

Sentiment and Semantic Analysis of the Performance and Content of News Websites and Social Media with a Focus on Health Messages

By: Masoud Ghayoomi Ph.D.✉

Abstract:

One of the most important issues in any society is attention to health, which is regarded as an indicator of development. Planning in this area should be based on scientific and practical criteria. Health communication is one dimension of health-related issues that focuses on informing and raising public awareness. The transformation of communication, which today is predominantly based on digital communication, has changed the process of health communication and public awareness through health messages, making such messages accessible via news websites and social media. This study examines health-related messages during the one-year period from 2019 to 2020, corresponding to the COVID-19 pandemic, drawing on two sources of information dissemination—news websites and social media—and conducts sentiment and semantic analyses of COVID-19-related two-word expressions (bigrams). The study employed both quantitative and qualitative methods to analyze linguistic data. Accordingly, it relied on the compilation of a linguistic corpus from news websites and social media during the aforementioned period. Following the collection and normalization of the data and the selection of health messages related to COVID-19, high-frequency COVID-19-related bigrams were extracted, and the data were subjected to statistical and linguistic analyses. The findings showed that: (a) as the dissemination of health messages decreased, the mortality rate of the pandemic increased; (b) health messages predominantly conveyed negative sentiment; (c) the intensity of negative sentiment conveyed through health messages was greater on news websites than on social media; (d) social media employed greater lexical diversity to express negative sentiment; and (e) the lexical similarity between the content words used and COVID-19-related keywords indicated that, in health messages published on news websites, specialized terms from the fields of “medical, biological, and health sciences” were more prevalent, whereas on social media, greater semantic similarity was observed among words related to “threat, danger, and negative consequences.” Accordingly, this study suggests that health messages should be disseminated in language capable of counteracting the negative sentiment conveyed by such messages, thereby establishing effective communication with audiences while fostering hope and positive feelings. Furthermore, these messages should be expressed in simple, non-specialized language that can effectively communicate with and engage the audience.

Keywords: *Health Communication, Health Messages, Pandemic, COVID-19, Corpus Linguistics*

✉ Associate of Prof. in Linguistics Research, Institute of Humanities and Cultural Studies, Tehran, Iran
Email: m.ghayoomi@ihcs.ac.ir

Received: 2026/05/02

Accepted: 2026/07/12

DOI: 10.22082/CR.2026.2087813.2921

تحلیل احساسی و معنایی عملکرد و محتوای وبگاه‌های خبری و شبکه‌های اجتماعی با محوریت پیام سلامت

✉ مسعود قیومی

چکیده

یکی از مسائل بسیار بااهمیت، در هر جامعه، توجه به سلامت است که به‌عنوان نشانه توسعه‌یافتگی در نظر گرفته می‌شود. برنامه‌ریزی در این حوزه، باید بر مبنای معیارهای علمی و کاربردی صورت پذیرد. ارتباطات سلامت یکی از ابعاد این موضوع است که به اطلاع‌رسانی و آگاهی‌بخشی عمومی می‌پردازد. تحول در ارتباطات که امروزه به‌طور عمده بر مبنای ارتباطات دیجیتال جریان دارد، فرایند ارتباطات سلامت و آگاهی‌بخشی با استفاده از پیام‌های مرتبط را دستخوش تغییر کرده و این پیام‌ها را از طریق وبگاه‌های خبری یا شبکه‌های اجتماعی در دسترس قرار داده است. در این پژوهش می‌کوشیم با بررسی پیام‌های سلامت‌محور در یک بازه زمانی یک‌ساله از ۱۳۹۸ تا ۱۳۹۹ در دو منبع انتشار اطلاعات، وبگاه‌های خبری و شبکه‌های اجتماعی به تحلیل احساسی و معنایی عبارات دو واژه‌ای کرونایی در دوران همه‌گیری ویروس بپردازیم.

این تحلیل، مبتنی بر بررسی کمی و کیفی داده‌های زبانی به‌عنوان شواهد بوده و به این ترتیب، تهیه پیکره زبانی در بازه زمانی یادشده از وبگاه‌های خبری و شبکه‌های اجتماعی ضرورت داشته است. پس از گردآوری و هنجارسازی داده‌ها و انتخاب پیام‌های سلامت مرتبط با بیماری کرونا، دو واژه‌ای‌های پرسامد کرونایی استخراج شده‌اند و تحلیل آماری و زبانی داده‌ها صورت گرفته است.

برخی از مهم‌ترین یافته‌های پژوهش عبارت از این موارد بوده‌اند: الف) با کاهش انتشار پیام‌های سلامت، میزان مرگ و میر همه‌گیری افزایش یافته است. ب) با پیام‌های سلامت به‌طور عمده حس منفی منتقل می‌شود. ج) شدت انتقال حس منفی از پیام‌های سلامت در وبگاه‌های خبری، بیشتر از شبکه‌های اجتماعی است. د) شبکه‌های اجتماعی از تنوع واژگانی بیشتری برای ابراز حس منفی استفاده می‌کنند. ه) تشابه واژگانی واژه‌های محتوایی به کار رفته با کلیدواژه‌های کرونایی مبین این نکته است که در پیام‌های سلامت وبگاه‌های خبری، واژه‌های تخصصی در حوزه «علوم پزشکی، زیستی و بهداشت» کاربرد بیشتری دارند در حالی که در شبکه‌های اجتماعی، تشابه معنایی در واژه‌های مربوط به «تهدید، خطر و پیامدهای منفی» بیشتر است. در نهایت، با بررسی یافته‌های پژوهش می‌توان نتیجه‌گرفت پیام‌های سلامت باید به زبانی ساده و غیر تخصصی و به شیوه‌ای که بر حس منفی محتوای آنها غالب باشد، منتشر شوند تا ارتباطی مؤثر همراه با امید و حس مثبت، با مخاطب برقرار کنند.

واژه‌های کلیدی: ارتباطات سلامت، پیام سلامت، همه‌گیری، کرونا، زبان‌شناسی پیکره‌ای

مقدمه

توجه به سلامت در هر جامعه، از اهمیت بسیار زیادی برخوردار است و از نشانه‌های توسعه‌یافتگی تلقی می‌شود. این موضوع در بند ۱۲ امور زیربنایی سیاست‌های کلی برنامه پنج‌ساله هفتم^۱ کشور ما، ایران نیز مطرح شده‌است. سلامت حوزه‌ای علمی و تخصصی است که هرگونه برنامه‌ریزی در آن باید بر مبنای معیارهای علمی و کاربردی صورت پذیرد؛ درعین حال، مفهومی پیچیده و چندبعدی است که علاوه بر حوزه بهداشت و پزشکی، می‌تواند در چهارچوب ارتباطات سلامت نیز مورد توجه قرار گیرد.

تحولات فناورانه اجتماعی، انتظارات رسانه را دستخوش تغییر کرده و بر رویکردهای ارتباطات سلامت تأثیرگذار بوده است. تحول شتابان در حوزه‌های ارتباطات و سلامت، چشم‌انداز تولید و انتشار پیام‌های مرتبط را نیز به‌طور اساسی متحول کرده است تا بتواند به بهبود سواد سلامت منجر شود. بهبود سواد سلامت برای افزایش توانمندسازی حیاتی است که می‌تواند در تولید محتوا، روش آموزش و ارتباطات سلامت اثرگذار باشد (ناتیم، ۲۰۰۰). ارتباطات سلامت حوزه‌ای میان‌رشته‌ای است که از دانش‌هایی مانند پزشکی، روان‌شناسی، جامعه‌شناسی، زبان‌شناسی و علوم ارتباطات بهره می‌گیرد. از آنجاکه سلامت فردی و جمعی، به یکی از دغدغه‌های جدی بشر تبدیل شده است و حتی در امریکا، از سواد پایین سلامت با عنوان «قاتل خاموش»^۳ یاد می‌شود، تولید پیام‌های سلامت برای افزایش آگاهی عمومی، به‌عنوان موضوعی میان‌رشته‌ای در چهارچوب ارتباطات سلامت از اهمیت شایانی برخوردار است.

کرونا به‌عنوان یک بیماری همه‌گیر، تأثیرات مختلفی بر کشورهای جهان داشته که بارزترین آنها در حوزه سلامت، بهداشت و درمان بوده است. یکی از مهم‌ترین مسائلی که در این حوزه، به‌هنگام مواجهه با همه‌گیری (پاندمی) باید به آن پرداخته شود، اطلاع‌رسانی به افراد جامعه در قالب تولید پیام‌های سلامت‌محور است. ممکن است علاوه بر اطلاع‌رسانی، اخبار و اطلاعات دیگری نیز در مورد این موضوع منتشر شود که از جنبه‌های مختلف در خور بررسی است. یکی از مهم‌ترین این جنبه‌ها، مربوط به حوزه زبان‌شناسی است؛ چراکه اساساً زبان به‌عنوان ابزار ارتباطی بین افراد جامعه مطرح است.

1. <https://rc.majlis.ir/fa/law/show/1796652>

2. Nutbeam

3. silent killer

در پژوهش حاضر تلاش شده است تا الگوهای زبانی که در تولید پیام سلامت در دوران شیوع بیماری کرونا در سال‌ها ۱۳۹۸ تا ۱۴۰۱ کارایی داشته‌اند، از پیکره زبانی مربوط به کرونا که ترکیبی از دو گونه زبانی رسمی، در اخبار خبرگزاری‌ها و غیررسمی، در دیدگاه‌های شبکه‌های اجتماعی اینستاگرام و توییتر بوده است، استخراج و بررسی شوند. از آنجاکه امروزه بسیاری از الگوهای زبانی، از فضای مجازی به زبان رسمی منتقل می‌شوند (مانند گرایش به هشتگ^۱ زنی) ضروری است که کاربرد زبان نیز در فضای مجازی مورد بررسی قرار گیرد تا تحلیل‌های دقیق‌تری ارائه شود. در همه‌گیری کرونا که در این بازه سه‌ساله بر کشور ایران حاکم بود، موضوع پیام‌های سلامت‌محور برای افزایش آگاهی جمعی نقش بسیار مهمی را برعهده داشت. همچنان که تولید محتوای پیام‌های سلامت ادامه می‌یافت، با گسترش و پیچیده‌تر شدن وضعیت همه‌گیری، محتوای پیام‌ها نیز توسعه پیدا می‌کرد، مانند پیام‌های زیر که در قالب هشتگ در رسانه و شبکه‌های اجتماعی مطرح بود و عظیمی‌فرد (۱۳۹۹) به آنها اشاره کرده‌است:

- کرونا را شکست می‌دهیم.

