

فصلنامه مطالعات مدیریت کسب و کار هوشمند سال ششم، شماره ۲۱، پاییز ۱۳۹۶
صفحات ۱۶۹ تا ۱۸۸

جایگاه کتاب و کتابخوانی در سایت تبیان با رویکرد متن کاوی و تحلیل شبکه‌های اجتماعی

سلما کشاورزبان *

حوریه براردخت **

چکیده

در این پژوهش با تمرکز بر روی مقوله مطالعه، سایت موسسه فرهنگی و اطلاع‌رسانی تبیان با استفاده از خوشه‌بندی موضوعی مورد متن‌کاوی دقیق قرار گرفته است. با استفاده از روش خوشه‌بندی موضوعی و متن‌کاوی، ضمن مشخص نمودن خوشه‌های موضوعی برجسته و کم‌اهمیت به بررسی اهمیت این موضوع در میان کاربران شبکه تبیان پرداخته شده است. با توجه به ماهیت فرهنگی تبیان و امر خواندن و مطالعه، جستجوی عبارات کلیدی مرتبط بیشترین نتایج را در طبقه موضوعی فرهنگ به ترتیب با ۲۶ و ۱۸ درصد در اختیار خواهد گذاشت. با توجه به ماهیت مجازی و الکترونیکی بودن سایت تبیان، لازم است بخش آموزش‌های الکترونیکی این سایت با تولید محتوای آموزشی بیشتر این کمبود را پوشش دهد. با وجود بالا بودن تعداد نتایج حاصل در طبقه

* کارشناسی ارشد، علم اطلاعات و دانش‌شناسی، دانشگاه شهید بهشتی، تهران. (نویسنده مسئول)؛

salmak1351@gmail.com,

** کارشناسی ارشد، علم اطلاعات و دانش‌شناسی، دانشگاه شیراز، شیراز.

تاریخ پذیرش: ۱۳۹۶/۱۰/۲۵

تاریخ دریافت: ۱۳۹۶/۰۴/۲۷

۱۷۰ فصلنامه مطالعات مدیریت کسب و کار هوشمند، سال ششم، شماره ۲۱، پاییز ۱۳۹۶

موضوعی جامعه، نتایج در کل شبکه برای مقوله کتاب‌خوانی بسیار اندک به نظر می‌رسد (تنها به ترتیب ۹۵۲ و ۱۱۹۹ نتیجه در جستجوی عبارت کلیدی خواندن و مطالعه) و این امر لزوم فعالیت بیشتر تبیان در راستای اعتلای این فرهنگ مغفول مانده در کشور را نمایان می‌سازد. جامعه جهانی هرروز در پرتو تحولات جدید، فناوری مدرن را به جامعه ارائه می‌دهد و در این میان اگر کسی از دانش روز عقب بماند هیچ فرصت جبرانی برای مهروموم‌های ازدست‌رفته ندارد؛ بنابراین، باید برای اهمیت و ارزش دادن افراد به خواندن و مطالعه و ترویج این امر حتی در شبکه‌های اجتماعی عمیقاً در اندیشه بود و به‌طور وسیعی برنامه‌ریزی کرد.

کلیدواژگان: تحلیل داده‌های عظیم؛ خواندن؛ داده‌های جستجویی؛ متن‌کاوی؛ خوشه‌بندی موضوعی.

مقدمه

پیشرفت‌های اخیر در فناوری اطلاعات منجر به تولید حجم عظیمی از اطلاعات توسط کاربران شامل جستجوهای آنلاین در شبکه‌های اجتماعی شده است (لی و همکاران^۱، ۲۰۱۷). شبکه‌های اجتماعی در سرتاسر جهان به صورت منابع مهمی از اطلاعات و محیطی جهت پیروی و گسترش تمایلات، اخبار و ایده‌ها درآمده‌اند (کانگ و لی^۲، ۲۰۱۷). شبکه‌های فرهنگی و اجتماعی همانند سایت موسسه فرهنگی و اطلاع‌رسانی تبیان به صورت مداوم باعث تشویق و جذب هزاران کاربر جهت به اشتراک گذاشتن و تبادل اطلاعات عقاید و دیدگاه‌هایشان و مشارکت در رویدادهای همه‌جانبه آنلاین می‌باشند. هم‌روزه شاهد بارگذاری پست‌ها، نظرات و مطالب بسیار زیادی از سوی مسئولین و کاربران در این شبکه‌ها هستیم و اغلب این اطلاعات به فرمت‌های مختلف متنی ذخیره‌سازی می‌شوند. تعیین علائق و سلیق کاربران بر اساس متون داده‌ای شبکه‌های اجتماعی و همچنین بررسی اهمیت موضوعی مطالب اخیراً به یکی از موضوعات جذاب و مهم تحقیقاتی مبدل شده است (هان و لی^۳، ۲۰۱۴). به‌طور ویژه، داده‌های حاصل از جستجوی کلمات یا عبارات کلیدی در شبکه‌های اجتماعی اطلاعات باارزشی در مورد سلیق، عقاید و افکار کاربران در مورد موضوع یا مطلبی خاص در اختیار می‌گذارد (اینجادات و همکاران^۴، ۲۰۱۶). در جایی که هدف بررسی و تحلیل اهمیت و نقش مطالعه و خواندن در میان کاربران و متصدیان شبکه‌های فرهنگی، اطلاع‌رسانی و اجتماعی است، حجم و میزان محتوای به اشتراک گذاشته‌شده مرتبط در این زمینه می‌تواند بیانگر میزان توجه و علاقه کاربران و اهمیت مقوله کتاب و کتابخوانی در میان مسئولین و متصدیان شبکه باشد. از سوی دیگر سازمان‌ها، ارگان‌ها و همچنین مسئولین علاقه‌مند هستند از قابلیت‌ها و قدرت فراگیر داده‌های عظیم موجود در شبکه‌های اجتماعی جهت اتخاذ تصمیمات مدیریتی و سیاست‌گذاری‌های پیش رو بهره ببرند (جانسن و همکاران^۵، ۲۰۱۷). چالش اصلی در این زمینه انتخاب کلمات یا

1. Li et al.
2. Kang & Lee
3. Han, H. Lee
4. Injadat et al.
5. Janssen et al.

عبارات کلیدی مناسب و جمع‌آوری داده‌های جستجو است (برینجولفسون و همکاران^۱، ۲۰۱۶). همچنین، برخی از شبکه‌های فرهنگی و اجتماعی همانند تیان این قابلیت را دارند که پس از جستجوی کلمات کلیدی مورد پژوهش در آن‌ها، اقدام به طبقه‌بندی موضوعی آن‌ها نموده و از این طریق سهم هر طبقه موضوعی به‌راحتی از قبل کلمه کلیدی جستجو شده قابل استخراج خواهد بود.

