

ارائه الگوی پیشنهادی برای نرم‌افزار استخراج داده از رسانه‌های اجتماعی

امیررضا اصنافی*^۱، مریم پاکدامن^۲، سمانه برجی^۳

بازیابی دانش و نظام‌های معنایی

سال هفتم، شماره ۲۵، زمستان ۹۹، ص ۳۱ تا ۳۱

نوع مقاله: مقاله پژوهشی

تاریخ دریافت: ۹۸/۱۲/۱۹

تاریخ پذیرش: ۹۹/۰۴/۰۷

چکیده

این پژوهش با هدف اصلی ارائه الگوی پیشنهادی برای نرم‌افزار استخراج و آرشیو داده از رسانه‌های اجتماعی تدوین شده است. در پژوهش حاضر ضمن مطالعه کتابخانه‌ای در خصوص استخراج داده و آرشیوسازی رسانه‌های اجتماعی به بررسی نرم‌افزارهای تولید شده در این زمینه پرداخته شده است؛ در نهایت براساس سیاه‌وارسی تهیه شده، یک الگوی پیشنهادی از این نوع نرم‌افزارها ارائه شد. این روند با همراهی متخصصان رسانه‌های اجتماعی در علم اطلاعات و دانش‌شناسی انجام گرفت. در این پژوهش، گردآوری اطلاعات، به روش مطالعه اسنادی و پیمایشی تحلیلی انجام شد. ابزار گردآوری اطلاعات پرسشنامه است. با نمونه‌گیری هدفمند و رصد افراد مطلع، تعداد ۳۱ نفر از متخصصان رسانه‌های اجتماعی در علم اطلاعات و دانش‌شناسی انتخاب شدند. قابلیت‌های به‌دست آمده از نرم‌افزارها در قالب طیف لیکرت در اختیار متخصصان قرار گرفت و از آن‌ها خواسته شد که میزان موافقت خود را با ضرورت وجود هر یک از این قابلیت‌ها مشخص نمایند. برای بررسی بودن یا نبودن هر یک از قابلیت‌های نرم‌افزارها از روش سی‌وی‌آر^۴، بار عاملی اکتشافی و بار عاملی تأییدی استفاده شد. میزان پذیرش هر یک از قابلیت‌ها، با میانه ارزیابی (معیار سنجش میانه^۳) شد. سپس، هر یک از گویه‌ها بر اساس میانگین موجود در آزمون فریدمن رتبه‌بندی شدند. در مرحله دوم سی‌وی‌آر هر یک از قابلیت‌ها، و در مرحله آخر بار عاملی تأییدی و بار عاملی اکتشافی بررسی شد. پرسش‌های پژوهش با استفاده از آزمون تی‌تک نمونه‌ای بررسی شدند. قابلیت‌های «امکان جستجوی پیشرفته آرشیو»، «دسته‌بندی آرشیوها (مرورآسان)» و «ارائه گزارش‌های بصری» نمره بالاتر از ۵۰ و معیارهای

۱. * استادیار گروه علم اطلاعات و دانش‌شناسی دانشگاه شهید بهشتی، aasnafi@gmail.com

۲. پژوهشگاه بین‌المللی زلزله‌شناسی و مهندسی زلزله، m.pakdaman@gmail.com

۳. کارشناس ارشد، علم اطلاعات و دانش‌شناسی دانشگاه شهید بهشتی، borji.samaneh@gmail.com

۴. CVR

«بازنشانی رکوردهای حذف شده»، «امکان ایجاد آرشیو باز برای دسترسی عام»، «اخذ گواهی‌های اعتباری از نهادهای دولتی و خصوصی»، «دریافت نسخه کامل پس از پرداخت هزینه» و «نسخه دمو» نمره کمتر از ۵۰ را در سی‌وی آر به دست آوردند. تمامی قابلیت‌ها بار عاملی تأییدی و اکتشافی بیشتر از ۰/۳ داشتند؛ بنابراین، در نظر گرفته شدند. اما با مقایسه بار عاملی قابلیت‌ها با سی‌وی آر امکان حذف قابلیت‌هایی با سی‌وی آر کمتر از ۵۰ وجود دارد. تمامی قابلیت‌های تخصصی نرم‌افزارهای فعلی از دیدگاه متخصصان ضروری هستند. قابلیت‌های «امکان جستجوی پیشرفته آرشیو»، «دسته‌بندی آرشیوها (مرور آسان)» و «ارائه گزارش‌های بصری» مهم‌ترین الزامات نرم‌افزارها از دیدگاه متخصصان هستند. قابلیت «اخذ گواهی‌های اعتباری از نهادهای دولتی و خصوصی» کمترین اهمیت را از دید متخصصان دارد. معیارهای «حذف خودکار آرشیوها در زمان تعریف شده»، «بازنشانی رکوردهای حذف شده»، «امکان ایجاد آرشیو باز برای دسترسی عام»، «اخذ گواهی‌های اعتباری از نهادهای دولتی و خصوصی»، «حذف خودکار آرشیوها در زمان تعریف شده»، «دریافت نسخه کامل پس از پرداخت هزینه» و «نسخه دمو» می‌توانند از قابلیت‌های نرم‌افزار حذف شوند. البته در بررسی نرم‌افزارهای فعلی استخراج داده از رسانه‌های اجتماعی، مشاهده شد که برخی از این قابلیت‌ها دارای اهمیت بالایی است. قابلیت «امکان ذخیره سازی محلی» توسط متخصصان به سیاهه‌وارسی نهایی اضافه گردید.

کلیدواژه‌ها: رسانه‌های اجتماعی، آرشیو رسانه‌های اجتماعی، استخراج داده از رسانه‌های اجتماعی

پژوهشگاه علوم انسانی و مطالعات فرهنگی
پرتال جامع علوم انسانی

مقدمه

به تعبیر مانوئل کاستلز^۱ (۲۰۰۷) فناوری‌های نوین اطلاعات، نقاط دور عالم را در شبکه‌های جهانی به همدیگر پیوند می‌دهند، ارتباطات رایانه‌ای مجموعه‌ای از جوامع مجازی را به وجود می‌آورد و در نتیجه آن همه ساختارها و فرایندهای مادی و معنوی بشری دگرگون می‌شود. در واقع به تناسب بهره‌مندی جوامع از فناوری‌های اطلاعاتی و ارتباطی و ایفای نقش در ارتباطات و تعامل بین مردم و تبادل اطلاعات و دسترسی جامعه به این فرایندها، تغییرات اجتماعی شتاب و شدت بیشتری می‌گیرد. از این‌رو، بیان اینکه بین فناوری‌های اطلاعاتی و ارتباطی و شکل‌گیری زندگی اجتماعی پیوندی استوار برقرار شده است، مفهومی اغراق‌آمیز نیست (گریپنبرگ^۲، ۲۰۰۵).

با نفوذ روزافزون رسانه‌های اجتماعی در زندگی افراد، مباحث درباره این رسانه‌ها در زمینه‌های مختلف از جمله علوم کامپیوتر و داده‌کاوی از اهمیت بسیار زیادی برخوردار شده است. رسانه‌های اجتماعی مرز بین دنیای واقعی و دنیای مجازی هستند. این رسانه‌ها افراد را قادر می‌سازند که در هر جا و هر زمانی بتوانند بایکدیگر در ارتباط باشند (ایروانچی‌زاده، ۱۳۹۴). این رسانه‌ها باعث فراهم کردن بستری هستند که در آن رفتار افراد را در مقیاسی بی‌سابقه، بدون هیچ حد و مرز جغرافیایی، با دید و رویکردهای جدید قابل مشاهده است (لیو^۳ و همکاران، ۲۰۱۴). در این میان رسانه‌های اجتماعی سهم بسزایی در انتقال، انتشار و ایجاد اطلاعات دارند، استخراج و آرشیو این اطلاعات از اهمیت ویژه‌ای برخوردار است. با استفاده از این داده‌ها می‌توان به دانش زیادی در زمینه تصمیم‌گیری، شخصی‌سازی، پیش‌بینی، سیستم‌های محاسباتی، سیستم‌های پیش‌گیری از بیماری و بسیاری از زمینه‌های دیگر دست‌یافت (روی^۴ و همکاران، ۲۰۱۵). داده‌های موجود در رسانه‌های اجتماعی قابلیت استخراج و آرشیوسازی دارند، بنابراین بهره‌مندی از یک نرم‌افزار قدرتمند استخراج داده از رسانه‌های اجتماعی بسیار مهم و حائز اهمیت است. این درحالی است که در دنیای مدیریت

1. Manuel Castells
2. Gripenberg
3. Liu
4. Roy

پیشینه‌ها، رسانه‌های اجتماعی هنوز جدید هستند. حتی برای ادارات ملی آرشیو پیشینه‌ها، چگونگی تعریف و تشخیص آنچه که موقتی یا ماندگار است و قالب فایل پیشینه‌های پیچیده، یک چالش به شمار می‌رود (آلن، ۲۰۱۸). در سال‌های اخیر نرم‌افزارهای گوناگونی با ویژگی‌های مختلف به منظور استخراج و آرشیوسازی داده‌ها از رسانه‌های اجتماعی ایجاد شدند؛ با این حال بررسی‌های محقق نشان داد در زمان انجام این پژوهش نرم‌افزار جامع این چنین در ایران ایجاد نشده است. این درحالی است که امروزه رسانه‌های اجتماعی سهم بسزایی در انتقال، انتشار و ایجاد اطلاعات دارند و بهره‌مندی از آرشیو آن‌ها در آینده از اهمیت ویژه‌ای برخوردار خواهد بود. آرشیوهای ایجاد شده از رسانه‌های اجتماعی توسط نرم‌افزار استخراج داده می‌تواند کمک بسزایی به سیاستمداران، دانشمندان، محققان و... کند. عدم دسترسی به آرشیوهای رسانه‌های اجتماعی ممکن است ضرر و زیان‌های اقتصادی، سیاسی، فرهنگی و... را به همراه داشته باشد. نرم‌افزارهای ایجاد آرشیو از رسانه‌های اجتماعی می‌توانند مانع بروز چنین مشکلاتی شوند. بهره‌مندی از الگوی ارائه شده در این پژوهش قابلیت‌های مورد نیاز برای ایجاد یک نرم‌افزار استخراج و آرشیو داده از رسانه‌های اجتماعی را مشخص می‌نماید.