- به مدد الهی؛ کرونا را شکست می‌دهیم.

- در خانه بمانید؛ کرونا را شکست می‌دهیم.

- در خانه بمانیم؛ کرونا را شکست می‌دهیم.

- در خانه می‌مانیم؛ به مدد الهی؛ کرونا را شکست می‌دهیم.

در پژوهش حاضر، سعی بر این بوده است که در یک بررسی پیکره‌بنیان، به تحلیل کمی و کیفی پیام‌های سلامت‌محور کرونا، از جنبه بیان احساسات در عبارات کرونایی و همچنین، خوشه‌بندی معنایی این عبارات و هشتگ‌های مرتبط با آنها، در وبگاه‌های خبری و شبکه‌های اجتماعی پرداخته شود.

پیشینه پژوهش

در حوزه پیام‌های سلامت‌محور، پژوهش‌های متنوعی از نظر موضوع انجام شده‌اند که به صورت فشرده به آنها می‌پردازیم.

موریسون و همکاران^۲ (۲۰۰۵) به تحلیل ساختار و محتوای پیام‌های سلامت پرداخته‌اند. در این مطالعه، ۷۲ پیام سلامت از منابع معتبر، اخذ و از افراد

1. Hash tag

2. Morrison et al.

متخصص خواسته شده به بازتولید ۱۰ پیام تصادفی بپردازند. پیام‌ها از نظر گفتمانی، بررسی شده‌اند و میزان تطابق معیار کاپای آنها ۰/۶۹ به دست آمده است: نتایج حاصل نشان داده است که ساختار پیام‌های سلامت قابلیت بازتولید دارد و می‌تواند برای تهیه ابزاری در این حوزه کارایی داشته باشد.

ادواردی^۱ (۲۰۱۹) به بررسی نقش زبان در ارتباطات سلامت محور میان بیمار و درمانگر پرداخته است. یافته‌ها بیانگر این نکته است که توجه به زبان‌شناسی در ارتباطات سلامت محور میان بیمار و درمانگر، بر نقش ارتباطی مدیریت سلامتی که متمرکز بر ارتباطات با بیمار است، تأثیرگذار است. دانش زبان‌شناسی نشان می‌دهد که معنا چگونه به صورت صریح یا ضمنی از طریق کدگذاری و کدگشایی بیان می‌شود. از نظر انتخاب واژه‌ها، ساختار و نوای گفتار، چگونه می‌توان معنای موردنظر را به بیمار منتقل کرد. داده‌های این پژوهش از طریق مصاحبه با بیماران به دست آمده است.

لی و همکاران^۲ (۲۰۲۵) به بررسی تأثیر سبک زبانی و ویژگی‌های آن، ویژگی‌های شخصیتی واکنش روانی فرد و متقاعدسازی در حق انتخاب تزریق واکسن آنفولانزا پرداخته‌اند. یافته‌های آنان نشان می‌دهد که سبک زبانی و ویژگی‌ها و واکنش‌های روانی فرد، اعتماد و حق انتخاب تزریق یک یا چندبار، در فرایند ترغیب‌سازی فرد به تزریق واکسن اهمیت شایانی دارد.

خانیکی و راثی‌تهرانی (۱۳۹۰) به بررسی نگرش مخاطبان درباره آن دسته از پیام‌های سلامت محور تلویزیون پرداخته‌اند که ممکن است منجر به تغییر رفتار مخاطب شود. در این پژوهش، آنان پیام‌های سلامت را به دو دسته، پیام‌های سلامت مستقیم و غیرمستقیم تقسیم کرده‌اند و با جمع‌آوری اطلاعات از ۶۰۰ نفر تهرانی بالای ۱۵ سال، تأثیرگذاری آنها را سنجیده‌اند. یافته‌ها نشان می‌دهد که پیام‌های سلامت تلویزیون برای تغییر رفتار فرد کافی نیستند و ارتباطات میان‌فردی، نقش مکمل را برای پیام رسانه‌ای برعهده دارد. در ادامه، پژوهشگران، بهترین الگو را برای ارائه پیام‌های سلامت در تلویزیون، شامل برنامه‌های مستند و گزارش‌های تصویری با حضور متخصصان، کارشناسان، صاحب‌نظران و استادان باتجربه دانسته‌اند.

1. Udvardi

2. Li et al.

رضوی طوسی و همکاران (۱۳۹۲) پیام‌های سلامت و پزشکی در رسانه و تأثیر آن بر آگاهی مردم را بررسی کرده‌اند. برای جمع‌آوری داده‌ها در این پژوهش، پرسشنامه‌ای در اختیار ۸۲۵ نفر از افراد بالای ۱۵ سال قرار گرفته است. حدود ۹۱ درصد از پاسخگویان رضایت زیاد و خیلی‌زیادی از پخش پیام‌های سلامت از تلویزیون داشته‌اند. به این ترتیب، یافته‌ها نشان می‌دهد، تلویزیون نقش مهمی در ایجاد، ترغیب، تشویق و تلقین برخی رفتارها، عادات و گرایش‌ها در حوزه سلامت ایفا می‌کند.

احمدی و المظهری (۱۳۹۷) به بازنمایی پیام‌های سلامت‌محور و تحلیل مضمونی مطالب در روزنامه‌های ایران، همشهری و جام‌جم پرداخته‌اند. در این پژوهش، موضوع‌های مطرح‌شده در اخبار و گزارش‌های این سه منبع، به‌طور دستی، بررسی و براساس طبقه‌بندی سه‌گانه پیام‌های سلامت محور نات‌بیم (۲۰۱۵) یعنی پیام‌های آگاهی‌بخش، آموزشی و ترغیب‌کننده، تحلیل شده‌اند. براین اساس، ارائه راه‌حل ساده برای کاهش تهدید که جزء پیام‌های ترغیب‌کننده است، بیشترین هدف را در روزنامه‌ها دنبال می‌کرده و پیام‌های سلامت، به سه دسته آگاهی‌بخش، آموزشی و اقناعی تقسیم‌بندی شده‌اند. مطالعات صورت‌گرفته‌ای که در بالا به آنها اشاره شد، مواردی هستند که به اهمیت کاربرد دانش زبان‌شناسی در پیام‌های سلامت‌محور و همچنین ساختار و بازنمایی این دسته از پیام‌ها اشاره کرده‌اند. در ادامه، به بررسی پژوهش‌هایی می‌پردازیم که با محوریت بیماری کرونا، به مطالعه در حوزه پیام‌های سلامت‌محور پرداخته‌اند.

چاکرابورتی و همکاران^۱ (۲۰۲۰) به بررسی دو دسته از توییت‌های منتشرشده به زبان انگلیسی درمورد کووید ۱۹ در بستر شبکه اجتماعی توییتر، از نظر احساسات پرداخته‌اند. براساس نتایج به‌دست‌آمده، بیشترین حجم توییت‌ها، دارای بار احساسی مثبت یا خنثی بوده است. با وجود این، توییت‌های دارای حس منفی، ویژگی بازنشرشدگی بیشتری نسبت به توییت‌های دارای حس مثبت یا خنثی داشته‌اند.

زوکودی و مک‌کینی^۲ (۲۰۲۱) به ارزیابی پیام‌های سلامت براساس تحلیل احساسات پرداخته‌اند. براساس نتایج این مطالعه، تفاوت‌های فردی و جنسیتی در واژه‌های به‌کاررفته، بر محتوای پیام سلامت اثرگذار است.

1. Chakraborty et al.

2. Szkody & McKinney

بولوت و پوث^۱ (۲۰۲۲) به تحلیل احساسات و عواطف در دو موج همه‌گیری کرونا در سال ۲۰۲۰ در آلبرتا، کانادا پرداخته‌اند. براساس این مطالعه، انتقال حس مثبت و منفی، با واژه‌های خاصی برای انتقال پیام سلامت کرونا همراه است. حس مثبت، در موج اول غالب بوده و با ورود به موج دوم، از شدت آن کاسته شده است. حس منفی نیز به همان نسبت کاهش یافته است. حس اعتماد و انتظار در موج دوم، نسبت به موج اول کاهش یافته است.

سلحشوری و همکاران (۲۰۲۴) به اندازه‌گیری کیفیت ۲۰ پیام سلامت مربوط به کرونا، از نظر نوع محتوا، صحت پیام، روانی و سادگی آن، طراحی بصری محتوا و شواهد علمی در وبگاه علوم پزشکی دانشگاه جندی شاپور اهواز پرداخته‌اند. براساس تحلیل صورت گرفته، ۷۳/۴ درصد پیام‌ها، بر پیشگیری و بقیه، بر مدیریت این بیماری متمرکز بوده‌اند. ۸۹ درصد پیام‌ها خطاب به عموم و بقیه، خطاب به بیماران منتشر شده‌اند. ۴۲ درصد پیام‌ها بر فاصله‌گذاری اجتماعی، ۳۸/۵ درصد بر ضدعفونی کردن دست‌ها و ۱۶/۵ درصد بر واکسیناسیون متمرکز بوده‌اند. ۴۵ درصد پیام‌ها لحن خبری، ۲۳ درصد لحن متقاعدسازی و ۹/۲ درصد لحن هشدار داشته‌اند. بیشترین و کمترین توجه پیام‌ها به سلامت جسمی (۴۶/۸ درصد) و سلامت روانی (۲/۸ درصد) معطوف بوده است.

عظیمی‌فرد (۱۳۹۹) در پژوهش خود، به بررسی پیام‌های سلامت‌محور در مورد کرونا پرداخته است. وی در یک تحلیل مقایسه‌ای، پیام‌ها و پوستره‌های سلامت‌محور فارسی را که در مورد روش‌های انتقال بیماری، توصیف علائم و رعایت فاصله‌گذاری هشدار داده‌اند با پیام‌های سلامت‌محور تهیه‌شده به زبان فرانسه و انگلیسی مقایسه کرده است. یافته‌های این پژوهش نشان می‌دهد که اغلب پیام‌های سلامت‌محور منتشرشده در ایران، جنبه آگاهی‌بخشی و آموزشی داشته‌اند و کمتر اقناعی بوده‌اند، ضمن اینکه در تهیه این پیام‌ها، اغلب مباحث زبان‌شناختی نادیده گرفته شده و پیام‌های مخاطب‌محور و مشارکتی محدود بوده است.