پیشینه پژوهش

متن‌کاوی اولین بار توسط فلدمن و همکاران^۲ (۲۰۰۶) مطرح گردید و پس‌از آن به‌سرعت کاربرد و جذابیت زیادی برای پژوهشگران، سیاست‌گذاران و بازاریان جهت کشف دانش نهفته جهت اتخاذ تصمیمات و برنامه‌ریزی‌ها پیدا نمود. به‌ویژه در مهر و موم‌های اخیر مطالعات فراوانی در خارج از کشور در زمینه کاربردهای متن‌کاوی در شبکه‌های مجازی و اجتماعی صورت پذیرفته است. به‌عنوان مثال، لازارد و همکاران^۳ (۲۰۱۵) در پژوهشی با حمایت مرکز حفاظت و کنترل بیماری‌های ایالات متحده آمریکا با استفاده از روش کاوش و تحلیل متن به بررسی میزان گسترش وحشت عمومی از طریق متن‌کاوی توئیت^۴ های زنده میان کاربران نمودند و به این نتیجه رسیدند که افکار عمومی به‌شدت در مورد علائم و بازه زمانی فعال بودن ویروس ابولا^۵، طریقه انتقال ویروس و چگونگی محافظت از بدن در زمان مسافرت نگران است. بانوس و همکاران (۲۰۱۶) در مطالعه‌ای به بررسی نحوه بهبود استراتژی‌های تولید اطلاعات جانبی از قبیل الگوریتم‌های خوشه‌بندی و طبقه‌بندی مؤثر بر اساس مدل احتمالاتی پرداختند و از تکنیک متن‌کاوی جهت کشف دانش ضمنی موجود در متون شبکه‌های اجتماعی استفاده نمودند. این مطالعه تمرکز عمده‌ای بر روی روش‌شناسی متن‌کاوی با تأکید بر تعداد خوشه‌های مورد استفاده، زمان اجرا و دقت فعالیت متن‌کاوی داشته است. کیم و سین^۶ (۲۰۱۶) در پژوهشی ضمن بررسی سامانه‌های شبکه

1. Brynjolfsson et al.
2. Feldman et al.
3. Lazard et al.
4. Tweet
5. Ebola
6. Kim & Sin

جایگاه کتاب و کتابخوانی در سایت... ۱۷۳

اجتماعی مختلف مورد استفاده توسط دانشجویان دوره کارشناسی و کتابداران کتابخانه‌های دانشگاهی جهت کسب اطلاعات و نحوه ارزیابی اطلاعات حاصل از این شبکه‌ها در محیط علمی به مقایسه عادات و روش‌های این دو مجموعه با هم پرداختند تا بتوانند فاصله ایجاد شده میان آن‌ها را موردسنجش قرار دهند. نمونه انتخابی دانشجویان ۱۳۵۵ نفر و نمونه انتخابی کتابداران ۱۸۹ نفر بودند و نتیجه اینکه سامانه شبکه اجتماعی مورد استفاده و هدف از استفاده از این شبکه برای هر دو مجموعه مشابه بوده ولی فاصله زیادی از جهت استراتژی‌های مورد استفاده جهت کسب اطلاعات وجود داشته است. الدایهانی و آبراهامز^۱ (۲۰۱۶) به بررسی پایگاه توئیت های ۱۰ کتابخانه دانشگاهی شامل ۲۳۷۰۷ توئیت، ۱۷۸۴۸ اشارک^۲، ۷۶۲۵ هشتک^۳ و ۵۹۷۴ توئیت مجدد^۴ با استفاده از روش متن‌کاوی پرداخته و بیان نمودند با این شیوه می‌توان به برنامه‌ریزی استراتژیک مناسب و اتخاذ تصمیمات مدیریت مقتضی جهت ارائه خدمات ایده آل به مراجعین روی آورد. بلو - ارگاز و همکاران^۵ (۲۰۱۶) در مطالعه‌ای با استفاده از روش متن‌کاوی به بررسی و پایش مباحث انجمن‌های شبکه اجتماعی توئیت^۶ در ارتباط با تأثیرات مباحث صورت پذیرفته بر تصمیمات اخذ شده توسط سازمان‌های عمومی بهداشت و سلامت پرداختند. کانگ و لی (۲۰۱۶) در مطالعه خود با استفاده از تکنیک متن‌کاوی و خوشه‌بندی و دسته‌بندی متون به بررسی مطالب به اشتراک گذاشته شده توسط کاربران شبکه‌های اجتماعی پرداختند و علائق و سلاقی آن‌ها را مورد تحلیل قرار دادند.

به‌طورکلی مطالعات متن‌کاوی در داخل کشور پیشینه طولانی نداشته و اندک پژوهش‌های صورت پذیرفته در این زمینه دچار ایراداتی از بابت روش‌شناسی و تحلیل نتایج بوده‌اند. کرمی (۱۳۸۶) در مطالعه‌ای مروری به بررسی کاربرد ابزارهای تحلیلگر داده‌کاوی و متن‌کاوی در چابکی سازمان‌های مراقب بهداشتی و درمانی پرداخته است. عظیمی همت و شمس عزت (۱۳۹۴) در پژوهشی ضمن تأکید بر تحقیقات اندک