همچنین به نظر می‌رسد تاکنون پژوهشی منسجم و کامل در خصوص قابلیت‌های نرم‌افزار استخراج و آرشیو داده از رسانه‌های اجتماعی در ایران انجام نگرفته است. همین‌طور پژوهشی به منظور تدوین الگو و سیاهه‌ای جامع و منسجمی برای قابلیت‌های تخصصی یک نرم‌افزار استخراج داده از رسانه‌های اجتماعی صورت نگرفته است. به این دلیل که سیاهه منسجمی برای قابلیت‌های این نوع نرم‌افزار وجود ندارد؛ ایجاد یکپارچگی و ارائه سیاهه‌ای کلی که بتواند تا حد امکان تمامی معیارها و قابلیت‌های تخصصی را تحت پوشش قرار دهد، مورد نیاز است. از این رو، پژوهش حاضر، با هدف ارائه الگو و سیاهه پیشنهادی برای نرم‌افزار کاربردی الگوی استخراج داده از رسانه‌های اجتماعی به بررسی قابلیت‌های تخصصی اینگونه

نرم‌افزارها پرداخته است. با اجرای الگوی پیشنهادی این پایان‌نامه می‌توان نیاز به استفاده از نمونه خارجی چنین نرم‌افزارهایی را کاهش داد.

اهداف پژوهش

پژوهش حاضر در نظر دارد ضمن بررسی متون در خصوص استخراج داده و آرشیوسازی رسانه‌های اجتماعی به عنوان ابزارهای جدید و تعاملی انتقال اطلاعات به بررسی برخی از نرم‌افزارهای این چنین پردازد؛ در نهایت نیز به ارائه یک الگوی مطلوب از این نوع نرم‌افزارها براساس سیاه‌وارسی تهیه شده به منظور استفاده کاربران داخلی رسانه‌های اجتماعی پردازد. هدف اصلی این پژوهش ارائه الگوی پیشنهادی برای نرم‌افزار استخراج و آرشیو داده‌های رسانه‌های اجتماعی است.

سایر اهداف پژوهش عبارت‌اند از:

۱. شناسایی ضرورت قابلیت‌های «مدیریت داده‌های آرشیوی» در نرم‌افزار استخراج داده از رسانه‌های اجتماعی، از نظر متخصصان
۲. شناسایی ضرورت قابلیت‌های «اعتباربخشی» در نرم‌افزار استخراج داده از رسانه‌های اجتماعی، از نظر متخصصان
۳. شناسایی ضرورت قابلیت‌های «داده‌کاوی» در نرم‌افزار استخراج داده از رسانه‌های اجتماعی، از نظر متخصصان
۴. شناسایی ضرورت قابلیت‌های «امنیتی» در نرم‌افزار استخراج داده از رسانه‌های اجتماعی، از نظر متخصصان
۵. شناسایی ضرورت قابلیت‌های «مدیریت کاربران» در نرم‌افزار استخراج داده از رسانه‌های اجتماعی، از نظر متخصصان
۶. شناسایی ضرورت قابلیت‌های «تعرفه مالی» در نرم‌افزار استخراج داده از رسانه‌های اجتماعی، از نظر متخصصان
۷. شناسایی ضرورت قابلیت‌های نرم‌افزارهای فعلی استخراج داده از رسانه‌های اجتماعی، از نظر متخصصان

سؤال‌های پژوهش

۱. از نظر متخصصان، قابلیت‌های «مدیریت داده‌های آرشیوی» در نرم‌افزار استخراج داده از رسانه‌های اجتماعی تا چه میزان ضرورت دارند؟
 ۲. از نظر متخصصان، قابلیت‌های «اعتباربخشی» در نرم‌افزار استخراج داده از رسانه‌های اجتماعی، تا چه میزان ضرورت دارند؟
 ۳. از نظر متخصصان، قابلیت‌های «داده‌کاوی» در نرم‌افزار استخراج داده از رسانه‌های اجتماعی، تا چه میزان ضرورت دارند؟
 ۴. از نظر متخصصان، قابلیت‌های «امنیتی» در نرم‌افزار استخراج داده از رسانه‌های اجتماعی، تا چه میزان ضرورت دارند؟
 ۵. از نظر متخصصان، قابلیت‌های «مدیریت کاربران» در نرم‌افزار استخراج داده از رسانه‌های اجتماعی، تا چه میزان ضرورت دارند؟
 ۶. از نظر متخصصان، قابلیت‌های «تعرفه مالی» در نرم‌افزار استخراج داده از رسانه‌های اجتماعی، تا چه میزان ضرورت دارند؟
 ۷. از نظر متخصصان، قابلیت‌های فعلی در نرم‌افزارهای استخراج داده از رسانه‌های اجتماعی تا چه میزان ضرورت دارند؟
 ۸. نرم‌افزار کاربردی الگو برای استخراج و آرشیو رسانه‌های اجتماعی از چه قابلیت‌های تخصصی‌ای برخوردار است؟
- اصنافی (۱۳۹۵) پژوهش خود را با عنوان «تأملی بر آرشیوسازی رسانه‌های اجتماعی» نگاشته است. نتایج این پژوهش نشان داد رسانه‌های اجتماعی خروجی فراتر از متن دارند، آن‌ها به محتوای گوناگون، نظیر وب سایت‌ها، عکس‌ها، ویدئوها و سایر رسانه‌ها پیوند می‌دهند. بنابراین، باید برنامه ویژه‌ای برای آرشیوسازی محتوای این منابع داشت. از یک سو با توجه به چالش‌هایی که در برابر محتوای تولیدشده در محیط وب و به خصوص در رسانه‌های استاندارد جهت آرشیوسازی اجتماعی وجود دارد و از سوی دیگر فقدان نرم‌افزار محتوای این منابع در داخل کشور، پیشنهاد می‌شود صاحبان محتوا در بازه‌های زمانی

مشخص، نسبت به تهیه نسخه‌های پشتیبان از مطالب تولیدشده اقدام نمایند تا در صورت وبلاگ، بتوان از یک رسانه اجتماعی نظیر مواجه شدن با آسیب‌های احتمالی سرور اصلی وجود نسخه آرشیو آن اطمینان حاصل نمود. در نهایت باید در نظر داشت که آرشیو کردن، رسالت ثبت حافظه مؤسسات باشد. آرشیویست‌ها، کتابداران، پژوهشگران، ناشران باید با یکدیگر برای توسعه مدل‌های جدید و ابزارهای اختصاصی آرشیو کردن رسانه‌های اجتماعی تلاش کنند.

ایروانچی (۱۳۹۴) پایان‌نامه خود را با عنوان «روشی مستقل از معنا برای استخراج کلمه کلیدی داده‌های متنی کاربران در یک رسانه اجتماعی» نگاشته است. این پژوهش به ارائه یک روش مستقل از معنی برای استخراج کلیدواژه از توییت می‌پردازد. روش‌های سنتی استخراج کلیدواژه معمولاً برای اسناد رسمی مانند مقالات خبری و علمی طراحی شده‌اند. اما این روش‌ها در داده‌های توییت که دارای چالش‌های مختلفی هستند، کارایی مناسبی ندارند. از جمله این چالش‌ها می‌توان به وجود کلمات هجو و بی‌معنی و محدودیت کاراکتر توییت‌ها اشاره نمود. این محدودیت کاربران را مجبور می‌کند که پیام خود را در جملات کوتاهی بیان کنند. به همین دلیل اکثر کلمات موجود در این توییت‌ها یک بار تکرار می‌شوند که باعث کاهش کارایی روش‌های مبتنی بر فرکانس سنتی می‌گردد. به‌منظور حل این چالش، روش ارائه شده در این پژوهش ترکیبی از دو روش مبتنی بر دیکشنری و مبتنی بر فرکانس است. روش مبتنی بر دیکشنری از تیر و متن خبر برای ساخت دیکشنری استفاده کرده است و روش مبتنی بر فرکانس از خوشه‌بندی توییت‌ها استفاده می‌کند. نتایج آزمایش‌ها نشان داد که روش ارائه شده روشی موفق و قوی برای استخراج کلیدواژه از متون کوتاه مانند توییت‌ها است.

آلن (۲۰۱۸) مدیر ارشد شاخه رسانه‌های اجتماعی در آرشیو ملی آمریکا در پژوهشی با عنوان «جمع‌آوری داده و بهره‌برداری از رسانه‌های اجتماعی» که در کنفرانس ایفلا ۲۰۱۸ ارائه داده است؛ می‌گوید: «در دنیای مدیریت پیشینه‌ها، رسانه‌های اجتماعی هنوز جدید

هستند. حتی برای ادارات ملی آرشیو پیشینه‌ها، چگونگی تعریف و تشخیص آنچه که موقتی یا ماندگار است و قالب فایل پیشینه‌های پیچیده، یک چالش به شمار می‌رود.» سازمان‌ها در درک با ارزش بودن داده‌هایشان بر روی رسانه‌های اجتماعی تردید دارند. در این ارائه گفته شد: «برای اولین بار با دولت باراک اوباما، داده‌های رسانه‌های اجتماعی توسط آرشیو ملی آمریکا حفظ شد. برای اطمینان از اینکه محتوای موجود از کاخ سفید بر روی رسانه‌های اجتماعی پس از انتقال در دسترس عموم قرار می‌گیرد، کارمندان بخش رسانه‌های اجتماعی از نزدیک با دفتر کتابخانه‌های ریاست جمهوری همکاری کردند. نهایتاً داده‌های کاخ سفید بر روی فلیکر، اینستاگرام، تامبلر، توییتر و حساب‌های یوتیوب استخراج و آرشیو شد. در سال ۲۰۱۷ رکوردهای آرشیو شده در یک پرتال عمومی^۱ بارگذاری شد.»