نکته قابل توجه در مورد پژوهش‌های زبان‌شناسی پیام‌های سلامت‌محور به‌طور کلی و در حوزه کرونا به‌طور خاص، این است که این موضوع به‌ندرت کانون مطالعه به‌خصوص در زبان فارسی بوده است. پژوهش حاضر می‌تواند گامی در راه توجه به اهمیت موضوع زبان در پیام سلامت در زبان فارسی و تنوع‌بخشی به این دسته از مطالعات باشد. عمده تمایز پژوهش حاضر با سایر

پژوهش‌ها این است که ضمن استفاده از یک پیکره زبانی به نسبت بزرگ در حوزه کرونا، از روش‌های محاسباتی و الگوریتمی مبتنی بر پردازش زبان طبیعی برای تحلیل احساسات و معنایی داده‌ها از نظر خوشه‌بندی معنایی عبارات کرونایی و بررسی هشتگ‌ها استفاده کرده است درحالی که داده‌های مورد نیاز پژوهش‌های زبانی ارتباطات سلامت محور به‌طور عمده بر مصاحبه استوارند و از روش‌های الگوریتمی پردازش زبان طبیعی برای آنها استفاده نمی‌شود.

چهارچوب نظری

در مکتب ساخت‌گرایی، سوسور^۱ (۱۹۱۶) معتقد است که زبان، دارای صورت^۲ و معنی^۳ است. دو سطح صورت و معنی، واقعیتی روان‌شناختی هستند که پیوند آنها، تمام جملات زبانی بیان‌شده گویشور یک زبان را شامل می‌شود. صورت می‌تواند به شکل یک زنجیره آوایی و یا نوشتاری که دارای مفهوم است، تجلی کند. صورت و معنی در یک نظام مشخص، واحدهای زبانی را ایجاد می‌کنند. این واحدهای زبانی که به‌عنوان واژه نیز تلقی می‌شوند، براساس روابط همنشینی^۴ در چهارچوب قواعد آن نظام مشخص، قابلیت ترکیب با یکدیگر را دارند تا واحدهای بزرگ‌تر را که جمله نامیده می‌شوند، بسازند. اساساً حوزه نحو، به بررسی و تحلیل واحدهای بزرگ‌تر از واژه در سطح جمله می‌پردازد و جملات و گروه‌ها^۵ را تحلیل و توصیف می‌کند. واژه، ممکن است بیش از یک معنا داشته باشد. بنابراین، رابطه همنشینی واژه‌ها با یکدیگر و ایجاد بافت زبانی، منجر به مشخص شدن معنای آنها می‌شود.

صورت زبانی می‌تواند به واسطه خط که دربرگیرنده توانش زبانی^۶ است، تجلی عینی پیدا کند. این تجلی عینی سبب می‌شود که زبان در حوزه رایانه مطرح شود و امکان پردازش آن میسر گردد تا رایانه به کمک روش‌های ماشینی و آمار، به مفهوم برسد.

یکی از اهداف مهم زبان‌شناسی رایانشی، مدل‌سازی توانایی‌های زبانی مغز است. در زبان‌شناسی رایانشی، استفاده از رویکرد قاعده‌مندی زبانی نسبت به رویکرد آماری-احتمالاتی بیشتر است و در پردازش زبان طبیعی، آمار و احتمال اهمیت بسزایی دارد. در مدل‌سازی آماری زبان^۷ تلاش می‌شود با کمک اطلاعات آماری استخراج‌شده از پیکره زبانی، مدلی آماری از زبان ارائه گردد. در معیار آماری اطلاعات دوسویه، میزان رابطه معنایی دو واژه با یکدیگر سنجیده می‌شود.

1. de Saussure

2. form

3. meaning

4. associative relation

5. phrase

7. statistical language modeling

از این رو، چرچ و همکاران^۱ (۱۹۹۱) از اطلاعات دوسویه نقطه‌به‌نقطه به‌عنوان معیار آماری برای تشخیص رابطه باهم‌آیی^۲ بین واژه‌ها استفاده کرده‌اند.^۳ منظور از باهم‌آیی، توالی دو یا چند واژه با یکدیگر است تا براساس تکرار کاربرد این توالی به‌صورت ثابت، یک الگوی ثابت زبانی شکل گیرد و به فسیل تبدیل شود. این باهم‌آیی‌ها تغییرناپذیر هستند و هرگونه تغییر در آنها، به ساخت نشاندار منجر می‌شود؛ مانند «کله پاچه» و نه «پاچه کله» و یا «خلیج فارس» و «خلیج همیشه فارس». چرچ و همکارانش معتقدند چنانچه عدد اطلاعات دوسویه نقطه‌به‌نقطه بالاتر از صفر باشد، بین دو واژه، رابطه معنایی وجود دارد. فرمول (۱) برای محاسبه اطلاعات دوسویه نقطه‌به‌نقطه بین دو واژه به کار می‌رود:

$$\begin{aligned}
 (1) \quad PMI &= \log \left(\frac{\frac{f(w_1, w_2)}{N}}{\frac{f(w_1)}{N} \times \frac{f(w_2)}{N}} \right) \\
 &= \log \left(\frac{f(w_1, w_2) \times N}{f(w_1) \times f(w_2)} \right) \\
 &= \log \left((f(w_1, w_2) \times N) \right) - \log (f(w_1) \times f(w_2)) \\
 &= (\log \cdot f(w_1, w_2) + \log \cdot N) - (\log \cdot f(w_1) + \log \cdot f(w_2))
 \end{aligned}$$

در این فرمول، $f(w_1, w_2)$ بسامد وقوع دو واژه w_1 و w_2 است. $f(w_1)$ بسامد واژه اول و $f(w_2)$ بسامد واژه دوم است. N تعداد کل واژه‌های پیکره است. اشکال فرمول (۱) برای محاسبه اطلاعات دوسویه نقطه‌به‌نقطه این است: چنانچه بسامد واژه‌ها و نیز تعداد توالی دو واژه پایین باشد، عدد اطلاعات دوسویه نقطه‌به‌نقطه بسیار بزرگ می‌شود که با توجه به نظر چرچ و دیگران (۱۹۹۱) بیانگر وجود رابطه معنایی زیاد بین این دو واژه است اما بسامد توالی این دو واژه گویای این ویژگی نیست؛ زیرا ممکن است دو واژه با بسامد پایین، عدد اطلاعات دوسویه نقطه‌به‌نقطه بالایی داشته باشند. برای پرهیز از این مشکل،

1. Church et al.

2. collocation

۳. تفاوت اطلاعات دوسویه نقطه‌به‌نقطه و اطلاعات دوسویه در این است که در اطلاعات دوسویه نقطه‌به‌نقطه فقط یک رخداد در هر لحظه مدنظر قرار می‌گیرد، درحالی‌که در اطلاعات دوسویه به میانگین تمام رخدادها توجه می‌شود.

بسامد توالی دو واژه را طبق فرمول (۲) در این محاسبات دخالت می‌دهیم و از این رابطه جدید در محاسبات استفاده می‌کنیم.

(۲)

$$LMI = \left((\log f_{(w_1, w_2)} + \log N) - (\log f_{(w_1)} + \log f_{(w_2)}) \right) \times f_{(w_1, w_2)}$$

فرمول (۲) اطلاعات دوسویه فرهنگ‌نگار^۱ (کیلگاریف و همکاران^۲، ۲۰۰۴) و یا اطلاعات دوسویه موضعی^۳ (اورت^۴، ۲۰۰۵) نامیده می‌شود. تضریب بسامد توالی دو واژه در عدد اطلاعات دوسویه نقطه‌به‌نقطه سبب می‌شود به‌هنگام مرتب‌سازی نزولی، مواردی که عدد اطلاعات دوسویه فرهنگ‌نگار آنها بالا است ولی بسامد پایین دارند، در جایگاه پایین‌تری نسبت به مواردی قرار گیرند که عدد اطلاعات دوسویه فرهنگ‌نگار و بسامد توالی بالایی دارند.

روش‌شناسی

برای انجام هر پژوهش، به داده نیاز است و در مطالعات زبان‌شناختی، باید از پیکره زبانی استفاده کرد. پیکره‌های زبانی موجود، به دلیل به‌روزرسانی نشدن در مطالعات مربوط به کرونا کارایی لازم را ندارند. از این رو، برای پرداختن به این پژوهش، به پیکره تخصصی زبانی که مرتبط با بیماری کرونا است، نیاز است تا بتوان اطلاعات زبانی مورد نظر را از این پیکره استخراج کرد. از آنجاکه در این پژوهش به بررسی پیام‌های سلامت‌محور در وبگاه‌های خبری و شبکه‌های اجتماعی پرداخته می‌شود، نیاز است که پیکره زبانی هدف، براساس ویژگی‌های گونه زبانی رسمی و غیررسمی در شبکه‌های اجتماعی تهیه گردد. این پیکره، در بازه زمانی دوازده‌ماهه تهیه می‌شود. از آن جاکه پدیده هشتگ‌زنی، از فضای مجازی آغاز شده و در نهادهای رسمی، مانند وزارت بهداشت نیز بسیار مورد استفاده قرار می‌گیرد، در پژوهش حاضر تلاش می‌شود تعدادی از هشتگ‌های پیام‌های سلامت مرتبط با بیماری کرونا در گونه رسمی، مانند وبگاه‌های خبری و گونه غیررسمی، مانند شبکه‌های اجتماعی بررسی شوند.

1. Lexicographic Mutual Information (LMI)

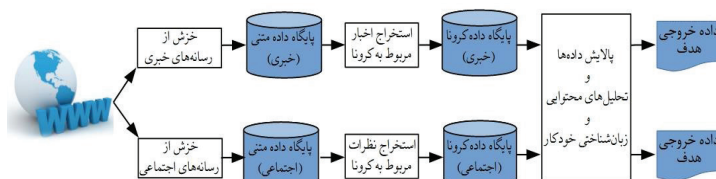
2. Kilgariff et al.

