or outgoing traffic to and from a website. Hyperlink analysis has been used to study a variety of topics including ranking of universities, understanding the blogosphere, scholarly websites (Vaughan and Thelwall 2003), and political networks (Park and Thelwall 2008), and to measure business competitiveness (Vaughan and You 2006). The case study included in this chapter demonstrates the importance of hyperlinks in viral phenomena and shows the valuable insights they carry for viral marketers in formulating viral marketing strategies. We must admit that, regardless of its importance, hyperlink analysis also has some limitations. Hyperlink networks, for example, does not provide any insight into the type or amount of traffic flowing among websites(Ackland 2010). When we talk about hyperlinks analytics, it mostly implies in-links, outlinks, and co-links analysis and does not include hyperlinks within a website between pages. Hyperlinks between pages within a website are created mostly for navigational purposes. Also, search engine ranking algorithms either ignore or give low importance to hyperlinks within a website (Thelwall 2014).

TYPES OF HYPERLINK ANALYTICS

Hyperlink analytics can take several forms, including: 1) hyperlink environment analysis, 2) link impact analysis, and 3) social media hyperlink analysis.

Hyperlink Environment Analysis

Hyperlink environment analyses deal with a particular website or set of websites. Hyperlinks (i.e., out-links, in-links, and co-links) of a website are extracted and analyzed to identify the sources of Internet traffic. Hyperlinks environment networks can take two forms: 1) co-links networks or 2) in-links and out-links networks.

Co-Link Networks

In co-links environment networks, nodes are websites and links that represent similarity between websites, as measured by co-link counts. With the Webometric Analyst tool, one can construct a co-link network diagram among a set of websites (Thelwall 2005; Thelwall 2014).

In-Links and Out-Links

Networks In-links and out-links hyperlink environment networks are constructed based on in-links and out-links from a website or set of websites. In such a network, nodes will be websites and links will present in-links and out-links. The tutorial provided in this chapter demonstrates constructs such as networks using the VOSON hyperlink analysis tool. Link Impact Analysis Link impact analysis investigates the web impact of a website address (or URL) in terms of citations or mentions it receives over the web. In a link impact analysis, statistics about web pages that mention the URL of a given website are collected and analyzed (Thelwall 2005; Thelwall 2014). The assumption is that a URL (or website address) cited frequently over the web is more important. Thus, measuring the web impact of URLs may provide an idea about the importance of a website. Social Media Hyperlink Analysis Social media hyperlink analysis deals with extraction and analysis of hyperlinks embedded within social media texts (.e.g., tweets and comments). These hyperlinks can be extracted and studied to identify the sources and destination of social media traffic. A good example of the usefulness of the hyperlink embedded in the social media text is the study by Khan et al. (2014), in which they extracted out-links from Korean and US government agencies' tweets. By extracting out-links and tracing them back to their sender, the authors were able to construct a map of the out-link structure (Figure 10). According to a comparison of out-links between tweets of the Korean and US governments, there were some differences in citation (i.e., out-link) patterns. The Korean government tended to cite domestic portals' news services and their own blogs (i.e., self-citation). Although there were SNSs and newspaper sites, most of the related out-links were for portals. On the other hand, the US government showed a more diverse pattern in terms of out-link destinations. US out-links were not concentrated in specific sites and tended to go directly to news agencies, not to secondary sources such as portals. These comparisons between the US and Korean governments suggest that social media out-links can carry valuable information and can help explain real-world phenomena and shed light on the disparities in social media use among different

cultures (Khan, Yoon et al. 2014). Figure 10. Twitter networks for Korea (left) and the US Governments (right)

Lecture 9

HYPERLINK ANALYTICS TOOLS

The following are some popular hyperlink analytics tools.

Webometric Analyst: Webometric Analyst is a web impact analysis tool and can conduct variety of analysis on social media platforms including hyperlink network analysis and web mentions: http://lexiurl.wlv.ac.uk/

VOSON: VOSON (http://www.uberlink.com/) is a hyperlink analytics tools for constructing and analyzing hyperlink networks. More details on VOSON are provided in the hyperlink analytics chapter. This chapter includes a detailed tutorial on using VOSON for hyperlink analysis.

Open Site Explorer: Open Site Explorer is a link analysis tool to research and compare competitor backlinks, identify top pages, view social activity data, and analyze anchor text: https://moz.com/researchtools/ose/

Link Diagnosis: Link Diagnosis (http://www.linkdiagnosis.com/) is a free online tool for analyzing and diagnosing links.

Advanced Link Manager: Advanced Link Manager provides a variety of link analysis capabilities, including the ability to track link-building progress over time, domain quality analysis, backlinks evolution, and website-crawling abilities: http://www.advancedlinkmanager.com/

Majestic: Majestic (https://majestic.com) provides a variety of link analysis tools, including link explorer, backlinks history, and link mapping tools.

Backlink Watch: Backlink Watch (http://backlinkwatch.com/) is a free tool for checking the quality and quantity of in-links pointing to a website.