1. Al-Daihani, Abrahams
2. Mention
3. Hashtag
4. Retweet
5. Bello-Orgaz et al.
6. Twitter

متن‌کاوی در زمینه متون فارسی، به دلیل پیچیدگی‌های ساختاری آن، مروری بر روش‌های اصلی متن‌کاوی و کاربرد آن‌ها در پردازش متون زبان فارسی صورت دادند. اسماعیلی و زاهد (۱۳۹۴) در فعالیتی مروری به‌طور اجمالی و مختصر به معرفی روش‌ها و چالش‌های پیش روی متن‌کاوی در شبکه‌های اجتماعی پرداختند. پیکری، یعقوبی و طاهری (۱۳۹۴) به‌عنوان یکی از محدود مطالعات پژوهشی به زبان فارسی در زمینه متن‌کاوی، با استفاده از تکنیک متن‌کاوی و تحلیل محتوا پدیده فوت مرتضی پاشایی را در شبکه اجتماعی توئیتر مورد مطالعه و بررسی قرار داده و توئیتهای انتشاریافته در این ارتباط شامل ۱۷۷۱۴ توئیتهای را در پنج مقوله رده‌بندی و همبستگی بین رده‌ها را با ویژگی کاربران مورد تحلیل قرار دادند. در پژوهش حاضر با توجه به سهم مورد غفلت قرار گرفته مطالعه و خواندن در میان کاربران شبکه‌های اجتماعی کشور، به کاوش و تحلیل متون حاصل از جستجوی عبارات کلیدی مرتبط با این مقوله همانند "خواندن" و "مطالعه" در سایت موسسه فرهنگی و اطلاع‌رسانی تبیان به‌عنوان بزرگ‌ترین فضای مجازی و اجتماعی کشور پرداخته شده است. در این بررسی تمامی نتایج جستجوی مرتبط شامل کلیه طبقات و دسته‌های ارائه‌شده توسط سایت در قالب جداول اکسل جمع‌آوری شده و پس از مرور و حذف اطلاعات نامرتبط^۱ توسط الگوریتم‌های متن‌کاوی مورد تجزیه و تحلیل قرار گرفته است.

فرآیند متن‌کاوی

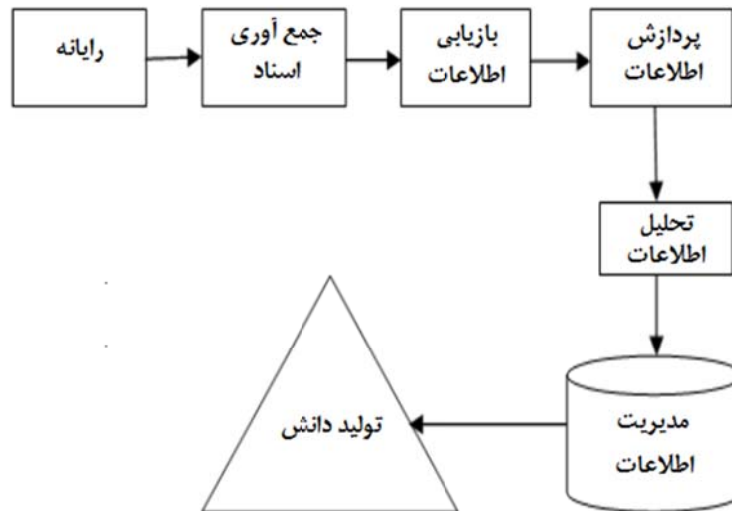
باوجود رشد روزافزون حجم اطلاعات متنی، وجود ابزارهایی جهت سازمان‌دهی، بازیابی و استخراج دانش مفید از این داده‌ها ضروری است. در همین راستا، تاکنون تحقیقات گسترده‌ای روی روش‌های متن‌کاوی صورت گرفته است. تکنیک‌های متن‌کاوی، با استفاده از شناسایی و کشف الگوها، به دنبال استخراج اطلاعات مفید از داده‌های متنی بدون ساختار هستند. متن‌کاوی به‌عنوان داده‌کاوی متن، کشف دانش در متن و همچنین تجزیه و تحلیل هوشمند متن نیز شناخته می‌شود. دلیل اصلی به کار بردن روش‌های داده‌کاوی برای اسناد متنی، ساختاربندی کردن آن‌هاست. روش‌های موجود برای ساختاربندی کردن متون عبارت‌اند از: روش‌های رده‌بندی و روش‌های

خوشه‌بندی. ترکیب روش‌های متن‌کاوی با روش‌های ساختاربندی، ابزارهای قدرتمندی برای کاوش الگوهای مفید در مجموعه‌های متنی فراهم می‌کند. بر اساس فلوجارت شکل ۱ متن‌کاوی و تحلیل داده‌های عظیم حاصل یک فرآیند کشف دانش از طریق پایگاه‌های اطلاعاتی بزرگ همانند شبکه‌های اجتماعی است که هدف آن شناخت الگوهای ناشناخته و تولید دانش است (بانوس و همکاران^۱، ۲۰۱۶). این فرآیند به این صورت تعریف می‌شود:

بازیابی اطلاعات^۲: فرآیند یافتن اسناد حاوی پاسخ به سؤالات پژوهش است و فرآیند یافتن خود پاسخ‌ها را شامل نمی‌شود. به‌منظور رسیدن به این هدف از روش‌های آماری برای پردازش خودکار داده‌های متنی و مقایسه با سؤال‌های ارائه‌شده استفاده می‌گردد. پردازش زبان طبیعی^۳: عبارت از امکان بهره‌گیری از مطالب نوشته شده به زبان‌های طبیعی (همانند انگلیسی و فارسی) و درک و فهم این نوشته‌ها توسط ماشین است. از این تکنیک‌ها می‌توان در خلاصه‌نویسی، متن‌کاوی و پرس‌وجوهای فارسی استفاده نمود.

استخراج اطلاعات^۴: که هدف از آن کسب اطلاعات خاص از اسناد متنی است که این اطلاعات در الگوهای شبیه به پایگاه داده‌ها ذخیره می‌شوند تا در آینده مورد استفاده قرار گیرند.

1. Bhanuse et al.
2. Information Retrieval
3. Natural linguistic processing
4. Information extraction



شکل ۱. فلوچارت روند متن کاوی.

متن کاوی در شبکه‌های اجتماعی یکی از موضوعات داغ در تحقیقات محققین پردازش زبان طبیعی است و با چالش‌هایی نظیر تناقض داده کلان (از طرفی داده‌ها در شبکه‌های اجتماعی کلان هستند و از طرفی ما برای تحلیل کاربر فردی نیاز به اطلاعات اندکی از او داریم)، پیدا کردن نمونه داده مناسب (داده‌های جمع‌آوری شده از شبکه‌های اجتماعی چندان قابل اعتماد نیستند و قوانین نمونه‌برداری در آن‌ها رعایت نمی‌شود. معلوم نیست داده‌ای که برنامه در اختیار ما قرار داده از کجا و در طی چه زمانی جمع‌آوری شده است)، از بین بردن اطلاعات پرت (اضافی) (طریقه حذف اطلاعات اضافی خود مناقشه ایست در کاوش در شبکه‌های اجتماعی) و مشکل ارزیابی (معمولاً از روش‌های سنتی استفاده می‌شود و برای ارزیابی از تقسیم داده به دو گروه آموزش و آزمون کمک می‌گیرند ولی وقتی این شرایط فراهم نباشد کار بسیار دشوار می‌شود) روبرو است (سانگ و همکاران، ۲۰۱۶).