3. ocal mutual information

4. Evert

الف) جمع‌آوری داده‌های پژوهش: برای تهیه پیکره زبانی با حجم زیاد، می‌توان از روش‌های رایانشی استفاده کرد. یکی از ساده‌ترین روش‌ها، گردآوری داده از طریق خزش^۱ در وب است. برای انجام این پژوهش، با خزش از ۶ وبگاه خبری پرمخاطب که نقش سازنده‌ای در اطلاع‌رسانی در جامعه دارند، مانند وبگاه وزارت بهداشت، خبرگزاری صداوسیما، شبکه خبر، باشگاه خبرنگاران جوان، ایرنا و ایسنا و دو رسانه اجتماعی تلگرام و توییتر، داده‌های زبانی گردآوری شده‌اند. به دلیل تنوع در موضوعات، متن‌هایی که واژه «کرونا» در آنها ذکر شده، در ساخت پیکره تخصصی کرونا مورد استفاده قرار گرفته‌اند. بازه زمانی که برای استخراج داده‌ها در نظر گرفته شده، از زمان آغاز همه‌گیری در جهان، از دی‌ماه ۱۳۹۸ تا آذرماه ۱۳۹۹ بوده است. در شکل (۱) فرایند جمع‌آوری داده‌ها ترسیم شده است. ابتدا، داده‌های به‌دست‌آمده، در دو پایگاه داده جداگانه که آنها را «پایگاه داده متنی (خبری)» و «پایگاه داده متنی (اجتماعی)» می‌نامیم، نگهداری شده‌اند و سپس، داده‌های مربوط به کرونا با استفاده از کلیدواژه‌های «کرونا»، «کورونا»، «کوید» و «کووید» از این دو پایگاه، استخراج و در دو پایگاه به نام‌های «پایگاه داده کرونا (خبری)» و «پایگاه داده کرونا (اجتماعی)» ذخیره شده‌اند. معماری معرفی‌شده در شکل (۱) در چهارچوب پژوهش حاضر برای گردآوری داده‌ها طراحی و پیاده‌سازی شده است. در فرایند خزش، از کتابخانه اسکریپی^۲ که به زبان برنامه‌نویسی پایتون^۳ بوده، استفاده شده است.

نمودار ۱. فرایند گردآوری داده‌ها



1. crawl
2. Scrappy
3. Python

ب) پیش‌پردازش داده‌ها: پس از جمع‌آوری داده، نیاز است که در قسمت پیش‌پردازش، دو اقدام انجام شود: (۱) بهنجارسازی داده که به نوعی، ایجاد یکدستی در داده است. در این قسمت، یکدستی از نظر حروفی مانند «ی» و «ک» عربی و تبدیل آنها به معادل فارسی انجام می‌پذیرد؛ (۲) واحدسازی داده که تا حدی مرز واژه‌ها را تنظیم می‌کند. واژه‌ها یک فاصله بیرونی دارند که واحدی واژگانی را ایجاد می‌کند. رایانه نیز با احتساب فاصله به عنوان مرز، واژه‌های زبان را تشخیص می‌دهد. نوع دیگری از فاصله که نیم‌فاصله نامیده می‌شود، در میان یک واژه ظهور پیدا می‌کند. رعایت نکردن اصول نوشتاری در درج صحیح فاصله یا نیم‌فاصله، موجب تغییر در شناخت واژه می‌شود، مانند «دانش آموز» و «ویا بهتر است». در این مرحله، تلاش می‌شود تا حدودی مشکل واحدسازی واژه‌ها برطرف شود. در ادامه، اطلاعات مربوط به آمار توصیفی مانند تعداد واژه‌ها و تعداد واژه‌های نوع، میزان تنوع واژگانی در کنار طول متوسط جمله پس از آماده‌سازی پیکره، کار استخراج الگوهای زبانی انجام می‌پذیرد. ج) پیکره زبانی کرونا: باتوجه به وجود بایگانی در وبگاه‌های خبری، خزش تمامی اخبار این وبگاه‌ها در بازه زمانی دی‌ماه ۱۳۹۸ تا آذرماه ۱۳۹۹ انجام شده است. شایان ذکر است که داده‌های شبکه‌های اجتماعی به‌طور کامل بایگانی نمی‌شود و به‌صورت خطی وجود دارد و بعد از مدتی از بین می‌روند. تلاش شده است دیدگاه‌های منتشرشده در این دو بستر، در بازه زمانی مورد نظر گردآوری شوند.

گزارش یافته‌ها

یافته‌های پژوهش را براساس اهداف مورد نظر، تحلیلی بر عملکرد و محتوای وبگاه‌های خبری و شبکه‌های اجتماعی مرتبط با موضوع کرونا، به دو دسته تقسیم می‌کنیم اما نخست، بر عملکرد کلی این دو دسته از منابع اطلاعاتی در بستر جامعه می‌پردازیم.

الف: تحلیلی بر عملکرد وبگاه‌های خبری و شبکه‌های اجتماعی

داده‌های متنی که طی فرایند خزش از وبگاه‌های خبری و شبکه‌های اجتماعی هدف، گرد آمده‌اند، در جدول‌های (۱) و (۲) گزارش شده‌اند. همان‌گونه که جدول (۱) نشان می‌دهد، بیش از یک و نیم میلیون سند خبری از وبگاه‌های خبری هدف خزش شده که پیکره‌ای در حدود ۴۷۴ میلیون واژه‌ای تشکیل می‌دهد. بیش از ۹۰ درصد این پیکره، متشکل از مستندات خبری

ایسنا، خبرگزاری صداوسیما و باشگاه خبرنگاران جوان است که دلیل آن، تولید حجم زیاد محتوا در این سه وبگاه خبری است. کمترین حجم داده نیز به خبرگزاری ایرنا تعلق دارد.^۱

طول متوسط هر سند خبری در تمام داده در حدود ۳۱۴ واژه است. از میان منابع خزش شده، به‌طور متوسط، هر سند خبری در وبگاه وزارت بهداشت، در حدود ۶۶۸ واژه دارد که در مقایسه با متوسط هر سند خبری در تمام داده، کمی بیش از دو برابر است. این نکته نشان می‌دهد که در این نوع سند خبری، اطلاعات بیشتری نسبت به سایر خبرگزاری‌هایی که مهم‌ترین وظیفه‌شان نشر خبر و اطلاعات است، وجود دارد. شایان ذکر است که طول متوسط اخبار در وبگاه‌های شبکه خبر و باشگاه خبرنگاران جوان، کمتر از متوسط طول سند خبری در تمام داده است.

جدول ۱. اطلاعات آماری از وبگاه‌های خبری خزش شده

نام وبگاه خبری	تعداد خبرها	تعداد واژه‌ها	تعداد واژه‌های یکتا	طول متوسط خبرها	تنوع واژگانی	حجم داده خزش شده
وزارت بهداشت	۴۳۵۸	۲۹۱۰۱۶۴	۷۲۶۶۵	۶۶۷/۷۸	۰/۰۲۵۰	۱۳۶ (مگابایت)
شبکه خبر	۲۵۸۵۲	۷۹۳۰۴۱۲	۲۳۳۷۲۵	۳۰۶/۷۶	۰/۰۲۹۵	۷۰ (مگابایت)
ایرنا	۵۰۵۷	۱۹۱۲۲۴۷	۱۰۶۳۲۸	۳۷۸/۱۴	۰/۰۵۵۶	۲۲ (مگابایت)
ایسنا	۲۷۳۷۲۴	۱۱۶۳۴۰۲۹۵	۲۵۳۵۹۶۳	۴۲۵/۰۳	۰/۰۲۱۸	۱/۲ (گیگابایت)
خبرگزاری صداوسیما	۴۳۲۳۷۶	۱۵۰۰۳۸۰۸۸	۱۸۷۳۸۵۳	۳۴۷/۰۱	۰/۰۱۲۵	۱/۳ (گیگابایت)
باشگاه خبرنگاران جوان	۷۶۹۲۸۴	۱۹۴۶۷۹۹۱۵	۱۴۷۷۳۸۵	۲۵۳/۰۷	۰/۰۰۷۶	۲/۲ (گیگابایت)
جمع کل	۱۵۱۰۶۵۱	۴۷۳۸۱۱۱۲۱	۴۰۴۳۲۱۸	۳۱۳/۶۵	۰/۰۰۸۶	۴/۷ (گیگابایت)

۱. به‌دلیل مشکلات نرم‌افزاری حاصل از مسدود شدن سرور از سمت خبرگزاری ایرنا و دسترسی نداشتن به این سرور، امکان خزش حجم تمام داده‌های این خبرگزاری میسر نشد.

از نظر تنوع واژگانی، به‌طور کلی، تنوع واژگانی به‌کاررفته در کل اخبار در حدود ۰/۸۶ درصد است که بسیار کم است. نبود تنوع واژگانی در داده، بیانگر کاربرد واژه‌های تکراری است. بنابراین، برای تولید و درک اخبار، به دایره واژگانی زیادی نیاز نیست. از میان خبرگزاری‌ها، تنوع واژگانی به‌کاررفته در اخبار به باشگاه خبرنگاران جوان با حدود ۰/۷۶ درصد تعلق دارد که از میانگین تنوع واژگانی کل داده پایین‌تر است. این نکته بیانگر سادگی مطالب از نظر واژگانی در این وبگاه است. در این میان، خبرگزاری ایرنا به نسبت سایر خبرگزاری‌ها با وجود حجم کمتر داده خزش‌شده، از واژگان متنوع‌تری (۵/۵۶ درصد) برای انتشار خبر استفاده کرده است. به عبارت دیگر، برای انتشار خبرهای این خبرگزاری، از دایره واژگانی بیشتری استفاده شده و خوانایی خبرهای آن در مقایسه با سایر خبرگزاری‌ها، دشوارتر و درک جملاتش سخت‌تر بوده است. تنوع واژگانی در وبگاه‌های خبری وزارت بهداشت، شبکه خبر و ایسنا در یک محدوده بین ۱/۵ تا ۲ برابر تنوع واژگانی میانگین و تنوع واژگانی در خبرگزاری صداوسیما، حدود ۴۰ درصد بیشتر از تنوع واژگانی میانگین بوده است.