هارور^۲ و رمضان‌زاده هروی (۲۰۱۵) به مطالعه امکان‌سنجی برای آرشیوسازی رسانه‌های اجتماعی در ایرلند پرداختند. آن‌ها عقیده داشتند که بین رویدادها و هشتک‌ها رابطه وجود دارد و برای تحلیل رویدادها می‌توان از آن‌ها بهره برد. تمرکز کار آن‌ها بر روی یادداشت‌های توییتر بود. این پژوهشگران مدلی را برای آرشیو کردن محتوای رسانه‌های اجتماعی بر اساس فعالیت‌های آرشیوی و رویکردهای رایانشی ارائه نمودند. ۱۲ دانشگاه کانادا از نرم‌افزارهای آرشیوسازی وب برای نگهداری از محتوای رسانه‌های اجتماعی خود استفاده می‌کنند. محتوای تولیدشده توسط سازمان، تهیه مستندات و پیشینه سازمان، رویدادها، اخبار و مکاتبات از مواردی هستند که توسط این دانشگاه‌ها آرشیو می‌شود. کتابخانه دانشگاه یوبی سی برای آرشیوسازی رسانه‌های اجتماعی خود، از نرم‌افزار آرکایو ایت و ماشین وی‌بک در پایگاه آرشیو اینترنتی استفاده می‌کند (رینگهام^۳ و رومان آمیگو^۴، ۲۰۱۶).

1. <https://www.obamalibrary.gov/research/archived-white-house-websites-and-social-media>

2. Harrower

3. Ringham

4. Roman Amigo

لیثام^۱ (۲۰۱۴) آرشیویست اداره آرشیو استرلیا، ابزار و روش‌هایی در مواجهه با آرشیو رسانه‌های اجتماعی را دسته‌بندی کرد. این مقاله یک طیف از راهبردهای مدیریت اطلاعات را ارائه داده است که می‌توان آن را به رسانه‌های اجتماعی خود براساس نیازهای خاص و ریسک‌های مخصوص به خود اعمال کرد. راهبردهایی که انتخاب می‌شود باید براساس تحلیل و درک خطرها باشند. ممکن است نیاز به انتخاب راهبردهای متفاوت مدیریت اطلاعات برای هر کانال رسانه اجتماعی براساس سطوح مختلف ترافیک، خطر و بحث در هر کدا باشد. جدول تهیه شده توسط لیثام در پیوست «الف» قابل مشاهده است. آرشیو ملی ایالات متحده آمریکا در سال ۲۰۱۳ گزارشی با عنوان «بهترین نمونه‌ها برای ضبط پیشینه‌های رسانه‌های اجتماعی» منتشر کرد. در این گزارش اعلام شد: کاربست سیاست‌ها و فنون ضبط رکوردهای^۲ رسانه‌های اجتماعی فدرال در شاخه اجرایی^۳، دوران ابتدایی و آغازین خود را می‌گذرانند. ارگان‌ها روش‌های مختلفی را به منظور ضبط کردن محتوای ایجاد شده در پلت‌فرم‌های رسانه‌های اجتماعی به کار می‌گیرند. در سال ۲۰۱۲ دفتر ارشد ضبط سوابق و تیم رسانه‌های اجتماعی آرشیو ملی با اعضا و مدیران وب مدیریت سوابق فدرال، فروم^۴ برگزار کردند تا درباره بهترین نمونه‌های^۵ ضبط رسانه‌های اجتماعی بحث نمایند. شرکت‌کنندگان استفاده و تست را مدنظر دارند و شماری از ابزارها و فنون ذیل را تأیید کردند. کلی^۶ و دیگران (۲۰۱۰) در انگلستان به آرشیوسازی پست‌های تویتر پرداختند. آن‌ها برای مدیریت توییت‌ها و دسترس‌پذیری به داده‌ها نرم‌افزار تی‌واپر^۷ را استفاده کردند. آن‌ها از سال ۲۰۱۰ برای آرشیوسازی اطلاعات، از این نرم‌افزار استفاده می‌کنند. البته چالشی که در این بین وجود دارد این است که کاربران می‌توانند توییت‌های خود را حذف کنند؛ حتی آن‌هایی که آرشیو شده است. این حق را توییت‌ها به آن‌ها داده است که هر زمان که مایل

1. Latham
2. record
3. Executive Branch
4. forum
5. Best practices
6. kelly
7. TWAPER

بودند توییت‌های خود را حذف نمایند؛ بنابراین آرشیوسازی دائمی آن‌ها بدون اجازه مالک توییت‌ها ممکن نیست. آرشیو کردن توییت‌ها براساس کلیدواژه‌ها انجام می‌گردد که حتی ممکن است به جای یک موضوع، به پروفایل شخصی یک فرد هم مربوط شود. در کتابخانه‌های دانشگاه جرج واشنگتن، ابزاری ایجاد شد که از طریق آن امکان برداشت پست‌های عمومی فراهم بود. این نرم‌افزار رابط برنامه نویسی کاربردی برنامه توییترا^۱ نام داشت. در این برنامه، امکان جستجوی کاربران خاص، جستجوی کلیدواژه‌گانی و مکانی و دریافت خروجی به صورت اکسل برای انجام تحلیل‌های بیشتر نیز در این برنامه فراهم بود (کلی و دیگران، ۲۰۱۰).

بنابراین، با نفوذ روزافزون رسانه‌های اجتماعی در زندگی افراد، بحث درباره این رسانه‌ها در زمینه‌های مختلف از جمله علوم کامپیوتر و داده کاوی از اهمیت بسیار زیادی برخوردار است. رسانه‌های اجتماعی مرز بین دنیای واقعی و دنیای مجازی هستند. این رسانه‌ها افراد را قادر می‌سازند که در هر جا و هر زمانی بتوانند باهم در ارتباط باشند (ایروانچی زاده، ۱۳۹۴). این رسانه‌ها باعث فراهم کردن بستری هستند که در آن رفتار افراد را در مقیاسی بی‌سابقه، بدون هیچ حد و مرز جغرافیایی، با دید و رویکردهای جدید قابل مشاهده است (لیو^۲ و همکاران، ۲۰۱۴). در این میان رسانه‌های اجتماعی سهم بسزایی در انتقال، انتشار و ایجاد اطلاعات دارند، استخراج و آرشیو این اطلاعات از اهمیت ویژه‌ای برخوردار است. با استفاده از این داده‌ها می‌توان به دانش زیادی در زمینه تصمیم‌گیری، شخصی‌سازی، پیش‌بینی، سیستم‌های محاسباتی، سیستم‌های پیش‌گیری از بیماری و بسیاری از زمینه‌های دیگر دست‌یافت (روی^۳ و همکاران، ۲۰۱۵). آرشیویست‌ها غالباً واژه آرشیو را انتقال داده‌ها از یک منشأ حقیقی یا حقوقی به یک مخزن می‌دانند که بتوان آن داده‌ها را سازماندهی و سپس دسترس‌پذیر کرد؛ اما اصطلاح آرشیوسازی در بافت رسانه‌های اجتماعی به معنای برداشت و ذخیره‌سازی محتوا در یک سپردنگاه دیجیتال بر اساس قواعدی که مالک محتوا ایجاد

1. Twitter's Application Programming
2. Liu
3. Roy

کرده است و قابلیت‌های ارائه‌دهنده خدمت تعریف می‌گردد. برداشت و حفظ محتوای رسانه‌های در حال حاضر به یک موضوع مورد توجه در زمینه آرشيو تبدیل شده است. شرکت‌ها و مؤسسات توسعه، برنامه‌هایی را که منجر به برداشت داده‌ها از محیط رسانه‌های اجتماعی می‌گردد را مورد توجه قرار داده‌اند (اصنافی، ۱۳۹۵). رسانه‌های اجتماعی، خروجی فراتر از متن دارند. آن‌ها به محتوای گوناگون، نظیر وب سایت‌ها، عکس‌ها، ویدئوها و سایر رسانه‌ها پیوند می‌دهند. بنابراین، باید برنامه ویژه‌ای برای آرشيو سازی محتوای این منابع داشت (اصنافی، ۱۳۹۵). در چند سال اخیر نرم‌افزارهای بسیاری برای استخراج داده از رسانه‌های اجتماعی تولید شدند. این نرم‌افزارها پس از استخراج داده و آرشيو سازی، خدمات فراتری برای مدیریت محتوای رسانه‌های اجتماعی در اختیار کاربر قرار می‌دهند. در تمامی این نرم‌افزارها اهمیت تک‌تک داده‌ها در نظر گرفته شده است و برای حفظ داده‌ها روش‌های کاربردی ای بکار می‌رود؛ به عنوان مثال عموم این نرم‌افزارها با آرشيو سازی خودکار، این اطمینان را به کاربر می‌دهند که داده‌های آن‌ها حفظ خواهند شد، حتی برخی از این نرم‌افزارها، آرشيو سازی خودکار بلادرنگ^۱ دارند. از دیگر ویژگی‌های قابل توجه این نرم‌افزارها می‌توان به تحقق بخشی دولت شفاف اشاره کرد. شهروندان با دسترسی آزادانه به پرتال بایگانی‌های محتوای صفحات رسانه‌های اجتماعی دولت‌ها به تمامی داده‌ها دسترسی دارند. بخش‌های دولتی مدعی مدیریت شفاف می‌توانند با راه‌اندازی صفحه پرتال عمومی خود، روند سیاست‌های خود را شفاف‌سازی نمایند.