یافته‌های پژوهش

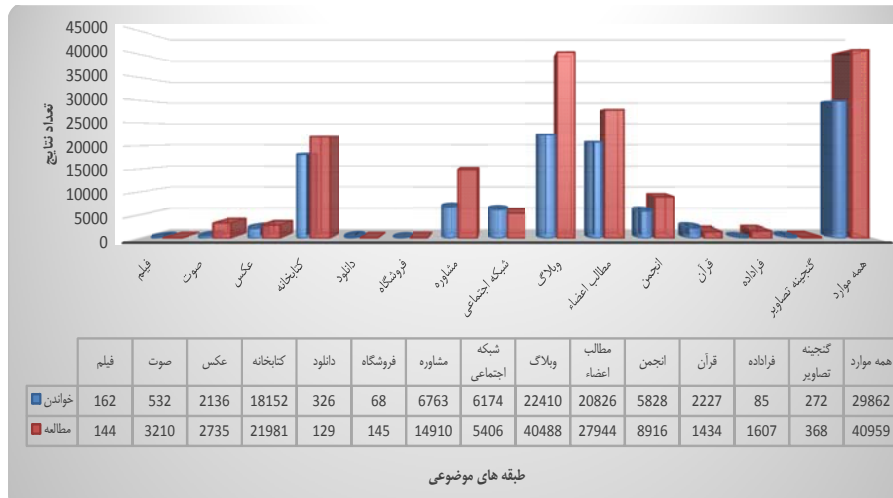
همان‌طور که پیش‌تر گفته شد در این پژوهش جهت بررسی تحلیلی میزان مطالعه در

میان کاربران شبکه‌های اجتماعی، به کاوش و تحلیل متون حاصل از جستجوی عبارات کلیدی مرتبط با این مقوله همانند "خواندن" و "مطالعه" در سایت موسسه فرهنگی و اطلاع‌رسانی تبیان پرداخته شده است. با جستجوی این عبارات به ترتیب ۲۹۸۶۲ و ۴۰۹۵۹ نتیجه حاصل شده است که نتایج در طبقات موضوعی به شرحی که در ادامه آمده قرار گرفته‌اند. همان‌طور که در شکل ۲ مشاهده می‌شود بیشترین نتایج حاصل به ترتیب در طبقات موضوعی وبلاگ، مطالب اعضاء، کتابخانه و مطالب سایت و کمترین نتایج در طبقات موضوعی فروشگاه، فیلم و دانلود قرار گرفته‌اند.

در طبقه فیلم نتایج بر اساس سایز در ۴ دسته از کمتر از ۵۰۰ کیلوبایت تا بیش از ۵ مگابایت، از نظر مدت‌زمان در ۴ دسته از کمتر از ۵ دقیقه تا بیش از ۳۰ دقیقه، از نظر موضوعی در ۶ دسته فیلم و سریال، طنز و سرگرمی، ورزشی، موزیک ویدئو، فیلم‌های آموزشی و فیلم‌های مذهبی قرار می‌گیرند. همچنین، در این طبقه نتایج قابلیت مرتب شدن بر اساس بیشترین دانلود، بیشترین امتیاز، جدیدترین فیلم و زمان بارگذاری فیلم دارند. در طبقه صوت نتایج بر اساس سایز در ۴ دسته از کمتر از ۵۰۰ کیلوبایت تا بیش از ۵ مگابایت و از نظر مدت‌زمان در ۴ دسته از کمتر از ۵ دقیقه تا بیش از ۳۰ دقیقه قرار می‌گیرند. همچنین، در این طبقه نتایج قابلیت مرتب شدن بر اساس بیشترین امتیاز، جدیدترین فیلم و زمان بارگذاری فیلم دارند. در طبقه عکس نتایج بر اساس حداکثر طول در ۸ دسته از کمتر از ۱۰۰ پیکسل تا ۲۲۷۲ پیکسل و از نظر حداکثر ارتفاع در ۷ دسته از ۱۰۰ پیکسل تا ۱۷۰۴ پیکسل قرار می‌گیرند. همچنین، طبقه گنجینه تصاویر از نظر سایز تصویر، حجم تصویر، رنگ تصویر، تاریخ بارگذاری تصویر، HD بودن یا نبودن و Wide بودن یا نبودن قابل دسته‌بندی است. در طبقه کتابخانه نتایج به تفکیک صفحه و تفکیک کتاب شامل عنوان و پدیدآورنده اثر به همراه خلاصه‌ای از متن کتاب قابل ارائه می‌باشند. در طبقه دانلود نتایج بر اساس سایز در ۴ دسته از کمتر از ۲ مگابایت تا بیش از ۱۵ مگابایت، از نظر تاریخ بارگذاری در ۵ دسته از روز جاری تا سال اخیر، از نظر موضوعی در ۴ دسته بازی، موبایل، کاربردی و امنیتی قرار می‌گیرند. علاوه بر این، در این طبقه نتایج قابلیت مرتب شدن بر اساس بیشترین دانلود، بیشترین بازدید، جدیدترین دانلود و محبوب‌ترین دانلود را دارند. در فروشگاه نتایج از نظر گروه کالایی قابل عرضه شامل چندرسانه‌ای، کتاب، کامپیوتر، لپ‌تاپ، موبایل، لوازم ورزشی،

لوازم التحریر، تجهیزات پزشکی و دکوراسیون می‌شود و کاربر قابلیت فیلتر نمودن نتایج برحسب حداقل تا حداکثر قیمت برحسب ریال را دارد. در طبقه مشاوره گروه‌بندی شامل رایانه و فناوری، مشاوره تحصیلی، روانشناسی، سلامتی، علمی و درسی، مذهبی، حقوقی، ورزشی و هنر در خانه شده و از نظر تاریخ به اشتراک‌گذاری نتایج در ۸ دسته از سه روز قبل تا یک سال قبل قرار می‌گیرند. همچنین، در این طبقه قابلیت جستجوی کد سؤال مشاوره‌ای در اختیار کاربر قرار دارد. در طبقه شبکه اجتماعی انواع پست شامل عکس، صوت، فیلم و نظرسنجی وجود داشته و از نظر تاریخ ثبت نتایج در ۸ دسته از روز جاری تا یک سال قبل قرار می‌گیرند. در طبقه وبلاگ می‌توان عبارت کلیدی را در فیلدهایی شامل برچسب‌ها، عنوان نوشته، متن، عنوان وبلاگ و آدرس وبلاگ جستجو نمود. در طبقه انجمن نتایج در قالب گروه‌هایی شامل اجتماعی، فرهنگ و هنر، تندرستی، دانش و فن، مذهبی، کاربران، مدیران انجمن و کسب‌وکار نمایش داده می‌شوند. در طبقه قرآن نتایج جستجو در قالب‌هایی شامل متن آیه، ترجمه و تفسیر نمایش داده می‌شوند. در طبقه فراداده نتایج جستجو از نظر فیلد با توجه به عنوان، پدیدآورندگان، موضوع، ویرایش، وضعیت نشر، فروست، یادداشت و شرح پدیدآورنده؛ از نظر نوع منبع به صورت کتاب یا مقاله و از نظر فرمت به صورت متن یا PDF قابل دسته‌بندی است. در طبقه مطالب اعضاء نتایج جستجو در دسته‌های موضوعی اعتقادی، تحقیقات و فناوری، سیاسی، طنز و سرگرمی، فرهنگی و هنری، سلامتی، اقتصاد، تاریخی، طبیعت، ورزشی، خانواده و حقوق قابل دسته‌بندی است. همچنین، نتایج از طریق جستجوی نام مستعار اعضاء و بر اساس تاریخ ثبت مطلب (در ۸ دسته از روز جاری تا یک سال قبل) قابل‌نمایش می‌باشند.