جدول ۲. اطلاعات آماری از شبکه‌های اجتماعی خزش‌شده

شبکه اجتماعی	تعداد دیدگاه‌های باز نشر نشده	تعداد واژه‌های دیدگاه‌ها	تعداد واژه‌های یکتا	طول متوسط دیدگاه‌های باز نشر نشده	تعداد دیدگاه‌های باز نشر شده	حجم داده
تویتر	۱۱۸۰۷۲۸	۲۵۲۸۶۲۶۸	۱۸۱۱۷۶۸	۲۲	۳۱۰۰۳۰	۳۴۰ (مگابایت)
اینستاگرام	۴۱۷۶۹۷	۲۹۰۵۵۱۰۱	۱۶۵۲۷۸۷	۷۰	۵۵۰۳۸۲	۹۱۹ (مگابایت)
جمع کل	۱۵۹۸۴۲۵	۵۴۳۴۱۳۶۹	۳۱۴۸۴۰۹	۳۴	۸۶۰۴۱۲	۳/۱ (گیگابایت)

در جدول ۲ حجم داده‌های خزش‌شده از دو شبکه اجتماعی تویتر و اینستاگرام گزارش شده است. در بازه زمانی دی‌ماه ۱۳۹۸ تا آذر ۱۳۹۹ تعداد کل ۱۴۹۰۷۵۸ نظر از تویتر و تعداد ۹۶۸۰۷۹ نظر از اینستاگرام گردآوری شده است. از نظر تعداد دیدگاه‌های منتشرشده در تویتر در مقایسه با اینستاگرام، ۶۰/۶۳ درصد از دیدگاه‌های گردآوری‌شده متعلق به تویتر بوده است. نکته درخور توجهی

که در داده‌های حاصل از شبکه‌های اجتماعی وجود دارد، امکان بازنشر یک نظر است که این ویژگی سبب بروز تکرار در داده می‌شود. همان‌گونه که در جدول (۲) مشخص است، حدود ۳۵ درصد از دیدگاه‌ها، بازنشر شده‌اند. شایان ذکر است که بازنشر دیدگاه‌ها در شبکه اجتماعی اینستاگرام، بسیار بیشتر از بازنشر آنها در شبکه اجتماعی توییتر بوده است. باتوجه به افزایش نیافتن محتوای زبانی در دیدگاه‌ها، در این پژوهش، موارد بازنشرشده را در پایگاه داده متنی اجتماعی، ذخیره اما در قسمت تحلیل داده‌ها، موارد بازنشرشده و تکراری را حذف می‌کنیم.

علاوه بر ویژگی بازنشر دیدگاه‌ها در شبکه‌های اجتماعی، کاربرد هشتگ برای بیان مفاهیم کوتاه نیز ویژگی دیگری است که به‌طور معمول، برای جریان‌سازی به‌کار می‌رود. دلیل نامگذاری هشتگ این است که برای بیان مفاهیم و برچسب‌گذاری آنها از نماد «#» که «هش» نامیده می‌شود بهره می‌برد. در داده‌های توییتر ۵۶/۲۵ درصد از کل هشتگ‌ها، به‌صورت تگ‌واژه‌ای به‌کار رفته‌اند و در اینستاگرام ۵۹/۰۶ درصد از کل آنها چندواژه‌ای بوده‌اند. در جدول (۳) اطلاعات آماری مربوط به هشتگ‌های به‌کاررفته در دیدگاه‌ها ارائه شده است. میانگین طول هشتگ‌های تولیدشده در توییتر اندکی طولانی‌تر از میانگین طول هشتگ‌های تولیدشده در اینستاگرام است؛ اما تکرار هشتگ‌ها در اینستاگرام سبب شده است که طول میانگین هشتگ‌های توییتر به نسبت کل داده در حدود یک حرف، کوتاه‌تر از طول میانگین هشتگ‌های اینستاگرام باشد.

جدول ۳. اطلاعات آماری هشتگ‌های تولیدشده در شبکه‌های اجتماعی خزش‌شده

پیکره	توضیحات		شبکه اجتماعی	
			اینستاگرام	توییتر
پیکره متنی (اجتماعی)	تعداد هشتگ		۱۱۴۱۷۲۹۹	۶۳۹۵۱۷
	تعداد یکنای هشتگ		۳۹۳۷۴۳	۸۸۵۵۱
	طول متوسط هشتگ (حروف)		۲۵/۶۴	۲۴/۴۶
	تعداد هشتگ تگ‌واژه‌ای		۷۶۰۹۳۹۲	۲۹۷۷۲۴
	تعداد یکنای هشتگ تگ‌واژه‌ای		۱۶۱۱۸۳	۳۸۷۳۸
	طول متوسط هشتگ تگ‌واژه‌ای (حروف)		۱۸/۰۹	۱۹/۸۸
	تعداد هشتگ چندواژه‌ای		۳۸۰۷۹۰۷	۳۴۱۷۹۳
	تعداد یکنای هشتگ چندواژه‌ای		۲۳۲۵۶۰	۴۹۸۱۳
	طول متوسط هشتگ چندواژه‌ای (حروف)		۳۰/۸۶	۳۳/۳۵

تا این مرحله، کار جمع‌آوری داده‌ها از منابع هدف انجام گرفته است. با اینکه این داده‌ها بسیار ارزشمندند، به دلیل متمرکز بودن پژوهش بر موضوع پیام‌های سلامت محور مربوط به «بیماری کرونا» به پالایش داده نیاز است. برای این هدف، از داده‌های خزش شده در این دو پایگاه داده، واژه‌های مربوط به بیماری کرونا، مانند کرونا، کورونا، کوید و کووید (Covid، CORONA، corona، Corona، COVID و covid) جست‌وجو و در دو پایگاه داده، به نام‌های «پایگاه داده کرونا (خبری)» و «پایگاه داده کرونا (اجتماعی)» ذخیره شده‌اند. داده‌های موجود در این دو پایگاه داده را می‌توان به عنوان پیکره زبانی قلمداد کرد؛ در نتیجه، این دو منبع داده را پیکره کرونای خبری و پیکره کرونای اجتماعی خطاب می‌کنیم.

جدول ۴. اطلاعات آماری از وبگاه‌های خبری خزش شده مربوط به کرونا

نام وبگاه خبری	تعداد اسناد	تعداد واژه‌ها	تعداد واژه‌های یکتا	طول متوسط سند	حجم داده خزش شده
وزارت بهداشت	۳۱۱۲	۱۹۸۹۰۶۴	۴۷۲۸۹	۶۳۹/۱۶	۲۱ (مگابایت)
شبکه خبر	۷۲۲۵	۲۵۶۱۶۰۸	۸۲۵۴۰	۳۵۴/۵۵	۲۸ (مگابایت)
ایرنا	۱۶۲۹	۷۸۷۵۷۸	۵۴۶۶۰	۴۸۳/۴۷	۸/۶ (مگابایت)
ایسنا	۸۸۷۶۲	۴۱۴۳۳۹۵۳	۷۹۲۷۲۰	۴۶۶/۸۰	۴۵۵ (مگابایت)
خبرگزاری صداوسیما	۱۱۳۳۶۲	۴۴۶۴۹۲۰۷	۷۰۶۴۴۲	۳۹۳/۸۶	۴۲۹ (مگابایت)
باشگاه خبرنگاران جوان	۱۷۲۹۸۵	۵۶۷۰۹۶۴۹	۵۳۳۶۵۶	۳۲۷/۸۳	۶۳۰ (مگابایت)
جمع کل	۳۸۷۰۷۵	۱۴۸۱۳۱۰۵۹	۱۵۸۶۲۷۷	۳۸۲/۶۹	۳/۱ (گیگابایت)

در مرحله بعد، پس از تحلیل‌های خودکار زبان‌شناختی و محتوایی، داده‌های تحلیل شده رایانشی برای فراتحلیل که هدف اصلی این پژوهش بوده، آماده شده است. در جدول‌های ۴ و ۵ حجم داده‌های موجود در دو پایگاه داده کرونا، حاصل از خزش وبگاه‌های خبری و شبکه‌های اجتماعی گزارش شده است. داده‌های جمع‌آوری شده مربوط به کرونا از وبگاه‌های خبری، حدود ۳۸۷ هزار سند خبری را شامل می‌شود که حاوی بیش از ۱۴۸ میلیون واژه است. حجم خبرهای مربوط به کرونا نیز ۲۵/۶۲ درصد از خبرهای خزش شده از وبگاه‌های خبری را دربرمی‌گیرد. داده‌های جمع‌آوری شده مربوط به کرونا از شبکه‌های

اجتماعی، در حدود ۷۷ هزار نظر است که حاوی بیش از ۴/۴ میلیون واژه است و دیدگاه‌های گردآوری شده مربوط به کرونا ۴/۸ درصد از حجم داده‌های خزش شده از شبکه‌های اجتماعی را دربرمی‌گیرد. اگرچه ۶۳/۶۸ درصد از دیدگاه‌ها به شبکه اجتماعی توییتر تعلق دارد، کوتاه بودن آنها سبب شده است که دیدگاه‌های حاصل از اینستاگرام ۶۵/۸۹ درصد از واژه‌های پیکره متنی به‌دست آمده از شبکه‌های اجتماعی را تشکیل دهد.

جدول ۵. اطلاعات آماری از شبکه‌های اجتماعی خزش شده مربوط به کرونا

شبکه اجتماعی	تعداد دیدگاه‌های باز نشر نشده	تعداد واژه‌های دیدگاه‌ها	تعداد واژه‌های یکتا	طول متوسط دیدگاه‌های باز نشر نشده	تعداد دیدگاه‌های باز نشر شده	حجم داده
توییتر	۴۸۸۲۳	۱۵۰۲۷۶۹	۱۶۶۷۷۹	۳۰/۷۸	۴۴۸۲۷	۱۷ (مگابایت)
اینستاگرام	۲۷۸۴۴	۲۹۰۳۴۶۹	۲۳۶۹۵۵	۱۰۴/۲۸	۱۰۹۳۷۹	۵۸ (مگابایت)
جمع کل	۷۶۶۶۷	۴۴۰۶۲۳۸	۳۶۲۷۳۹	۵۷/۴۷	۱۳۳۳۱۸	۷۵ (مگابایت)

گفته شد یکی از خصوصیات دیدگاه‌های شبکه‌های اجتماعی، استفاده از هشتگ برای برچسب‌گذاری مفاهیم است که امروزه به حوزه رسانه‌های خبری نیز تسری پیدا کرده است. هشتگ‌های تولیدشده در دیدگاه‌های پالایش و محدودشده به دیدگاه‌های مربوط به کرونا، بسیار متنوع بود. گاه در محتوای نظر، از واژه‌های هدف، مانند کرونا، کورونا، کوید، کووید (CORONA, corona, Corona)، COVID و covid استفاده نشده ولی یکی از این واژه‌ها در هشتگ مربوط به آن نظر به کار رفته است. باین‌حال، از این دسته از دیدگاه‌ها نیز در تحلیل خود استفاده کرده‌ایم. از میان هشتگ‌های استخراج شده از دیدگاه‌ها، مواردی را که از نظر محتوایی، به موضوع کرونا مرتبط بوده‌اند، استخراج و به‌همراه بسامد در جدول ۶ گزارش کرده‌ایم. هشتگ‌های تک‌واژه‌ای موجود در دیدگاه‌های مربوط به کرونا در اینستاگرام و توییتر، حدود ۴۷ درصد است و بقیه، هشتگ‌های چندواژه‌ای را شامل می‌شود. طول میانگین هشتگ‌های پیکره کرونا در مقایسه با پیکره متنی تقریباً یکسان است.