روش پژوهش

این پژوهش ضمن مطالعه کتابخانه‌ای در خصوص استخراج داده و آرشيو سازی رسانه‌های اجتماعی به عنوان ابزارهای جدید و تعاملی انتقال اطلاعات به بررسی نرم‌افزارهای تولید شده در این زمینه پرداخته است؛ در نهایت نیز براساس سیاهه‌وارسی تهیه شده، یک الگوی پیشنهادی از این نوع نرم‌افزارها برای استفاده کاربران داخلی ارائه داده است. ضمناً این روند با همراهی متخصصان رسانه‌های اجتماعی در علم اطلاعات و دانش‌شناسی انجام گرفته

1. In real time

است. در این پژوهش، گردآوری اطلاعات، به روش مطالعه اسنادی (کتابخانه‌ای) و پیمایشی تحلیلی انجام شد. از مطالعه اسنادی برای مرور مبانی نظری و دستیابی به چارچوب اولیه، با استفاده از کتاب‌ها، مقالات معتبر در پایگاه‌های اطلاعاتی امرالد، ساینس دایرکت، پروکوئست، اسکوپوس، تیلور و فرانسیس، بلکول-وایلی و مجلات معتبر علمی-پژوهشی، مقالات همایش‌ها و کنفرانس‌های ملی و بین‌المللی و بررسی نرم‌افزارهای فعلی استخراج و آرشیو داده از رسانه‌های اجتماعی به گردآوری داده‌ها پرداخته است. در روش پیمایشی نیز از پرسشنامه استفاده می‌شود. پرسشنامه‌ها به روش حضوری و یا از طریق رایانامه ارسال و گردآوری شدند. در این پژوهش، با نمونه‌گیری هدفمند و رصد افراد مطلع، تعداد ۳۱ نفر از متخصصان رسانه‌های اجتماعی در علم اطلاعات و دانش‌شناسی انتخاب شدند. ابزار گردآوری اطلاعات پرسشنامه است. قابلیت‌های به دست آمده از نرم‌افزارها در قالب طیف لیکرت با طیف‌های از کاملاً موافقم تا کاملاً مخالفم در اختیار متخصصان قرار گرفت و از آن‌ها خواسته شد که میزان موافقت خود را با ضرورت وجود هر یک از این قابلیت‌ها مشخص نمایند. پرسشنامه اولیه که استخراج قابلیت‌ها بود جهت بررسی روایی در اختیار ۴ نفر از کارشناسان و خبرگان و در مرحله دوم در اختیار ۳۱ خبره قرار گرفت. برای تعیین میزان اعتماد یا پایایی ابزار گردآوری داده‌ها نیز روش‌های مختلفی از جمله بازآزمایی، روش تنصیف، روش استفاده از آزمون‌های هم‌تا و نیز روش آلفای کرونباخ، وجود دارد که در اینجا از آلفای کرونباخ استفاده شد. پایایی پرسشنامه ۹۵ درصد به دست آمد که حاکی از پایایی بالای آن است. برای بررسی بودن یا نبودن هر یک از قابلیت‌های نرم‌افزارها از روش سی‌وی‌آر^۱، بار عاملی اکتشافی و بار عاملی تأییدی استفاده شد؛ بدین منظور، ابتدا از نرم‌افزار اس.پی.اس.اس. برای آمار توصیفی (فراوانی، درصد فراوانی، میانه و میانگین) استفاده شد. میزان پذیرش هریک از قابلیت‌ها با میانه (به دلیل رتبه‌ای بودن داده‌ها) ارزیابی (معیار سنجش میانه ۳) شد. اگر میانه بیشتر از ۳ باشد، آن گویه مقبول است. اگر کمتر از ۳ باشد، نامقبول و برابر ۳ متوسط است. سپس، هریک از گویه‌ها بر اساس میانگین موجود

در آزمون فریدمن (چرا که آزمون فریدمن برای داده‌های رتبه‌ای است) رتبه‌بندی شدند. در مرحله دوم سی‌وی‌آر هر یک از قابلیت‌ها و در مرحله آخر بار عاملی تأییدی با استفاده از نرم‌افزار ایموس¹ و بار عاملی اکتشافی با استفاده از اس.پی.اس.اس. بررسی شد. تحلیل عاملی به منظور تلخیص و کاهش تعداد زیادی از متغیرها به تعداد کوچک‌تری از عامل‌ها با هدف مدل‌سازی به کار می‌رود. بدین منظور از بار عاملی استفاده می‌شود. بار عاملی بیانگر میزان همبستگی متغیرهای بین متغیرهای مشاهده شده و عامل‌ها یا همان گویه‌هاست. هر بار عاملی در یک متغیر نشان می‌دهد که آن متغیر چقدر با عامل مربوطه همبستگی دارد. مقدار بار عاملی ۰/۳، سطح معناداری مقبول؛ بار عاملی ۰/۴، سطح معناداری بیشتر مقبول و بار عاملی ۰/۵ به بالا، سطح معناداری بسیار مقبول است (حبیب‌پورگتابی و صفری‌شالی، ۱۳۸۸). پرسش‌های پژوهش با استفاده از آزمون تی تک نمونه‌ای بررسی شدند. قبل از استفاده از این آزمون، چگونگی توزیع داده‌ها بررسی شد که همه قابلیت‌های نرم‌افزارها دارای توزیع نرمال بودند. این پژوهش از خردادماه سال ۱۳۹۸ آغاز شده و در بهمن‌ماه همان سال به اتمام رسیده است.

یافته‌های پژوهش

پرسش ۱. از نظر متخصصان، قابلیت‌های «مدیریت داده‌های آرشیوی» در نرم‌افزار استخراج داده از رسانه‌های اجتماعی تا چه میزان ضرورت دارند؟

در بررسی نمونه نرم‌افزارهای استخراج داده از رسانه‌های اجتماعی همان‌طور که در پیوست «د» آمده است، ۱۳ قابلیت تخصصی با مفهوم مدیریت داده‌های آرشیوی ذکر شد. این قابلیت‌ها عبارتند از:

۱. امکان آرشیو انواع قالب داده‌ها
۲. آرشیوسازی خودکار ۲۴/۷
۳. ارائه آرشیو با رابط بصری
۴. آرشیوسازی در قالب بومی (همراه با فایل‌ها، لینک‌ها و...)

۵. دسته‌بندی آرشیوها (مرورآسان)
۶. امکان جستجوی پیشرفته آرشیو
۷. بازنشانی رکوردهای حذف شده
۸. امکان اشتراک آرشیوها
۹. حذف خودکار آرشیوها در زمان تعریف شده
۱۰. بهره‌مندی از داده‌ها آرشیوی در نرم‌افزارهای مدیریت شبکه‌های اجتماعی
۱۱. بهره‌مندی از فناوری ابر
۱۲. امکان ایجاد آرشیو باز برای دسترسی عام
۱۳. استخراج داده از دیگر منابع وب (وبلاگ، وب سایت و ...)

از آنجایی که داده‌های آرشیوی دارای توزیع نرمال است، برای تفسیر آن از آزمون تی تک‌نمونه‌ای/تک گروهی استفاده می‌کنیم. این آزمون، زمانی استفاده می‌شود که می‌خواهیم آن را با یک حالت معمول و رایج، استاندارد و یا حتی یک عدد فرضی و مورد انتظار مقایسه کنیم (حبیب پورگنجابی و صفری شالی، ۱۳۸۸). عدد مفروض در اینجا ضرب تعداد پرسش‌های داده‌های آرشیوی (۱۳ سؤال) در ۳ (حد وسط طیف لیکرت) است که برابر عدد ۳۹ می‌شود، که این مقدار با میانگین واقعی مقایسه می‌شود. در تفسیر نتایج این آزمون همواره باید به اختلاف این دو میانگین یا اختلاف میانگین^۱ و مقدار آزمون تی توجه کرد. هر چه تفاوت بین دو میانگین بیشتر باشد مقدار تی نیز بیشتر خواهد بود و هر چه اختلاف میانگین بیشتر باشد، سطح معنی‌داری کوچکتر (کوچکتر از ۰/۰۵) می‌شود. توجه به مقادیر فاصله اطمینان (برآورد فاصله‌ای) در آزمون‌های تی، به تفسیر بهتر نتایج آزمون کمک می‌کند. در این برآورد، با توجه به مقدار حد بالا^۲ و حد پایین^۳، باید گفت که:

هرگاه حد پایین و بالا مثبت باشد، میانگین مورد انتظار از مقدار آزمون بزرگ‌تر است و تفاوت میانگین معنی‌دار است. هرگاه حد پایین و بالا منفی باشد، میانگین از مقدار آزمون

1. Mean Difference
2. Upper
3. Lower

ارائه الگوی پیشنهادی برای نرم‌افزار استخراج داده ...

کوچکتر است و تفاوت میانگین معنی دار است. هر گاه حد پایین منفی و حد بالا مثبت باشد، میانگین با مقدار مورد آزمون تفاوت معنی دار ندارد (حبیب پور کتابی و صفری شالی، ۱۳۸۸).

جدول ۱. آمار تک‌نمونه‌ای داده‌های آرشیوی

قابلیت نرم‌افزارها	تعداد	میانگین واقعی	میانگین مفروض
داده‌های آرشیوی	۳۱	۵۶/۴۵	۳۹

جدول ۲. آزمون تک‌نمونه‌ای داده‌های آرشیوی

قابلیت نرم‌افزار	مقدار آزمون تی	درجه آزادی	سطح معنی داری	اختلاف میانگین		برآورد فاصله‌ای
				حد بالا	حد پایین	
داده‌های آرشیوی	۱۲/۳۷	۳۰	۰	۱۷/۴۵	۱۴/۵۷	۲۰/۳۳

مقدار آزمون تی تک‌نمونه‌ای (۱۲/۳۷) در جدول ۴-۶، نشان می‌دهد که با اطمینان ۹۵ درصد و سطح خطای کوچکتر از ۰/۰۵ تفاوت معنی داری بین دو میانگین واقعی و مفروض وجود دارد. همچنین حد پایین و بالا هر دو مثبت است؛ بنابراین میانگین از مقدار آزمون بزرگتر و تفاوت میانگین معنی دار است. ضمن آنکه براساس جدول ۴-۵ مقدار میانگین واقعی (۵۶/۴۵) از مقدار میانگین مفروض (۳۹) بزرگتر است؛ بنابراین، میزان موافقت متخصصان با داده‌های آرشیوی زیاد است.