با توجه به اینکه نتایج موجود در طبقه مطالب سایت به نحو سازمان‌دهی و مرتب‌شده‌ای دسته‌بندی موضوعی شده‌اند، عبارات کلیدی مورد جستجو در این طبقه مورد متن‌کاوی دقیق و تحلیل واقع شده‌اند. در شکل ۳ نتایج حاصل برای طبقه موضوعی مطالب سایت به تفکیک دسته‌بندی‌های مختلف موجود ارائه شده است. بر این اساس بیشترین نتایج حاصل به دسته موضوعی فرهنگ و کمترین نتایج در دسته حوزه علمیه قرار گرفته است که این نتیجه با توجه به ماهیت فرهنگی امر خواندن و مطالعه این نتیجه دور از ذهن نیست.

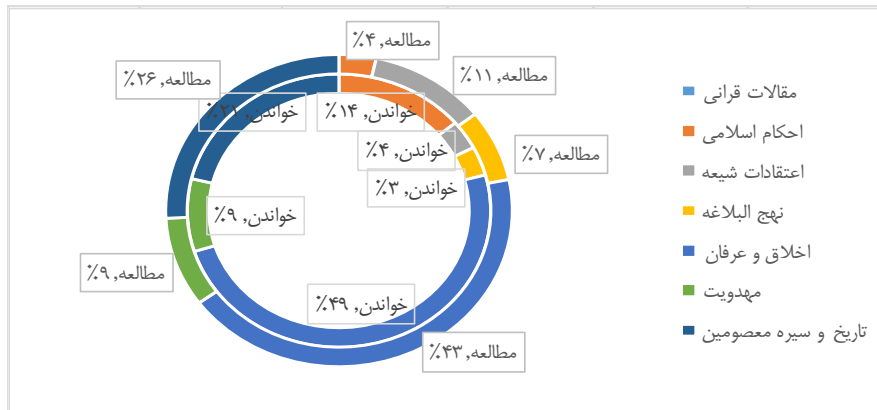


شکل ۲. نتایج جستجوی عبارات کلیدی در طبقات موضوعی.

بر اساس نتایج دسته‌بندی‌های مختلف طبقه موضوعی دین که در شکل ۴ ارائه شده است، دو موضوع اخلاق و عرفان و تاریخ و سیره معصومین از اهمیت زیادی برخوردارند و متأسفانه با وجود حساسیت زیاد موضوع مهدویت در عصر حاضر و نقش بسزای شبکه اجتماعی و اطلاع‌رسانی تبیان؛ این موضوع آن‌چنان‌که شایسته است در میان مسئولین و کاربران این شبکه مورد توجه واقع نشده است.

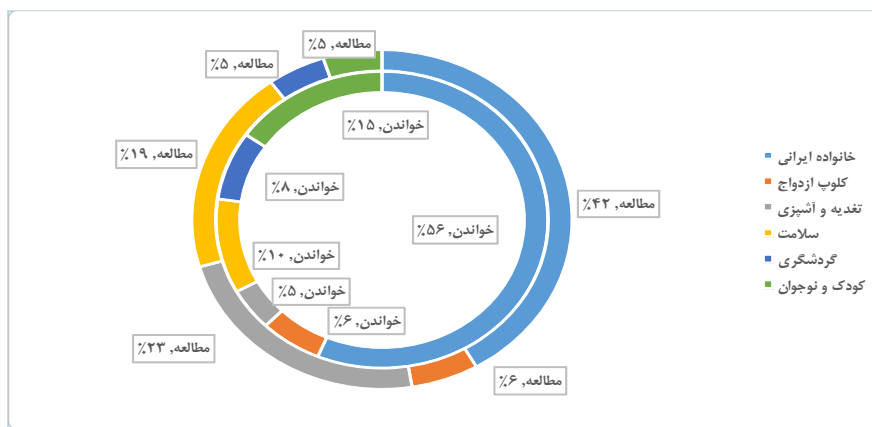


شکل ۳. نتایج حاصل برای طبقه موضوعی مطالب سایت به تفکیک دسته‌بندی‌های مختلف.



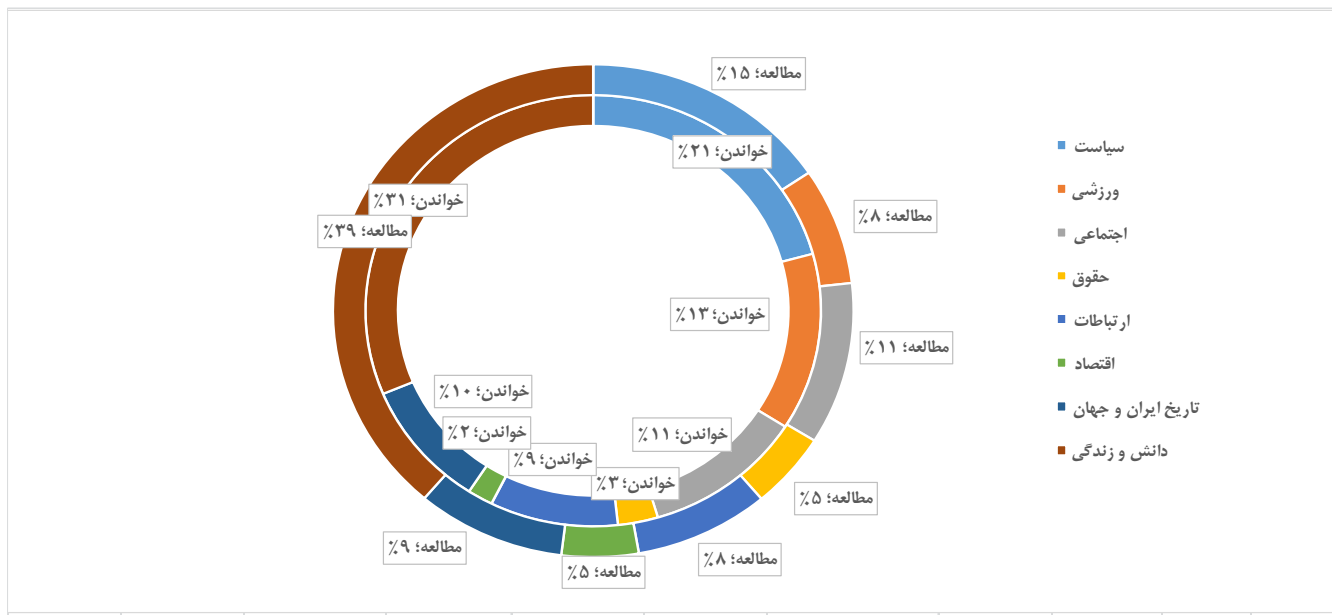
شکل ۴. نتایج دسته‌بندی‌های مختلف طبقه موضوعی دین.