تحلیل احساسی و معنایی
عملکرد و محتوای
وبگاه‌های خبری و
شبکه‌های اجتماعی با
محوریت پیام سلامت

جدول ۶. اطلاعات آماری هشتگ‌های تولیدشده در پیکره کرونای اجتماعی

پیکره	توضیحات	شبکه اجتماعی	
		اینستاگرام	توییتر
پیکره کرونای اجتماعی	تعداد هشتگ	۶۱۰۳۱۷	۳۴۰۰۳
	تعداد یکتای هشتگ	۵۶۵۴۹	۹۴۴۵
	طول متوسط هشتگ (حروف)	۲۳/۳۹	۲۶/۳۴
	تعداد هشتگ تک‌واژه‌ای	۴۱۲۰۴۱	۲۰۵۸۱
	تعداد یکتای هشتگ تک‌واژه‌ای	۲۶۵۸۴	۴۴۵۰
	طول متوسط هشتگ تک‌واژه‌ای (حروف)	۱۶/۳۴	۱۷/۷۴
	تعداد هشتگ چندواژه‌ای	۱۹۸۲۷۶	۱۳۴۲۲
	تعداد یکتای هشتگ چندواژه‌ای	۲۹۹۶۵	۴۹۹۵
	طول متوسط هشتگ چندواژه‌ای (حروف)	۲۹/۶۵	۳۴/۰۰

به‌منظور تحلیل داده‌های خزش‌شده از وبگاه‌های خبری و شبکه‌های اجتماعی، بازه زمانی دی‌ماه ۱۳۹۸ تا آذر ۱۳۹۹ را ابتدا براساس ماه انتشار، دسته‌بندی و سپس، داده‌ها را به چهار دوره زمانی تقسیم کرده‌ایم. در جدول ۷ داده‌های متعلق به هر ماه و در جدول ۸ داده‌های متعلق به دوره چهارماهه توضیح داده شده است. این تقسیم‌بندی، موجب می‌شود که ضمن فراهم‌آمدن امکان پردازش داده‌ها، امکان مقایسه اطلاعات به‌دست‌آمده در بازه‌های زمانی مختلف نیز میسر شود. همچنین، تقسیم‌بندی صورت‌گرفته به‌لحاظ زمانی می‌تواند ضمن بیان توزیع آماری محتوای تولیدشده، با شروع همه‌گیری کرونا و فراز و فرود آن در کشور ایران منطبق گردد.

برای داشتن یک نگاه کلی نسبت به تولید محتوا در بازه زمانی دی‌ماه ۱۳۹۸ تا آذر ۱۳۹۹، اطلاعات جدول ۷ در نمودار ۱ ترسیم شده است.

پژوهش‌های ارتباطی،
سال سی و سوم،
شماره ۱ (پیاپی ۱۲۵)،
بهار ۱۴۰۵

جدول ۷. اطلاعات آماری داده‌های کرونا در بازه‌های یک‌ماهه

ایرنا	سند	واژه	ایستا	شبکه خبر		وزارت بهداشت		
				سند	واژه	سند	واژه	
۰	۰	۹۲۷۸۲	۲۱۴	۱۵۴۱۹	۳۱	۴۳۱۱	۸	دی ۱۳۹۸
۰	۰	۱۲۷۸۸۰۹	۳۴۱۵	۸۶۰۸۷	۲۳۹	۳۴۷۲۹	۷۵	بهمن ۱۳۹۸
۰	۰	۵۲۴۴۵۲۳	۱۳۷۹۸	۲۰۳۶۶۵	۶۲۴	۸۰۶۶۱	۱۹۹	اسفند ۱۳۹۸
۰	۰	۵۰۵۷۳۹۰	۱۱۵۹۲	۲۴۶۶۳۳	۶۵۸	۱۰۶۷۹۷	۲۰۲	فروردین ۱۳۹۹
۰	۰	۴۰۸۱۰۴۲	۸۵۱۳	۲۰۳۶۵۵	۵۴۷	۹۸۴۸۰	۱۷۵	اردیبهشت ۱۳۹۹
۰	۰	۳۳۳۵۵	۶۶۲۲	۱۹۶۵۸۸	۴۲۵	۸۴۹۳۰	۱۴۸	خرداد ۱۳۹۹
۰	۰	۳۷۳۴۸	۷۸۰۶	۱۹۳۱۸۶	۴۷۳	۱۰۴۸۲۲	۱۷۶	تیر ۱۳۹۹
۱۶۹۷۴	۳۷	۳۷۳۴۹	۷۰۴۶	۲۱۷۳۹۳	۵۳۴	۱۱۸۱۶۷	۲۰۱	مرداد ۱۳۹۹
۱۸۵۴۳	۴۲	۳۱۹۳۱	۶۳۸۷	۱۷۱۳۵۶	۴۴۵	۱۲۲۷۶۱	۲۲۶	شهریور ۱۳۹۹
۹۳۱۴۸	۲۲۰	۳۴۳۴۹	۷۰۱۰	۳۱۶۹۰۳	۹۸۱	۲۷۳۸۱۸	۴۴۷	مهر ۱۳۹۹
۱۲۸۷۵۷	۲۶۷	۴۰۰۰۷	۸۲۸۶	۳۷۳۰۶۲	۱۱۱۸	۳۲۱۱۹۱	۴۹۱	آبان ۱۳۹۹
۵۳۰۵	۱۰۶۳	۴۲۴۷۸	۸۰۷۲	۳۳۳۹۷۲	۱۱۵۰	۴۱۸۴۳۸	۶۹۳	آذر ۱۳۹۹

تحلیل احساسی و معنایی
عملکرد و محتوای
وبگاه‌های خبری و
شبکه‌های اجتماعی با
محوریت پیام سلامت

ادامه جدول ۷.

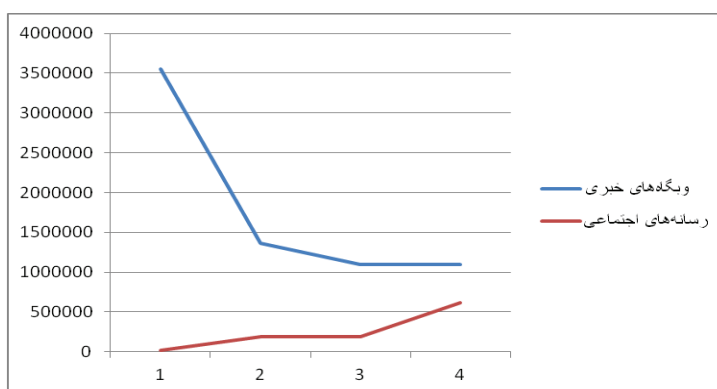
شبکه‌های اجتماعی		باشگاه خبرنگاران جوان		خبرگزاری صداوسیما		
واژه	سند	واژه	سند	واژه	سند	
۶۱۷	۱۵	۱۳۶۶۰۵	۴۹۱	۷۳۶۰۷	۱۱۷	دی ۱۳۹۸
۱۷۵۲۸	۳۹۵	۱۶۴۲۵۵۵	۶۲۷۴	۷۳۳۷۱۹	۲۲۴۷	بهمن ۱۳۹۸
۳۷۹۲۲۹	۷۸۳۷	۸۰۳۵۳۰۰	۲۸۴۹۲		۱۰۱۶۱	اسفند ۱۳۹۸
۹۰۶۲۴۹	۱۸۲۴۴	۸۹۸۴۰۴۶	۲۷۲۷۱	۲۸۷۶۵۰۰	۹۴۹۲	فروردین ۱۳۹۹
۷۰۳۵۸۷	۷۴۱۳	۶۴۲۶۸۸۵	۱۹۱۸۱	۲۶۸۰۵۷۵	۸۱۶۶	اردیبهشت ۱۳۹۹
۱۸۶۷۹۲۷	۱۶۰۸۴	۴۸۸۲۸۱۱	۱۳۹۸۲	۲۰۹۷۱۴۶	۶۰۹۳	خرداد ۱۳۹۹
۲۸۰۵۱۹	۳۷۴۵	۵۲۳۷۴۱۵	۱۵۹۶۳	۲۴۳۷۳۸۶	۷۶۹۸	تیر ۱۳۹۹
۶۴۰۲۲۷	۱۱۸۸۵	۴۶۹۱۷۷۸	۱۳۶۷۹	۲۵۶۵۲۷۳	۸۳۱۶	مرداد ۱۳۹۹
۲۶۱۲۱۶۷	۲۳۷۲۷	۴۱۶۴۶۹۵	۱۱۸۳۷	۲۱۲۲۸۹۶	۷۰۷۷	شهریور ۱۳۹۹
۵۵۸۸۴۵۱	۶۲۷۸۶	۴۱۱۴۰۱	۱۱۸۳۳	۲۳۸۳۱۳۲	۷۸۷۴	مهر ۱۳۹۹
۶۶۹۱۱۹۳	۵۷۶۲۷	۴۱۳۰۴۸	۱۲۴۱۲	۲۰۵۰۲۹۲	۷۰۱۳	آبان ۱۳۹۹
۴۷۵	۱۰	۴۰۳۲۲	۱۱۵۶۹	۲۰۸۱۸	۷۱۱۶	آذر ۱۳۹۹

جدول ۸. اطلاعات آماری داده‌های کرونا در بازه‌های چهارماهه

منبع	سه ماهه اول		سه ماهه دوم	
	سند	واژه	سند	واژه
وبگاه‌های خبری	۳۵۴۹۴۲۲	۱۲۵۴۹۸۳۰۷	۱۳۶۰۷۷۶	۴۸۳۳۸۸۹۰
شبکه‌های اجتماعی	۲۳۰۱۳	۶۴۰۵۸۳	۱۸۵۲۶۷	۵۲۹۵۹۰۸