پرسش ۲. از نظر متخصصان، قابلیت‌های «اعتباربخشی» در نرم‌افزار استخراج داده از رسانه‌های اجتماعی، تاچه میزان ضرورت دارند؟

در بررسی نمونه نرم‌افزارهای استخراج داده از رسانه‌های اجتماعی همان‌طور که در پیوست «د» آمده است، ۲ قابلیت تخصصی با مفهوم اعتبار نرم‌افزارها ذکر شد. این قابلیت‌ها عبارتند از:

- اخذ گواهی‌های اعتباری از نهادهای دولتی و خصوصی
- درج امضای دیجیتال

از آنجایی که اعتبار نرم‌افزارها دارای توزیع نرمال است، برای تفسیر آن از آزمون تی تک‌نمونه‌ای/تک گروهی استفاده می‌کنیم.

جدول ۳- آمار تک‌نمونه‌ای اعتبار نرم‌افزار

قابلیت نرم‌افزارها	تعداد	میانگین واقعی	میانگین مفروض
اعتبار نرم‌افزار	۳۱	۵۶/۴۵	۳۹

جدول ۴- آزمون تک‌نمونه‌ای اعتبار نرم‌افزار

اعتبار نرم‌افزار	مقدار آزمون تی	درجه آزادی	سطح معنی‌داری		اختلاف میانگین	برآورد فاصله‌ای
			حد پایین	حد بالا		
۱۲/۳۷	۳۰	۰	۱۷/۴۵	۱۴/۵۷	۲۰/۳۳	

مقدار آزمون تی تک‌نمونه‌ای (۱۲/۳۷) در جدول ۴-۸ نشان می‌دهد که با اطمینان ۹۵ درصد و سطح خطای کوچکتر از ۰/۰۵ تفاوت معنی‌داری بین دو میانگین واقعی و مفروض وجود دارد. همچنین حد پایین و بالا هر دو مثبت است؛ بنابراین میانگین از مقدار آزمون بزرگتر و تفاوت میانگین معنی‌دار است. ضمن آنکه براساس جدول ۴-۷ مقدار میانگین واقعی (۵۶/۴۵) از مقدار میانگین مفروض (۳۹) بزرگتر است؛ بنابراین، میزان موافقت متخصصان با قابلیت اعتبار نرم‌افزار زیاد است.

پرسش ۳. از نظر متخصصان، قابلیت‌های «داده‌کاوی» در نرم‌افزار استخراج داده از رسانه‌های اجتماعی، تا چه میزان ضرورت دارند؟ در بررسی نمونه نرم‌افزارهای استخراج داده از رسانه‌های اجتماعی همانطور که در پیوست «د» آمده است، ۳ قابلیت تخصصی با مفهوم داده‌کاوی ذکر شد. این قابلیت‌ها عبارتند از:

۱. تحلیل داده‌های استخراجی

۲. ارائه گزارش‌های بصری

۳. اشتراک گزارش‌ها

از آنجایی که داده‌کاوی دارای توزیع نرمال است، برای تفسیر آن از آزمون تی تک‌نمونه‌ای/تک‌گروهی استفاده می‌کنیم.

جدول ۵- آمار تک‌نمونه‌ای داده‌کاوی

قابلیت نرم‌افزارها	تعداد	میانگین واقعی	میانگین مفروض
داده‌کاوی	۳۱	۲/۰۴	۶

جدول ۶- آزمون تک‌نمونه‌ای داده‌کاوی

قابلیت نرم‌افزار	مقدار آزمون تی	درجه آزادی	سطح معنی‌داری	اختلاف میانگین	برآورد فاصله‌ای
داده‌کاوی	۵/۴۳	۳۰	۰	۲	حد پایین: ۱/۲۴ حد بالا: ۲/۷۵

مقدار آزمون تی تک‌نمونه‌ای (۵/۴۳) در جدول ۴-۱۰ نشان دهنده آن است که با اطمینان ۹۵ درصد و سطح خطای کوچکتر از ۰/۰۵ تفاوت معنی‌داری بین دو میانگین واقعی و مفروض وجود دارد. همچنین حد پایین و بالا هر دو مثبت است؛ بنابراین میانگین از مقدار آزمون بزرگتر و تفاوت میانگین معنی‌دار است. ضمن آنکه براساس جدول ۴-۹ مقدار میانگین واقعی (۲/۰۴) از مقدار میانگین مفروض (۶) بزرگتر است؛ بنابراین، میزان موافقت متخصصان با قابلیت داده‌کاوی نرم‌افزارها زیاد است. پرسش ۴. از نظر متخصصان، قابلیت‌های «امنیتی» در نرم‌افزار استخراج داده از رسانه‌های اجتماعی، تاچه میزان ضرورت دارند؟ در بررسی نمونه نرم‌افزارهای استخراج داده از رسانه‌های اجتماعی همانطور که در جدول ۲-۶ آمده است، ۲ قابلیت تخصصی با مفهوم امنیت داده‌ها ذکر شد. این قابلیت‌ها عبارتند از:

۱. رمزگذاری داده‌ها (استخراج، ذخیره، اشتراک)

۲. تخصیص سطوح دسترسی

از آنجایی که امنیت داده‌ها دارای توزیع نرمال است، برای تفسیر آن از آزمون تی تک‌نمونه‌ای/تک‌گروهی استفاده می‌کنیم.

جدول ۷-آمار تک‌نمونه‌ای امنیت داده‌ها

قابلیت نرم‌افزارها	تعداد	میانگین واقعی	میانگین مفروض
امنیت داده‌ها	۳۱	۱۳/۳۵	۹

جدول ۸-آزمون تک‌نمونه‌ای امنیت داده‌ها

قابلیت نرم‌افزار	مقدار آزمون تی	درجه آزادی	سطح معنی‌داری	اختلاف میانگین	برآورد فاصله‌ای
امنیت داده‌ها	۱۰/۹۵	۳۰	۰	۴/۳۵	حد پایین ۳/۵۴ حد بالا ۵/۱۶

مقدار آزمون تی تک‌نمونه‌ای (۱۰/۹۵) در جدول ۴-۱۲، بیانگر آن است که با اطمینان ۹۵ درصد و سطح خطای کوچکتر از ۰/۰۵ تفاوت معنی‌داری بین دو میانگین واقعی و مفروض وجود دارد؛ همچنین حد پایین و بالا هر دو مثبت است؛ بنابراین میانگین از مقدار آزمون بزرگتر و تفاوت میانگین معنی‌دار است. ضمن آنکه براساس جدول ۴-۱۱ مقدار میانگین واقعی (۱۳/۳۵) از مقدار میانگین مفروض (۹) بزرگتر است؛ لذا، میزان موافقت متخصصان با امنیت داده‌ها زیاد است.

پرسش ۵. از نظر متخصصان، قابلیت‌های «مدیریت کاربران» در نرم‌افزار استخراج داده از رسانه‌های اجتماعی، تا چه میزان ضرورت دارند؟

در بررسی نمونه نرم‌افزارهای استخراج داده از رسانه‌های اجتماعی همانطور که در پیوست «د» آمده است، ۲ قابلیت تخصصی با مفهوم مدیریت کاربران ذکر شد. این قابلیت‌ها عبارتند از:

۱. تعریف کاربر و مدیر

۲. تخصیص نقش به کاربران

از آنجایی که مدیریت کاربران دارای توزیع نرمال است، برای تفسیر آن از آزمون تی تک‌نمونه‌ای/تک گروهی استفاده می‌کنیم.

جدول ۹- آمار تک‌نمونه‌ای مدیریت کاربران

قابلیت نرم‌افزارها	تعداد	میانگین واقعی	میانگین مفروض
مدیریت کاربران	۳۱	۹/۹	۶

جدول ۱۰- آزمون تک‌نمونه‌ای مدیریت کاربران

قابلیت نرم‌افزار	مقدار آزمون تی	درجه آزادی	سطح معنی‌داری	اختلاف میانگین	برآورد فاصله‌ای
مدیریت کاربران	۱۱/۵۵	۳۰	۰	۲/۹۰	حد پایین: ۲/۳۹ حد بالا: ۳/۴۱

مقدار آزمون تی تک‌نمونه‌ای (۱۱/۵۵) در جدول ۴-۱۴، نشان می‌دهد که با اطمینان ۹۵ درصد و سطح خطای کوچکتر از ۰/۰۵ تفاوت معنی‌داری بین دو میانگین واقعی و مفروض وجود دارد. از سویی حد پایین و بالا هر دو مثبت است؛ از این‌رو، میانگین از مقدار آزمون بزرگتر و تفاوت میانگین معنی‌دار است. ضمن آنکه براساس جدول ۴-۱۳ مقدار میانگین واقعی (۵۶/۴۵) از مقدار میانگین مفروض (۳۹) بزرگتر است. بنابراین، میزان موافقت متخصصان با قابلیت مدیران کاربران نرم‌افزارها زیاد است.

پرسش ۶. از نظر متخصصان، قابلیت‌های «تعرفه مالی» در نرم‌افزار استخراج داده از رسانه‌های اجتماعی، تا چه میزان ضرورت دارند؟ در بررسی نمونه نرم‌افزارهای استخراج داده از رسانه‌های اجتماعی همانطور که در پیوست «د» آمده است، ۳ قابلیت تخصصی با مفهوم تعرفه مالی ذکر شد. این قابلیت‌ها عبارت‌اند از:

۱. دریافت نسخه کامل پس از پرداخت هزینه

۲. نسخه دمو

۳. بسته‌هایی با امکانات و تعرفه‌های متفاوت از آنجایی که تعرفه مالی دارای توزیع نرمال است، برای تفسیر آن از آزمون تی تک‌نمونه‌ای/تک گروهی استفاده می‌کنیم.