بر اساس نتایج دسته‌بندی‌های مختلف طبقه موضوعی زندگی که در شکل ۵ ارائه شده است، دو موضوع خانواده ایرانی و سلامت در جستجوی عبارات کلیدی خواندن و مطالعه از اهمیت بیشتری برخوردار بوده‌اند ولی متأسفانه با وجود مشکلات عدیده جامعه از نظر عدم تمایل جوانان به مقوله ازدواج، کلپ ازدواج کمترین میزان نتایج حاصل متون کاوش شده را به خود اختصاص داده و فرهنگ‌سازی لازم در شبکه اجتماعی و اطلاع‌رسانی تبیان به‌عنوان بزرگ‌ترین شبکه اجتماعی و فرهنگی کشور صورت نپذیرفته است. همچنین، علیرغم تبلیغات فراوان در زمینه لزوم بهبود وضعیت گردشگری کشور به‌عنوان منبع درآمد غیرنفتی، سهم موضوعی دسته گردشگری در این شبکه بسیار ناچیز است.

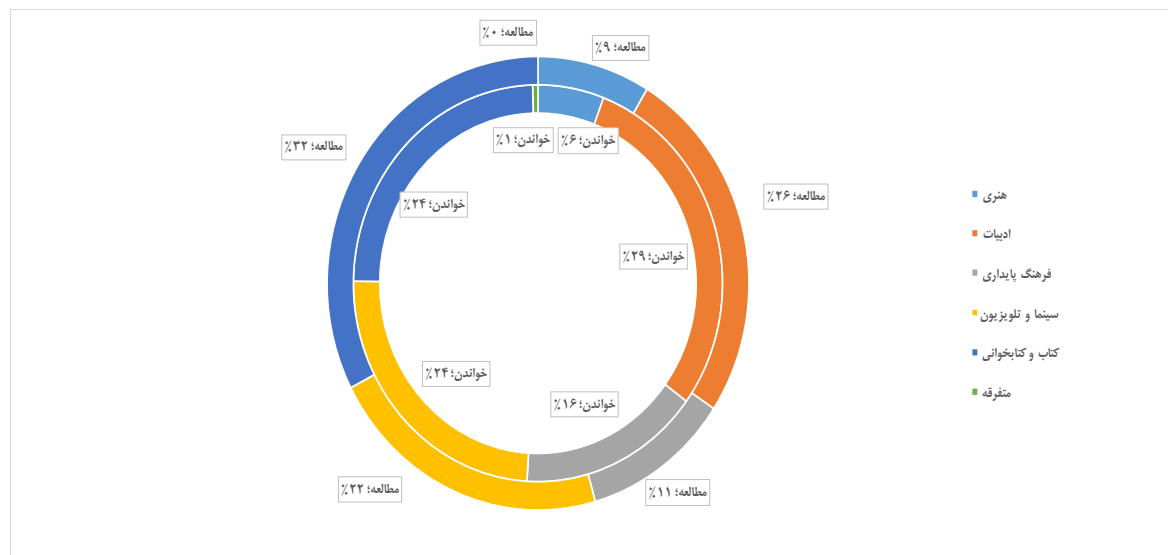


شکل ۵. نتایج دسته‌بندی‌های مختلف طبقه موضوعی زندگی.