منبع	سه ماهه سوم		سه ماهه چهارم	
	سند	واژه	سند	واژه
وبگاه‌های خبری	۱۰۹۶۶۰۰	۳۸۴۴۱۴۰۶	۱۰۹۲۰۴۶	۳۸۸۱۸۰۱۱
شبکه‌های اجتماعی	۱۹۲۴۰۷	۵۵۸۵۵۵۶	۶۱۵۰۷۰	۱۸۳۵۱۲۴۲

نمودار ۱. انتشار محتوای مربوط به کرونا در بازه‌های چهارماهه از وبگاه‌های خبری و اجتماعی



برای نشان دادن اهمیت توجه به تولید محتوای سلامت‌محور، به مقایسه‌ای صرفاً آماری میان تولید محتوا و میزان ابتلا و مرگ و میر ناشی از کرونا پرداخته‌ایم. همان‌گونه که در نمودار ۱ مشخص است، روند تولید محتوا در وبگاه‌های خبری در سه‌ماهه اول، یعنی دی‌ماه تا اسفندماه ۱۳۹۸، در مقایسه با سه‌برش زمانی دیگر، بسیار زیاد است؛ درحالی که در سه‌برش زمانی بعدی، در طول ۹ ماه، تولید محتوا در وبگاه‌های خبری روند نزولی به خود گرفته و تا حدودی در برش زمانی چهارم به‌نسبت برش زمانی سوم، ثابت باقی مانده است. شایان ذکر است که در همین برش زمانی، دیدگاه‌های شبکه‌های اجتماعی رو

به افزایش بوده است. اگرچه دیدگاه‌های منتشرشده در دو برش زمانی دوم و سوم تا حدودی ثابت بوده؛ ولی در برش زمانی سه‌ماهه چهارم رو به رشد بوده و بیشترین حجم دیدگاه‌های منتشرشده را به خود اختصاص داده است.

در نمودارهای ۲ و ۳، روند دچارشدن به بیماری کرونا و جان‌باختن افراد در جامعه ایران در بازه زمانی دی‌ماه ۱۳۹۸ تا آذرماه ۱۳۹۹ برگرفته از وبگاه <https://www.worldometers.info/coronavirus> ترسیم شده است. در این وبگاه، آمار روزانه کشورهای مختلف در مورد تعداد افراد مبتلا یا فوت‌شده بر اثر بیماری کرونا، گزارش و ذخیره شده است. شایان ذکر است که دسترسی به آمار کشورهای مختلف، امکان مقایسه را فراهم می‌آورد. از آنجاکه این مسئله، جزء اهداف پژوهش حاضر نبوده است، بررسی این موضوع به پژوهش‌های آتی محول می‌شود.

با نگاه کلی به نمودار ۲ مشخص می‌شود که تعداد افراد تحت تأثیر این بیماری، رو به افزایش بوده‌اند و بیشترین تعداد بیمار، به بازه زمانی مهر تا آبان ۱۳۹۹ تعلق داشته‌اند. روند مرگ و میر ناشی از بیماری کرونا در نمودار ۳ با فراز و فرود ولی در کل، افزایشی بوده است. در این نمودار، سه قله وجود دارد که بلندین آن متعلق به بازه زمانی مهر تا آذر ۱۳۹۹ است. نتیجه‌ای که از انطباق نمودارهای ۲ و ۳ به‌دست می‌آید، این است که در سه ماهه مهر تا آذر ۱۳۹۹، تعداد افراد دچار بیماری کرونا و مرگ و میر حاصل از این بیماری بیشتر از ماه‌های دیگر بوده است.

حال اگر سه نمودار ۱، ۲ و ۳ را با یکدیگر مقایسه کنیم، می‌بینیم که با وجود افزایش تعداد افراد بیمار و فوت‌شده تحت تأثیر بیماری کرونا، متأسفانه تولید محتوای سلامت‌محور در رسانه‌های خبری رو به نزول بوده است و رسانه‌های خبری به رسالت وجودی خود که همان اطلاع‌رسانی، افزایش آگاهی عمومی و نشر اطلاعات در خصوص این موضوع خاص بوده به‌درستی عمل نکرده‌اند. شاید یکی از دلایل افزایش بیماری و مرگ، عادی‌انگاری و جدی‌نگرفتن بیماری، همین ناآگاهی و منتشر نکردن پیام‌های سلامت‌محور باشد. البته ممکن است مسئله آگاهی‌بخشی و هشداردهی در بازه زمانی دی‌ماه تا اسفندماه ۱۳۹۸ در دستور کار رسانه‌های خبری قرار گرفته باشد و پس از این بازه، مطالب تکراری باشند اما به‌هرحال، اطلاع‌رسانی نباید دچار اختلال شود. به‌عبارت دیگر،

پاشنه آشیل رسانه‌های خبری در مورد پیام‌های سلامت‌محور بیماری کرونا، آگاه نکردن افراد جامعه به عنوان مخاطبان اصلی بوده است. شایان ذکر است که شبکه‌های اجتماعی نقش پررنگ‌تری در این زمینه ایفا کرده‌اند، چنان‌که می‌توان انعکاس افزایش دیدگاه‌ها را با افزایش بیماری و مرگ افراد بر اثر این بیماری همسو دانست.

نمودار ۲. روند ابتلا به بیماری کرونا در ایران



نمودار ۳. روند مرگ افراد دچار بیماری کرونا در ایران



ب) تحلیلی بر محتوای وبگاه‌های خبری و شبکه‌های اجتماعی

از آنجاکه موضوع مورد بحث این پژوهش، بیماری کرونا بوده است، ابتدا نیاز است که بسامد و رتبه این واژه در مجموع داده‌های خزش شده مشخص شود. کلیدواژه‌هایی که برای معرفی بیماری کرونا به کار رفته‌اند عبارت‌اند از: کرونا، کوید، کووید، کورونا و کروناویروس. پس از استخراج واژگان از دو پیکره متنی و اجتماعی، این کلیدواژه‌ها در پیکره جست‌وجو و بسامد آنها به همراه رتبه هریک در جدول (۹) گزارش شده است.

جدول ۹. بسامد و رتبه واژه‌های مرتبط با کرونا در وبگاه‌های خبری و شبکه‌های اجتماعی

واژه	پیکره کرونا (خبری)				پیکره کرونا (اجتماعی)			
	رتبه	بسامد	بسامد نسبی	تضریب بسامد نسبی و رتبه	رتبه	بسامد	بسامد نسبی	تضریب بسامد نسبی و رتبه
کرونا	۱۵	۹۳۱۹۸۸	۰/۰۰۶۶۱۱	۶۱۶۱/۴۳۲۶	۲۱	۶۴۳۳۶	۰/۰۰۰۴۵۶	۲۹/۳۶۰۹
کووید	۱۷۴	۱۰۲۹۱۹	۰/۰۰۰۷۳۰	۷۵/۱۳۶۷	۸۲۳	۸۳۸۱	۰/۰۰۰۰۵۹	۰/۴۹۸۳
کرونایی	۲۵۷	۷۳۱۶۹	۰/۰۰۰۵۲۰	۳۷/۹۷۶۵۳	۱۲۲	۳۴۸۲	۰/۰۰۰۰۲۵	۰/۰۸۶۰
کروناویروس	۸۵۴	۲۲۳۸۶	۰/۰۰۰۱۵۹	۳/۵۵۴۸	۶۵۷۱	۷۵۳	۰/۰۰۰۰۰۵	۰/۰۰۰۴۰
پسا کرونا	۳۶۲۸	۲۶۱۶	۰/۰۰۰۰۱۸	۰/۰۴۸۵	۴۴۵۴	۱۲۱	۰/۰۰۰۰۰۹	۰/۰۰۱۰
کورونا	۴۸۸	۸۳۶۸	۰/۰۰۰۰۰۴	۰/۰۳۳۵	۸۷۷	۶۱۵۲	۰/۰۰۰۰۴۳	۰/۲۶۴۵
کوید	۴۲۰۲	۱۹۶۹	۰/۰۰۰۰۱۴	۰/۰۲۷۵	۱۸۵۱	۳۳۷	۰/۰۰۰۰۰۲	۰/۰۰۰۰۸

همان‌گونه که در جدول (۹) دیده می‌شود، واژه «کرونا» پرکاربردترین واژه در وبگاه‌های خبری و شبکه‌های اجتماعی است. واژه «کووید» پرکاربردترین واژه بعدی است. واژه «کروناویروس» که معادل واژه انگلیسی «coronavirus»، است سومین واژه پرکاربرد در وبگاه‌های خبری و چهارمین واژه در شبکه‌های اجتماعی از میان کلیدواژه‌های استخراجی است. نگارش املایی «کورونا» و «کوید» چهارمین و پنجمین واژه‌های پرکاربرد از مجموعه کلیدواژه‌ها در وبگاه‌های خبری و واژه‌های «کروناویروس» و «کوید» چهارمین و پنجمین واژه‌های پرکاربرد از مجموعه کلیدواژه‌ها، در شبکه‌های اجتماعی هستند. دو واژه اشتقاقی و تصریفی ساخته‌شده از واژه پایه «کرونا» نیز که در وبگاه‌های خبری و شبکه‌های اجتماعی پرکاربردند، عبارت‌اند از: «کرونایی» که صفت است و «پساکرونا» که اسم است. از میان این دو، «کرونایی» بسیار پرکاربرد است. دو-واژه‌ای‌های با بسامد بالاتر از ۱۰۰ که حاوی یکی از چهار کلیدواژه مربوط به کرونا بوده‌اند و از پیکره کرونای خبری و اجتماعی استخراج شده‌اند در جدول (۱۰) آمده‌اند. نکته در خور توجه در جدول یادشده این است که در مورد نام بیماری «کرونا» تنوع بسیاری وجود دارد و همین امر سبب می‌شود که در نحوه نگارش آن نتوان به انسجام زبانی دست یافت. دو فعل سبک «گرفتن» و «داشتن» نیز برای دچارشدن یا دچاربودن به این بیماری به کار می‌روند.