جدول ۱۱- آمار تک‌نمونه‌ای تعرفه مالی

قابلیت نرم‌افزارها	تعداد	میانگین واقعی	میانگین مفروض
تعرفه مالی	۳۱	۸/۹۳	۶

جدول ۱۲- آزمون تک‌نمونه‌ای تعرفه مالی

قابلیت نرم‌افزار	مقدار آزمون تی	درجه آزادی	سطح معنی‌داری	اختلاف میانگین	
				حد پایین	حد بالا
تعرفه مالی	۱۰/۸۶	۳۰	۰	۲/۹۳	۳/۴۸

مقدار آزمون تی تک‌نمونه‌ای (۱۰/۸۶) در جدول ۴-۱۶، گویای آن است که با اطمینان ۹۵ درصد و سطح خطای کوچکتر از ۰/۰۵ تفاوت معنی‌داری بین دو میانگین واقعی و مفروض وجود دارد. همچنین حد پایین و بالا هر دو مثبت است؛ بنابراین میانگین از مقدار آزمون بزرگتر و تفاوت میانگین معنی‌دار است. ضمن آنکه براساس جدول ۴-۱۵ مقدار میانگین واقعی (۸/۹۳) از مقدار میانگین مفروض (۶) بزرگتر است؛ بنابراین، میزان موافقت متخصصان با قابلیت نرم‌افزارها زیاد است.

پرسش ۷. از نظر متخصصان، قابلیت‌های فعلی در نرم‌افزارهای استخراج داده از رسانه‌های اجتماعی تا چه میزان ضرورت دارند؟

در بررسی نمونه نرم‌افزارهای استخراج داده از رسانه‌های اجتماعی همان‌طور که در پیوست «ج» آمده است، ۶ مفهوم از قابلیت‌های تخصصی ذکر شد. این مفاهیم عبارت‌اند از:

۱. داده‌های آرشیوی
۲. اعتبار نرم‌افزار
۳. داده‌کاوی

۴. امنیت داده‌ها

۵. مدیریت کاربران

۶. تعرفه مالی

از آنجایی که این مفاهیم دارای توزیع نرمال است، برای تفسیر آن از آزمون تی تک‌نمونه‌ای/تک گروهی استفاده می‌کنیم.

جدول ۱۳- آمار تک‌نمونه‌ای همه قابلیت‌ها

قابلیت نرم‌افزارها	تعداد	میانگین واقعی	میانگین مفروض
همه قابلیت‌ها	۳۱	۱۰۷/۵۸	۷۵

جدول ۱۴- آزمون تک‌نمونه‌ای تعرفه مالی

قابلیت نرم‌افزار	مقدار آزمون تی	درجه آزادی	سطح معنی‌داری	اختلاف میانگین	
				حد پایین	حد بالا
همه قابلیت‌ها	۱۰/۸۹	۳۰	۰	۲۶/۴۷	۳۸/۶۸

مقدار آزمون تی تک‌نمونه‌ای (۱۰/۸۹) در جدول ۴-۱۸، بیانگر آن است که با اطمینان ۹۵ درصد و سطح خطای کوچکتر از ۰/۰۵ تفاوت معنی‌داری بین دو میانگین واقعی و مفروض وجود دارد. همچنین حد پایین و بالا هر دو مثبت است. لذا، میانگین از مقدار آزمون بزرگتر و تفاوت میانگین معنی‌دار است. از طرفی براساس جدول ۴-۱۷ مقدار میانگین واقعی (۱۰۷/۵۸) از مقدار میانگین مفروض (۷۵) بزرگتر است. بنابراین، میزان موافقت متخصصان با همه قابلیت‌های نرم‌افزارها زیاد است.

پرسش ۸. نرم‌افزار کاربردی الگو برای استخراج و آرشیو رسانه‌های اجتماعی از چه قابلیت‌های تخصصی‌ای برخوردار است؟

براساس یافته‌های بخش تجزیه و تحلیل پرسشنامه در همین فصل، داده‌های جدول ۴-۱، نشان داد تمامی قابلیت‌های تخصصی نرم‌افزار از نظر متخصصان لازم هستند. قابلیت‌های «امکان جستجوی پیشرفته آرشیو»، «دسته‌بندی آرشیوها (مرورآسان)» و «ارائه گزارش‌های

بصری» مهم‌ترین الزامات نرم‌افزارها از نظر متخصصان هستند. قابلیت «اخذ گواهی‌های اعتباری از نهادهای دولتی و خصوصی» کمترین اهمیت را از دید متخصصان دارد. داده‌های جدول ۴-۲ بیانگر آن است که معیارهای «بازنشانی رکوردهای حذف شده»، «امکان ایجاد آرشیو باز برای دسترسی عام»، «اخذ گواهی‌های اعتباری از نهادهای دولتی و خصوصی»، «دریافت نسخه کامل پس از پرداخت هزینه» و «نسخه دمو» نمره کمتر از ۵۰ را به دست آوردند.

داده‌های جدول ۴-۳ نیز نشان‌دهنده آن است که تمامی قابلیت‌ها بار عاملی تأییدی و اکتشافی بیشتر از ۰/۳ دارند. بنابراین، در نرم‌افزارها در نظر گرفته می‌شوند. اما با مقایسه بار عاملی قابلیت‌ها با سی‌وی آر قابلیت‌ها می‌توانیم قابلیت‌هایی که سی‌وی آر کمتر از ۵۰ داشتند را حذف کنیم.

بحث و نتیجه‌گیری

در سال‌های اخیر نرم‌افزارهای گوناگونی با ویژگی‌های مختلف به منظور استخراج و آرشیو سازی داده‌ها از رسانه‌های اجتماعی ایجاد شدند. این نرم‌افزارها توسط شرکت‌های مختلفی در دنیا ارائه شده‌است. بنا بر یافته‌های این پژوهش نرم‌افزارهای استخراج داده از رسانه‌های اجتماعی ساخت ایران یافت نشد. این درحالی است که عموم مردم، سازمان‌های دولتی و خصوصی، شرکت‌ها و نهادهای ایرانی در رسانه‌های اجتماعی مختلف فعالیت دارند. در این پژوهش سعی شده است با ارائه یک الگو پیشنهادی برای ایجاد نسخه ایرانی نرم‌افزار استخراج داده از رسانه‌های اجتماعی، راه‌حلی برای این خلاء ارائه شود.

با گسترش روزافزون انواع رسانه‌های اجتماعی و افزایش روزانه حجم داده‌های بارگذاری شده در آن‌ها، بدیهی است که تعجیل در راه‌اندازی این نوع نرم‌افزارها می‌تواند اهمیت داشته باشد. ضمن اینکه کاربران همیشه ترجیح می‌دهند از نمونه‌های پیشگام استفاده کنند. با گسترش استفاده از وب و پایگاه‌های اینترنتی، این نوع نرم‌افزارها می‌توانند راه‌حل‌های کاربردی و نوآورانه‌ای را در زمینه آرشیو وب به ویژه آرشیو رسانه‌های اجتماعی، ارائه دهند. نرم‌افزار ارائه شده باید بتواند برای طیف گسترده‌ای از صنایع از جمله

ارائه الگوی پیشنهادی برای نرم‌افزار استخراج داده ...

امور مالی، حقوقی، مخابرات، خرده‌فروشی، خدمات آب و برق، آموزش و پرورش دولتی، تحصیلات تکمیلی و... خدمات آرشیوی ارائه دهد. این نوع نرم‌افزارها از نوع نرم‌افزارهای خدماتی است که سازمان‌ها و شرکت‌ها از هر اندازه‌ای را قادر می‌سازند تا به طور دائمی وب‌سایت و محتوای رسانه‌های اجتماعی خود را با کیفیت معنی‌دار حفظ کنند و سپس به آن آرشیو دسترسی پیدا کنند.

حفظ محتوای رسانه‌های اجتماعی و سپس آرشیو آن مستلزم ثبت داده‌ها است. یکی از ویژگی‌های حائز اهمیت این نرم‌افزارها استخراج و ثبت تمامی داده‌های موجود در رسانه‌های اجتماعی است. تمامی داده‌ها از نظر این نرم‌افزار با ارزش و حساس به حساب می‌آیند و تمامی پیام‌ها، فعالیت‌های کاربران، تغییرات و... بایگانی خواهند شد.

پژوهش حاضر ضمن بررسی متون در خصوص استخراج داده و آرشیوسازی رسانه‌های اجتماعی به عنوان ابزارهای جدید و تعاملی انتقال اطلاعات به بررسی برخی از نرم‌افزارهای این چنین پرداخت. سپس سه نرم‌افزار پیچ‌فریز، آرکایوسوشال و اسمارش به طور ویژه مورد بررسی قرار گرفتند. پس از آن قابلیت‌های تخصصی این سه نرم‌افزار در قالب یک پرسشنامه در اختیار متخصصان و اساتید حوزه علم اطلاعات و دانش‌شناسی قرار گرفت. در نهایت نیز با ارائه یک الگو کاربردی از این نوع نرم‌افزارها بر اساس سیاهه‌وارسی تهیه شده، برای استفاده کاربران داخلی رسانه‌های اجتماعی می‌پردازد. با بررسی نرم‌افزارهای فعلی استخراج داده از رسانه‌های اجتماعی مشاهده شد این نرم‌افزارها دارای قابلیت‌های گسترده و کاربردی می‌باشند. برخی از این قابلیت‌ها به طور عمومی در انواع نرم‌افزارها یافت می‌شود. اما برخی دیگر از این قابلیت‌ها به طور تخصصی و ویژه در نرم‌افزارهای استخراج داده از رسانه‌های اجتماعی وجود دارد. این قابلیت‌ها در ۶ مفهوم دسته‌بندی شدند که عبارت‌اند از: داده‌های آرشیوی، اعتبار نرم‌افزار، داده‌کاوی، امنیت داده‌ها، مدیریت کاربران، تعرفه مالی.