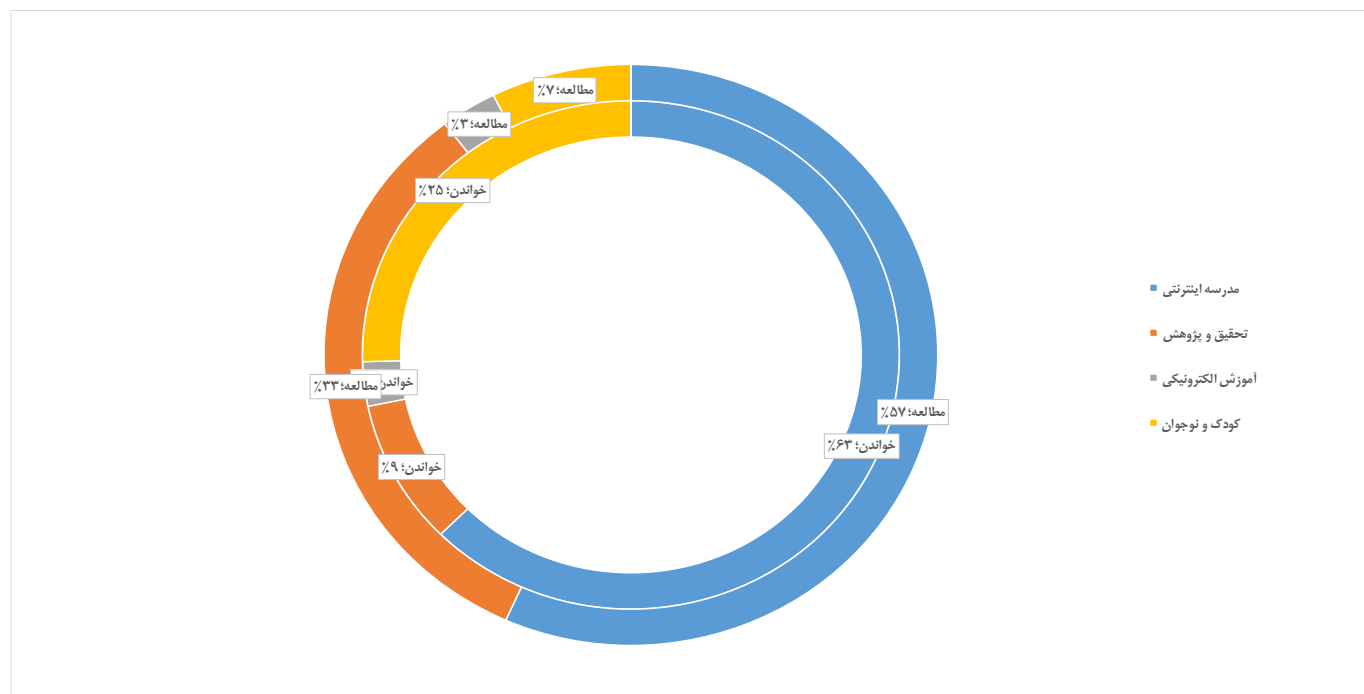
در شکل ۶ نتایج حاصل برای طبقه موضوعی جامعه ارائه شده است. بر این اساس بیشترین نتایج حاصل به دسته‌های موضوعی دانش و زندگی و سیاست و کمترین نتایج در دسته‌های حقوق و اقتصاد قرار گرفته است. به نظر می‌رسد شبکه اجتماعی تبیان علاقه زیادی به ارائه مطالب خواندنی و انتقال اطلاعات در ارتباط با موضوعات حقوقی و اقتصادی با وجود مشکلات عدیده جامعه از این بابت ندارد. همچنین بر اساس نتایج دسته‌بندی‌های مختلف طبقه موضوعی فرهنگ که در شکل ۷ ارائه شده است، به ترتیب سه موضوع کتاب و کتابخوانی، ادبیات و سینما و تلویزیون در جستجوی عبارات کلیدی خواندن و مطالعه از اهمیت بیشتری برخوردار بوده‌اند ولی موضوعات هنری کمترین میزان نتایج حاصل متون کاوش شده را به خود اختصاص داده است. با وجود بالا بودن تعداد نتایج حاصل در طبقه موضوعی جامعه، تعداد نتایج حاصل در کل شبکه برای مقوله کتاب و کتابخوانی بسیار اندک به نظر می‌رسد (تنها به ترتیب ۹۵۲ و ۱۱۹۹ نتیجه در جستجوی عبارت کلیدی خواندن و مطالعه) و این امر لزوم فعالیت بیشتر شبکه اجتماعی تبیان در راستای اعتلای این فرهنگ مغفول مانده در کشور را نمایان می‌سازد. در شکل ۸ نیز نتایج حاصل برای طبقه موضوعی مرکز یادگیری ارائه شده است. بر این اساس بیشترین نتایج حاصل به دسته موضوعی مدرسه اینترنتی و کمترین نتایج در دسته آموزش الکترونیکی قرار گرفته است. به نظر می‌رسد با توجه به ماهیت مجازی و الکترونیکی بودن سایت تبیان، لازم است بخش آموزش‌های الکترونیکی این سایت با تولید محتوای آموزشی بیشتر و حتی ترجمه و عرضه محتواهای آموزشی الکترونیکی خارجی این کمبود را جهت بالا بردن سطح آموزشی و علمی کاربران خود پوشش دهد.



شکل ۶. نتایج دسته‌بندی‌های مختلف طبقه موضوعی جامعه.



شکل ۷. نتایج دسته‌بندی‌های مختلف طبقه موضوعی فرهنگ.



شکل ۸. نتایج دسته‌بندی‌های مختلف طبقه موضوعی مرکز یادگیری.

بحث و نتیجه‌گیری

رشد فزاینده پایگاه داده‌ها در تقریباً هر ناحیه از فعالیت انسان باعث شده است که نیاز برای ابزارهای قدرتمند جدید برای تغییر دادن داده به دانش مفید افزایش یابد. برای برآوردن این نیاز محققان در ناحیه‌های مختلف مانند یادگیری ماشین، شناسایی الگو، آنالیز داده آماری، بصری سازی داده، شبکه‌های عصبی، اقتصادسنجی، بازیابی اطلاعات، استخراج اطلاعات و... روش‌ها و ایده‌هایی را کاوش کرده‌اند. ذات زیرساخت یافته‌ی متون، حاصل از این پایگاه‌های اطلاعاتی اعمال همان روش‌هایی را که ما در مورد پایگاه داده‌های غیرمتنی بکار می‌بریم، غیرممکن می‌سازد؛ بنابراین، روش‌ها و الگوریتم‌های پردازش (پیش‌پردازش) خاصی برای استخراج الگوهای مفید موردنیاز است. متن‌کاوی اطلاعات متنی غیر ساخت‌یافته را استفاده می‌کند و آن را برای کشف ساختار و معناهای ضمنی پنهان در متن بررسی می‌کند. با توجه به تولید میلیاردها متن در شبکه‌های اجتماعی به‌صورت روزانه، پژوهشگران نسبت به استخراج دانش از این متون علاقه‌مند شده‌اند. به همین دلیل متن‌کاوی در شبکه‌های اجتماعی موضوعی داغ در حوزه داده‌کاوی تبدیل شده است. در این پژوهش با توجه به سهم مورد غفلت قرار گرفته مطالعه و خواندن در میان کاربران شبکه‌های اجتماعی کشور، به کاوش و تحلیل متون حاصل از جستجوی عبارات کلیدی مرتبط در سایت موسسه فرهنگی و اطلاع‌رسانی تبیان پرداخته شده است.

نتایج حاکی از آن است که با توجه به ماهیت فرهنگی این موسسه و امر خواندن و مطالعه، جستجوی عبارات کلیدی مرتبط بیشترین نتایج را در طبقه موضوعی فرهنگ در اختیار خواهد گذاشت. با وجود حساسیت زیاد موضوع مهدویت در عصر حاضر؛ این موضوع آن‌چنان‌که شایسته است در میان مسئولین و کاربران این شبکه موردتوجه واقع نشده است. به نظر می‌رسد با توجه به ماهیت مجازی و الکترونیکی بودن سایت تبیان، لازم است بخش آموزش‌های الکترونیکی این سایت با تولید محتوای آموزشی بیشتر و حتی ترجمه و عرضه محتوای آموزشی الکترونیکی خارجی این کمبود را جهت بالا بردن سطح آموزشی و علمی کاربران خود پوشش دهد. با وجود بالا بودن تعداد نتایج حاصل در طبقه موضوعی جامعه، تعداد نتایج حاصل در کل شبکه برای مقوله کتاب و

کتابخوانی بسیار اندک به نظر می‌رسد (تنها به ترتیب ۹۵۲ و ۱۱۹۹ نتیجه در جستجوی عبارت کلیدی خواندن و مطالعه) و این امر لزوم فعالیت بیشتر شبکه اجتماعی تبیان در راستای اعتلای این فرهنگ مغفول مانده در کشور را نمایان می‌سازد. نتایج این پژوهش همانند مطالعه انجام‌شده توسط بانوس و همکاران (۲۰۱۶) و کیم و سین (۲۰۱۶) نشان داد که استفاده از روش‌های متن‌کاوی می‌تواند ابزار مناسبی جهت تشخیص میزان علائق و سلايق مخاطبين و همچنين برآورد روند تغييرات موضوعی رویکردهای ایشان در شبکه‌های اجتماعی باشد. علاوه بر این همان‌طور که از نتایج این پژوهش و بر طبق مطالعات گسترده صورت پذیرفته در خارج از کشور توسط الدایهانی و آبراهامز (۲۰۱۶)، بلو-ارگاز و همکاران (۲۰۱۶) و کانگ و لی (۲۰۱۶) و مطالعه داخلی پیکری، یعقوبی و طاهری (۱۳۹۴) برمی‌آید، می‌توان با استفاده از شیوه متن‌کاوی شبکه‌های اجتماعی پرمخاطب؛ به برنامه‌ریزی استراتژیک مناسب و اتخاذ تصمیمات مدیریتی مقتضی جهت ارائه خدمات ایده آل به مراجعین و مخاطبین روی آورد. به‌عنوان پژوهش‌های آتی پیشنهاد می‌شود اطلاعات حاصل از این تحقیق از طریق تهیه پرسشنامه و کسب رویکردهای کاربران شبکه اجتماعی تبیان مورد تحقیق و صحت‌سنجی قرار گیرد. همچنین، تحقیقات پیرامون محبوب‌ترین موضوعات میان کاربران و روندهای تغییر در گرایش به موضوعات مختلف می‌تواند در تصمیم‌گیری‌ها و سیاست‌گذاری‌ها و برنامه‌ریزی‌های آینده مسئولین فضای مجازی مؤثر باشد. علاوه بر این، انجام تحقیقاتی مشابه در زمینه متن‌کاوی و تحلیل محتوا در سایر شبکه‌های اجتماعی فراگیر در کشور نظیر تلگرام و واتساپ می‌تواند بسیار جذاب باشد.

منابع

- اسماعیلی، مهدی، زاهد، عطیه. (۱۳۹۴). مروری بر متن‌کاوی؛ تکنیک‌ها و چالش‌ها، سومین کنفرانس بین‌المللی پژوهش‌های کاربردی در مهندسی کامپیوتر و فناوری اطلاعات. پیکری، ناصر، یعقوبی، سیدعلی اصغر، طاهری، حمیدرضا. (۱۳۹۴). تحلیل احساسات در شبکه اجتماعی توئیتر با تکنیک متن‌کاوی، کنفرانس بین‌المللی وب پژوهی. عظیمی همت، منیره، و فاطمه. شمس عزت، (۱۳۹۴). مروری بر متن‌کاوی متون فارسی، دومین کنفرانس بین‌المللی و سومین همایش ملی کاربرد فناوری‌های نوین در علوم مهندسی، مشهد، دانشگاه تربت حیدریه، دانشگاه فردوسی مشهد.
- کرمی، مهتاب. (۱۳۸۶). کاربرد ابزارهای تحلیلگر داده‌کاوی و متن‌کاوی در چابکی سازمان‌های مراقب بهداشتی و درمانی، فصلنامه علمی - پژوهشی مدیریت سلامت، ۱۰ (۳۰)، ۱۵-۲۱.
- Sultan M. Al-Daihani, Alan Abrahams. (2016). A Text Mining Analysis of Academic Libraries' Tweets, *the Journal of Academic Librarianship*.
- G. Bello-Orgaz, J. Hernandez-Castro, D. Camacho, (2016). *Detecting discussion communities on vaccination in twitter*, Future Generation Computer Systems.
- Brynjolfsson, E., Geva, T., & Reichman, S. (2016). Crowd-squared: Amplifying the predictive power of search trend data. *MIS Quarterly*, 40(3).
- Shraddha S.Bhanuse, Shailesh D.Kamble, Sandeep M. Kakde. (2016). Text Mining using Metadata for Generation of Side information, *Procedia Computer Science*, 78, 807 – 814.
- Feldman, R., and Sanger, J. (2006). *The Text Mining Handbook*. New York: Cambridge University Press. ISBN 978-0-521-83657-9.
- Marijn Janssen, Haiko van der Voort, Agung Wahyudi. (2017). Factors influencing big data decision-making quality, *Journal of Business Research*, Volume 70, January 2017, Pages 338–345.
- Jaeyong Kang, Hyunju Lee, (2017). Modeling user interest in social media using news media and Wikipedia, *Information Systems*, Volume 65, April 2017, Pages 52–64.
- Kyung-Sun Kim, Sei-Ching Joanna Sin. (2016). Use and Evaluation of Information from Social Media in the Academic Context: Analysis of Gap between Students and Librarians, *the Journal of Academic Librarianship*, 42, 74–82.
- Allison J. Lazard, Emily Scheinfeld, Jay M. Bernhardt, Gary B. Wilcox, Melissa Suran. (2015). detecting themes of public concern: A text mining

analysis of the Centers for Disease Control and Prevention's Ebola live Twitter chat, *American Journal of Infection Control*, 1-3.

J. Han, H. Lee. (2014). characterizing user interest using heterogeneous media, in: *Proceedings of the 23rd international conference on World Wide Web*, pp. 289-290.

Xin Li, Bing Pan, Rob Law, Xiankai Huang, (2017). Forecasting tourism demand with composite search index, *Tourism Management*, 59, 57-66.

Juyoung Song, Tae Min Song, Dong-Chul Seo, Jae HyunJin. (2016). Data Mining of Web-Based Documents on Social Networking Sites That Included Suicide-Related Words among Korean Adolescents, *Journal of Adolescent Health*, Volume 59, Issue 6, Pages 668-673.

MohammadNoor Injadat, Fadi Salo, Ali Bou Nassif. (2016). Data mining techniques in social media: A survey, *Neurocomputing*, Volume 214, Pages 654-670.