نتایج این پژوهش نشان داد متخصصان علم اطلاعات و دانش‌شناسی قابلیت‌های «امکان جستجوی پیشرفته آرشیو»، «دسته‌بندی آرشیوها (مرور آسان)» و «ارائه گزارش‌های بصری» را مهم‌ترین الزامات نرم‌افزارهای استخراج داده از رسانه‌های اجتماعی می‌دانند. از این موارد دو مورد اول یعنی «امکان جستجوی پیشرفته آرشیو» و «دسته‌بندی آرشیوها (مرور آسان)» در

مفهوم داده‌های آرشیوی قرار گرفتند. قابلیت «ارائه گزارش‌های بصری» مربوط به مفهوم داده‌کاوی است. قابلیت مهم «امکان جستجوی پیشرفته آرشیو» در تمامی نرم‌افزارهای مورد پژوهش این تحقیق مشاهده شد. الگوی نرم‌افزار استخراج داده از رسانه‌های اجتماعی باید یک موتور جستجوی پیشرفته و سیستم برچسب‌گذاری سفارشی برای سازماندهی و فیلتر کردن محتوای بایگانی خود داشته باشند. عدم وجود به سیستم جستجوی پیشرفته، جستجو در میان انبوهی از پست‌ها، نظرات، توئیٹ‌ها و... که هر یک می‌تواند داده‌ای با ارزش محسوب شود، فرآیندی بسیار دشوار خواهد بود. براساس یافته‌های این پژوهش همانطور که پیشینه‌ها نیز بر اهمیت دسته‌بندی آرشیوها اشاره داشتند، متخصصان علم اطلاعات نیز این قابلیت را از موارد مهم دانستند. نرم‌افزارهای فعلی با گسترش دامنه شاخص‌ها، دیگر قابلیت تخصصی مهم یعنی «دسته‌بندی آرشیوها (مرورآسان)» محقق می‌سازند. بررسی نرم‌افزارها نشان داد آن‌ها با گسترش دامنه شاخص‌هایی همچون نوع محتوا، محدوده، تاریخ، شبکه‌ها، کلمات کلیدی، شرکت‌کننده‌ها و برچسب‌ها، آرشیوهای ایجاد شده را دسته‌بندی می‌کنند. قابلیت «ارائه گزارش‌های بصری» نیز از دیگر قابلیت‌های مهم از دیدگاه متخصصان بود. تهیه گزارش از داده‌ها برای به دست آوردن بینش در استفاده از داده‌ها در روند کسب و کارها، مراکز علمی، نهادهای دولتی و... از قابلیت‌های تخصصی مهم به شمار می‌رود. نرم‌افزارهایی که در این پژوهش بررسی شدند گزاره‌ای بصری جامع را از داده‌های استخراج آرشیو شده به کاربران خود ارائه می‌دهند.

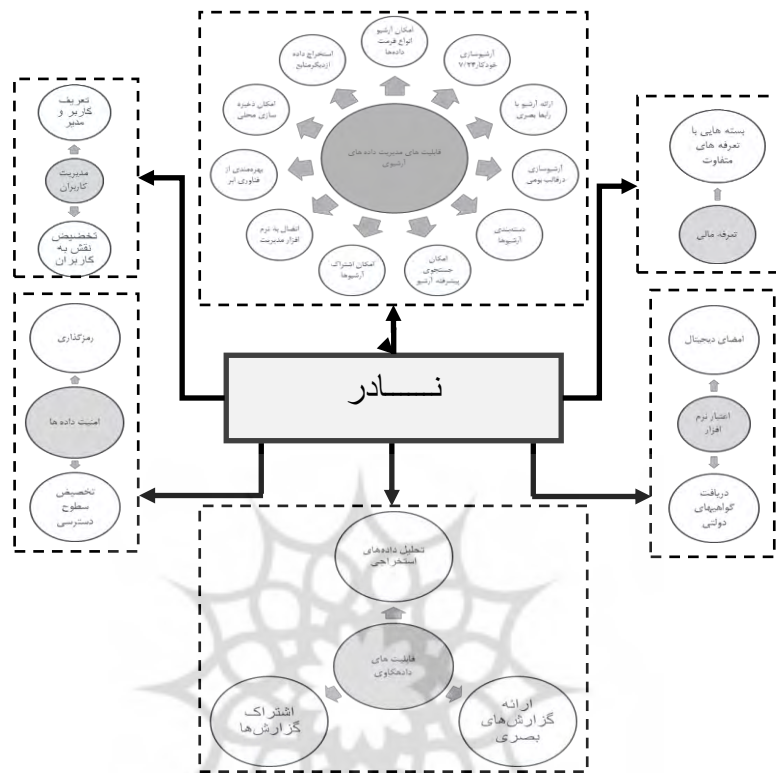
علاوه بر این متخصصان علم اطلاعات و دانش‌شناسی برای یک نرم‌افزار استخراج داده از رسانه‌های اجتماعی قابلیت‌های دیگری را به سیاهه واری اولیه اضافه کردند. برخی از این قابلیت‌ها مربوط به قابلیت‌های عمومی انواع نرم‌افزارها است که انتظار می‌رود در نرم‌افزار کاربردی الگو استخراج داده از رسانه‌های اجتماعی نیز در نظر گرفته شود. برخی دیگر از موارد ذکر شده در قابلیت‌های سیاهه واری اولیه گنجانده شده بود. از این میان یک قابلیت «امکان ذخیره‌سازی محلی» به سیاهه واری نهایی اضافه گردید.

جدول ۱۵- سیاه‌واری نهایی از قابلیت‌های نرم‌افزار الگو برای استخراج داده

از رسانه‌های اجتماعی	
مفاهیم	قابلیت‌ها
داده‌های آرشیوی	۱. امکان آرشیو انواع فرمت داده‌ها
	۲. آرشیوسازی خودکار ۲۴/۷
	۳. ارائه آرشیو با رابط بصری
	۴. آرشیوسازی در قالب بومی (همراه با فایل‌ها، لینک‌ها و...)
	۵. دسته‌بندی آرشیوها مرور آسان
	۶. امکان جستجوی پیشرفته آرشیو
	۷. امکان اشتراک آرشیوها
	۸. بهره‌مندی از داده‌ها آرشیوی در نرم‌افزارهای مدیریت شبکه‌های اجتماعی
	۹. بهره‌مندی از فناوری ابر
	۱۰. امکان ذخیره سازی محلی
	۱۱. استخراج داده از دیگر منابع وب (وبلاگ، وب سایت و...)
اعتبار نرم‌افزار	۱۲. درج امضای دیجیتال
	۱۳. دریافت گواهی‌های دولتی
داده کاوی	۱۴. تحلیل داده‌های استخراجی
	۱۵. ارائه گزارش‌های بصری
	۱۶. اشتراک گزارش‌ها
امنیت داده‌ها	۱۷. رمزگذاری داده‌ها (استخراج، ذخیره، اشتراک)
	۱۸. تخصیص سطوح دسترسی
مدیریت کاربران	۱۹. تعریف کاربر و مدیر
	۲۰. تخصیص نقش به کاربران
تعرفه مالی	۲۱. بسته‌هایی با امکانات و تعرفه‌های متفاوت

در تکمیل پاسخ به پرسش نهم، که برخاسته از هدف اصلی این پژوهش است، در این بخش یک الگوی مفهومی ارائه می‌گردد. الگوی مفهومی برای فهمیدن، شبیه سازی و درک

بهرتر مفاهیم موردنظر است. همچنین این الگو تعیین‌کننده چارچوب برای طراحی و ایجاد قابلیت‌های ضروری یک نرم‌افزار استخراج داده از رسانه‌های اجتماعی است. همان‌طور که پیش‌تر گفته شد، نرم‌افزار استخراج داده از رسانه‌های اجتماعی، در این پژوهش به اختصار «نادر» نامگذاری گردید. بررسی نرم‌افزارهای فعلی نشان داد، نادر باید ۶ مفهوم را با قابلیت‌های تخصصی خود پیاده‌سازی کند. این مفاهیم که عبارت‌اند از: داده‌های آرشیوی، اعتبار نرم‌افزار، داده‌کاوی، امنیت داده‌ها، مدیریت کاربران و تعرفه مالی، در الگوی مفهومی مشخص شدند. نادر پس از استخراج داده و آرشیوسازی آن، خدمات فراتری برای مدیریت محتوای رسانه‌های اجتماعی در اختیار کاربر قرار می‌دهند. این خدمات که از نظر متخصصان نیز از اهمیت بالایی برخوردار بودند، با مفهوم «قابلیت‌های مدیریت داده‌های آرشیوی» در الگوی مفهومی نمایش داده شدند؛ این قابلیت‌ها اصلی‌ترین مأموریت نرم‌افزار نادر، یعنی حفاظت و آرشیو داده‌ها را برعهده دارند. در این نرم‌افزار اهمیت تک‌تک داده‌ها در نظر گرفته شده است و برای حفظ داده‌ها روش‌های کاربردی‌ای بکار می‌رود؛ برای مثال با قابلیت آرشیوسازی خودکار، این اطمینان را به کاربر می‌دهد که داده‌های آن‌ها حفظ خواهند شد. مفهوم «داده‌کاوی» نیز در الگوی مفهومی نادر گنجانده شد. نتایج نظرسنجی از متخصصان نشان داد قابلیت‌هایی که در گروه این مفهوم قرار گرفته بودند، برای نرم‌افزار نادر از ضرورت بالایی برخوردار هستند. با توجه به ضرورت قابلیت‌هایی که در مفاهیم «داده‌کاوی» و «داده‌های آرشیوی» قرار دارند، بخش عمده‌ای از الگوی مفهومی به این دو مفاهیم اختصاص یافت. نتایج بررسی نرم‌افزارهای فعلی و نظرسنجی از متخصصان نشان داد، ۳ مفهوم دیگر یعنی اعتبار نرم‌افزار، داده‌کاوی، امنیت داده‌ها، مدیریت کاربران و تعرفه مالی نیز باید در الگوی مفهومی نادر قرار بگیرند. هر کدام از این مفاهیم با قابلیت‌هایشان در الگوی مفهومی مشخص شدند. همچنین نتایج نشان داد قابلیت‌هایی که در مفهوم «امنیت داده‌ها» قرار گرفته بودند، برای نرم‌افزار نادر از ضرورت بالایی برخوردار هستند. این مفهوم نیز در الگوی مفهومی نمایش داده شد.



شکل ۱: الگوی مفهومی نرم‌افزار نادر

نتایج پژوهش‌های پیشین نشان می‌دهد، رسانه‌های اجتماعی، نسل بعدی آرشیوها هستند. آن‌ها معتقدند آرشیو داده‌های رسانه‌های اجتماعی به محققان این امکان را می‌دهد تا مجموعه داده‌های قابل استفاده مجدد در رسانه‌های اجتماعی را کشف کنند؛ وسیله‌ای برای ارزیابی و یا تکرار تحقیقات براساس داده‌های رسانه‌های اجتماعی فراهم کنند و بینش‌های جدید، مطالعات یا تجزیه و تحلیل‌های مقایسه‌ای که غیرممکن است را ممکن سازند.

نتایج یافته‌های این پژوهش نیز همچون پژوهش‌های پیشین اهمیت این مسأله را تأیید می‌کند. در بررسی قابلیت‌های نرم‌افزارهای فعلی استخراج داده از رسانه‌های اجتماعی اهمیت این مسئله مشخص شد. عموم این نرم‌افزارها از قابلیت‌هایی برای داده‌کاوی برخوردار هستند. نتایج نظرسنجی از متخصصان نیز نشان دهنده ضرورت وجود قابلیت‌های داده‌کاوی

در نرم‌افزارهای اینگونه است. توجه به قابلیت داده‌کاوی تأیید بر با ارزشمندی داده‌های موجود در رسانه‌های اجتماعی دارد که در پژوهش‌های پیشین به دفعات مطرح شده است. به اعتقاد پژوهش‌های پیشین باید برنامه ویژه‌ای برای آرشیوسازی محتوای رسانه‌های اجتماعی داشت. تلاش‌هایی برای ساماندهی و ارائه راهبرد برای آرشیو منابع رسانه‌های دیجیتال توسط پژوهشگران پیشین شده است. اما این راهبردها از نظر آرشیوسازی و نرم‌افزارهای تخصصی آرشیو داده مورد بررسی قرار نگرفته‌اند. این پژوهش توانست با بررسی نرم‌افزارهای فعلی استخراج داده از رسانه‌های اجتماعی و اعمال نظر متخصصان، قابلیت‌های تخصصی نرم‌افزار استخراج و آرشیو داده از رسانه‌های اجتماعی را مشخص کند.

فهرست منابع

اصنافی، امیررضا (۱۳۹۵). تأملی بر آرشیوسازی رسانه‌های اجتماعی. *گنجینه اسناد*، ۱۰۲، ۹۶-۱۱۳

ایروانچی، عارفه (۱۳۹۴). روشی مستقل از معنا برای استخراج کلمه کلیدی داده‌های متنی کاربران در یک شبکه اجتماعی. پایان‌نامه کارشناسی ارشد مهندسی کامپیوتر گرایش نرم‌افزار. دانشکده فنی و مهندسی، دانشگاه علم و فرهنگ.

حبیب‌پورگتابی، کرم؛ و صفری‌شالی، رضا (۱۳۸۸). *راهنمای جامع کاربرد SPSS در تحقیقات پیمایشی: (تحلیل داده‌های کمی)*. تهران: لویه.

Allen, Dana (2018). Collecting and using social media. IFLA WLIC 2018, August 26, 2018

Archive Social Software. Accessed Jun/Jul 2019. From *Archive Social* website: <http://www.ArchiveSocial.com>

Archived White House Websites and Social Media. Accessed Jun/Jul 2019. From: <https://www.obamalibrary.gov/research/archived-white-house-websites-and-social-media>

Banks, Marcus (2009). Blog posts and tweets: the next frontier for grey literature. In: Farace, Dominic & Schopfel, Joachim, editors. *Grey Literature in Library and Information Studies*. Berlin: De Gruyter, :217-226.

Boyd, danah m. Ellison, Nicole B (2007). Social Network Sites: Definition, History, and Scholarship. *Journal of Computer-Mediated Communication*, 13, (1) :210-230

- Brent, Lewers (2013). Senior Policy Analyst Inland Revenue ,*New Zealand Government Controller and AuditorGeneral*, Learning from public entities' use of social media, June 2013
- Butterfield, Andrew. Ngondi, Gerard Ekembe. Kerr ,Anne (2016). *Dictionary of Computer Science* (7 ed.). Oxford University Press.
- Castells, Manuel (2007). Communication, Power and Counter-power in the Network Society. *International Journal of Communication*,1: 238-266.
- Castells, Manuel (2009). *Communication Power*. Oxford University Press. New York, 2009
- Chen ,Hsinchun. H. L. Chiang ,Roger. C. Storey ,Veda (2012). BUSINESS INTELLIGENCE AND ANALYTICS: FROM BIG DATA TO BIG IMPACT. *MIS Quarterly*,Vol.. 36(4): 1165-1188
- Diga, M., & Kelleher, T. (2009). Social media use, perceptions of decision-making power, and public relations roles. *Public Relations Review*, 35, 440–442.
- digitalinformationworld. Accessed Jun/Jul 2019. From: <https://www.digitalinformationworld.com/2019/01/most-popular-global-social-networks-apps-infographic.html>
- Doane, David P.; & Seward, Lori E. (2011). Measuring Skewness: A Forgotten Statistic? *Journal of Statistics Education*, 19(2), 1-18.
- DOMO(2018), Data Never Sleeps 5.0, from: <https://www.domo.com/learn/data-never-sleeps-5>
- Gripenberg, Pernilla. (2005). ICT and the shaping of society: Exploring human-ICT relationships in everyday life. (Ph.D. thesis) Helsinki: Edita Prima Ltd.
- Harrower, N.; Heravi, B. R.(2015). How to Archive an Event: Reflections on the Social Repository of Ireland. *New Review of Information Networking*, 20(1-2):104-116.
- Hemphill, Libby. Leonard, Susan.H. Hedstrom, Margaret. (2018) DevelopingaSocialMediaArchiveatICPSR. WADL'18, June 2018, Fort Worth, Texas USA
- Huang, Z., & Benyoucef, M. (2013). User-centered investigation of social commerce design. Paper presented at the International Conference on Online Communities and Social Computing
- Kaplan, A. M., & Haenlein, M. (2010). Users of the world, unite! The challenges and opportunities of Social Media. *Business Horizons*, 53(1), 59—68.
- Kelly, B.; Hawksey, M.; O'Brien, J.; Guy, M.;Rowe, M.(2010). *Twitter archiving using Twapper Keeper: technical and policy challenges*. In 7th International Conference on Preservation of Digital Objects (iPRES 2010). University of Bath.

- Kietzmann, J. H., Hermkens, K., McCarthy, I. P., & Silvestre, B. S. (2011). Social media? Get serious! Understanding the functional building blocks of social media. *Business Horizons*, 54(3), 241-251
- Latham, Ross (2014). Information Management Advice 57 Managing Social Media Records Part 4: How to Capture Social Media Records, Tasmanian Archive+ Heritage Office, Issued: June 2014. From: <https://www.informationstrategy.tas.gov.au>
- LIU, Yili. YING, Xiangxiang (2014). *A Review of Social Network Sites: Definition, Experience and Applications*. The Conference on Web Based Business Management.
- Lyman, Peter (2002). *Archiving the World Wide Web*. In Preserving Our Digital Heritage: Plan for the National Digital Information Infrastructure and Preservation Program. Washington, DC: Library of Congress. Retrieved Dec 25, 2011. From: http://www.digitalpreservation.gov/about/ndiipp_appendix.pdf.
- Mangus, Que (2014). *Why do I need to archive social media? Here are 4 good reasons!* From: <https://blog.microfocus.com/author/qmangus>
- Marr, Bernard (2018). *How Much Data Do We Create Every Day? The Mind-Blowing Stats Everyone Should Read*. forbes, May 21, 2018. From: <https://www.forbes.com/sites/bernardmarr/2018/05/21/how-much-data-do-we-create-every-day-the-mind-blowing-stats-everyone-should-read/#14e5662860ba>
- Musser, L. (2016). Preserving the Digital Record of Science and Engineering: the Challenge of New Forms of Grey Literature. *Issues in Science and Technology Librarianship*. Retrieved 29 May 2016. From: <http://istl.org/16-winter/short.html>.
- NARA (2010). *A report on federal web 2.0 use and record value*. Retrieved July 07, 2015, from NARA website: <http://www.archives.gov/recordsmgmt/resources/web2.0-use.pdf>
- NARA, (2013). *National Archives and Records Administration, White Paper on Best Practices for the Capture of Social Media Records*. May 2013. From NARA website: <https://www.archives.gov/files/recordsmgmt/resources/socialmediacapture.pdf>
- Page freezer Software. Accessed Jun/Jul 2019. From *Page freezer* website: <https://www.pagefreezer.com/>
- Ringham, Larissa; Roman Amigo, Carolina (2016). *UBC Library Web Archiving 2016: Digital Projects Librarian at University of British Columbia Library*. Retrieved 28May2016. From: http://www.slideshare.net/LarissaRingham/ubc-library-web-archiving2016?qid=2da7072b-2939-499d-bdf1-5c7d2f51a370&v=&b=&from_search=1.

- Roy, Rishiraj Saha . Padmakumar, Aishwarya. Prasaad Jeganathan ,Guna. Kumaraguru ,Ponnurangam (2015).Automated Linguistic Personalization of Targeted Marketing Messages Mining User-Generated Text on Social Media. *Springer International Publishing Switzerland*, CICLing 2015, Part II, LNCS 9042,:203-224
- Smarsh Software. Accessed Jun/Jul 2019. From *Smarsh* website: <http://www.smarsh.com/>
- TAHO (2014), *Information Management Advice 57 Managing Social Media Records* Part 4: How to Capture Social Media Records ,Tasmanian Archive+ Heritage Office. From: <https://www.informationstrategy.tas.gov.au>
- The British Library,(2010). The British Library Collection Development Policy for Websites. Accessed 9 June 2016. From: <http://www.bl.uk/reshelp/pdfs/modbritcdpwebsites.pdf>.
- Thompson, Dave (2008). Archiving Web Resources. *DCC / Digital Curation Manual*, December 2008, Version 1.0
- Thomson, S. D.; Kilbride, W.(2015). Preserving Social Media: The Problem of Access. *New Review of Information Networking*, 20(1-2);:261-275.