جدول ۱۰. واژه‌های مرکب با استفاده از کلیدواژه‌های مربوط به کرونا

	واژه اول	واژه دوم	بسامد	اطلاعات دوسویه فرهنگ‌نگار
پیکره کرونای خبری	کووید	۱۹	۱۰۰۵۸۲	۶۹۸۰۹۰/۹۴
	کرونا	ویروس	۲۳۷۰۸	۵۰۴۶۶/۰۶
	کوید	۱۹	۱۸۹۹	۱۳۱۵۵/۳۲
	کرونا	مثبت	۲۵۱۴	۵۱۰۴/۵۶
	کرونا	داشتن	۹۱۲۲	۲۵۱۲/۲۷
	کووید	نوزده	۲۸۵	۱۹۲۳/۷۴
	کرونا	گرفتن	۳۸۰۴	۱۶۱۸/۹۲
	کروناویروس	داشتن	۲۶۴	۱۲۱/۸۸
پیکره کرونای اجتماعی	کووید	۱۹	۷۹۸۲	۵۵۳۶۵/۵۹
	کوید	۱۹	۳۱۹۵	۲۲۱۶۳/۷۰
	کرونا	گرفتن	۷۲۲۱	۱۸۳۶۸/۲۲
	کرونا	ویروس	۱۶۴۲	۲۱۰۹/۶۱
	کوید	نوزده	۱۲۸	۱۰۲۰/۵۹
	کرونا	مثبت	۳۰۹	۴۹۹/۷۷

نکته قابل توجه در این جدول، وجود تنوع در نگارش «کووید ۱۹»، «کوید ۱۹» است که با درج فاصله کامل میان این دو عنصر در رایانه، دو واژه تشخیص داده می‌شود. نگارش «کووید نوزده» و «کوید نوزده» نیز بر این تنوع افزوده است. نگارش «کرونا ویروس» به جای «کرونا ویروس» نوع دیگری از این تنوع است. آنچه می‌توان نتیجه گرفت، این است که در نگارش واژه‌های مربوط به بیماری کرونا، تشتت آرای زیادی بین گویشوران وجود دارد و متأسفانه فرهنگستان زبان و ادب فارسی که وظیفه معیارسازی خط زبان فارسی را برعهده دارد وحدت رویه‌ای را در آن زمان اعلام نکرده است تا ضمن کاستن از تنوع نگارشی، از دشواری ارتباط مخاطب عام با محتوای تولیدشده بکاهد؛ چراکه ممکن است مخاطب عام به این جمع‌بندی نرسد که تمامی اصطلاحات و شیوه‌های نگارش مختلف مربوط به این بیماری در رسانه یکسان است و تفاوتی بین آنها وجود ندارد. به عبارت دیگر، این تنوع، سبب مشوش شدن ذهن مخاطب و ارتباط برقرار نکردن او با محتوای تولیدشده می‌گردد. شایان ذکر است که «کوید ۱۹» و «کرونا ویروس» اصطلاحاتی علمی هستند که در جامعه پزشکی معنادارند، همان چیزی که در جامعه به‌طور عام، با نام «بیماری کرونا» شناخته می‌شود.

برای تحلیل در سطح واژه، از داده‌های بن‌واژه‌سازی شده استفاده می‌کنیم و همسو با این هدف، دوواژه‌ای‌ها را از دو پیکره کرونای خبری و اجتماعی به‌دست می‌آوریم. عبارات پربسامد دوواژه‌ای اولیه ترکیبی از حروف اضافه و فعل بود که نمی‌تواند به‌خوبی محتوای سلامت‌محور مربوط به کرونا را بیان کند. بنابراین، در این سطح از تحلیل، عباراتی که کلیدواژه‌های مربوط به کرونا نقش وابسته را در آنها بازی می‌کنند در دو پیکره بن‌واژه‌سازی شده کرونا بررسی شده‌اند. تعداد دوواژه‌ای‌های استخراج‌شده از پیکره کرونای خبری، ۶۳۳۳۲۳۷ مورد و تعداد دوواژه‌ای‌ها در پیکره کرونای اجتماعی، ۱۱۶۹۴۸۶ مورد بوده است که در جدول ۱۱ تعداد ۵۰ دو واژه‌ای پربسامد با بالاترین امتیاز اطلاعات دوسویه فرهنگ‌نگار گزارش شده است.

پژوهش‌های ارتباطی،
سال سی و سوم،
شماره ۱ (پیاپی ۱۲۵)،
په‌ار ۱۴۰۵

جدول ۱۱. دوواژه‌ای‌های پربسامد استخراج‌شده از دو پیکره بن‌واژه‌سازی‌شده کرونا

پیکره کرونا (اجتماعی)				پیکره کرونا (خبری)			
اطلاعات فرهنگ‌نگار	بسامد	واژه دوم	واژه اول	اطلاعات فرهنگ‌نگار	بسامد	واژه دوم	واژه اول
۱۰۷۳۷۴/۵۳	۲۶۴۲۶	کرونا	ویروس	۱۱۴۸۲۲۱/۲۹	۲۵۵۰۵۰	کرونا	ویروس
۲۱۹۸۷/۴۱	۱۷۰۶۹	کرونا	به	۱۹۲۹۲۲/۰۱	۱۰۱۸۶۷	کرونا	بیماری
۲۰۶۳۹/۵۱	۱۲۱۷۸	کرونا	با	۱۰۸۴۱۱/۳۸	۸۹۷۹۱	کرونا	به
۳۱۳۴۸/۵۴	۹۰۳۶	کرونا	بیماری	۲۵۶۸۳۷/۴۵	۶۹۹۱۶	کرونا	شیوع
۲۶۳۸۰/۵۶	۷۶۷۳	کرونا	شیوع	۱۲۰۵۹/۸۰	۳۰۸۹۹	کرونا	از
۳۵۶۲/۹۰	۶۷۴۸	کرونا	از	۵۴۸۲۳/۱۰	۲۵۴۸۴	کووید۱۹	به
۱۷۹۵۵/۲۶	۴۲۹۵	کرونا	طنز	۱۰۱۰۳۱/۲۴	۲۲۵۳۸	کووید۱۹	بیماری
۱۱۸۸۵/۹۱	۳۱۳۶	کرونا	تست	۵۱۹۰۷/۸۳	۱۴۶۷۱	کرونا	بحران
۸۵۶۰/۰۵	۲۶۳۴	کرونا	دوران	۵۲۳۹۱/۹۱	۱۳۸۵۵	کرونا	تست
۹۱۹۴/۶۹	۲۵۹۷	کرونا	واکسن	۲۰۹۲۵/۵۳	۹۲۶۱	کرونا	ستاد
۱۳۷۷۶/۱۴	۲۵۳۵	کووید۱۹	بیماری	۲۸۱۶۵/۸۵	۸۹۶۴	کرونا	واکسن
۷۵۴۸/۲۱	۲۲۳۴	کرونا	بحران	۲۷۸۶۶/۲۳	۷۱۹۶	کووید۱۹	بیمار
۵۶۹۶/۳۰	۱۸۰۶	کرونا	به‌خاطر	۲۲۲۸۷/۳۱	۷۱۲۲	کووید۱۹	ویروس
۴۰۱۶/۱۶	۱۷۷۶	کووید۱۹	به	۱۶۶۶۵/۲۰	۶۷۷۳	کرونا	مدیریت
۶۴۷۸/۷۳	۱۷۳۹	کرونا	قربانی	۱۷۳۶۲/۷۹	۵۶۰۱	کرونا	دوران
۳۲۵۰/۲۰	۱۷۰۸	کرونا	اثر	۱۰۵۳۶/۹۵	۵۰۹۷	کرونا ویروس	به
۳۸۵۱/۷۴	۱۶۰۹	کرونا	ستاد	۱۳۴۳۱/۳۷	۴۶۵۳	کرونا	آزمایش
۲۱۵۴/۶۷	۱۲۳۹	کرونا	بدلیل	۱۶۸۹۷/۲۸	۳۷۷۰	کرونا ویروس	شیوع
۴۲۰۳/۶۸	۱۱۸۳	کرونا	بخاطر	۴۳۹۱/۳۹	۳۱۹۶	کرونا	درمان
۱۲۴۲/۹۳	۸۵۹	کرونا	وضعیت	۸۶۹۸/۶۱	۳۱۴۶	کرونا	تشخیص
۲۱۲۱/۶۲	۸۴۱	کرونا	مدیریت	۱۵۵۲۸/۸۹	۳۱۲۹	کووید۱۹	تشخیص
۱۳۶۶/۵۵	۷۵۰	کرونا	آمار	۲۳۷۱/۸۸	۲۹۹۴	کرونا	بیمار
۲۵۲۹/۰۳	۷۰۰	کرونا	پاندمی	۳۳۸۷/۵۱	۲۷۴۰	کرونا	موضوع
۱۳۰۱/۱۶	۶۸۲	کرونا	درمان	۳۰۲۲/۰۴	۲۶۹۹	کرونا	وضعیت
۲۳۵۹/۲۱	۶۵۵	کووید۱۹	ویروس	۷۹۶۰/۷۷	۲۶۹۱	کرونا	قربانی
۱۴۳۵/۰۴	۶۱۶	کرونا	وضع	۸۶۴۶/۸۴	۲۵۳۴	کرونا	اپیدمی
۲۳/۰۶۷	۵۹۰	کرونا	گرفتن	۵۷۵۷/۳۱	۲۳۲۷	کووید۱۹	شیوع
۱۵۸۶/۶۶	۵۷۹	کرونا	کنترل	۳۸۹۰/۰۰	۲۲۸۹	کرونا	انتقال

تحلیل احساسی و معنایی
عملکرد و محتوای
وبگاه‌های خبری و
شبکه‌های اجتماعی با
محوریت پیام سلامت

ادامه جدول ۱۱.

پیگیره کرونا (خبری)				پیگیره کرونا (اجتماعی)			
واژه اول	واژه دوم	بسامد	اطلاعات	واژه اول	واژه دوم	بسامد	اطلاعات
با	کووید ۱۹	۲۲۴۲	۷۳۲/۰۱۳	همه‌گیری	کرونا	۵۵۱	۱۷۸۴/۴۶
پاندمی	کرونا	۲۲۰۸	۸۸۷۳/۵۹	خطر	کرونا	۵۱۱	۱۱۵۰/۱۳
کنترل	کرونا	۲۱۳۱	۳۲۴۹/۶۱	اپیدمی	کرونا	۵۰۷	۱۹۳۶/۶۷
علامت	کرونا	۲۰۶۶	۳۸۶۰/۰۱	موضوع	کرونا	۴۸۳	۸۰۷/۶۶
همه‌گیری	کرونا	۲۰۰۱	۶۹۳۳/۰۳	آزمایش	کرونا	۴۶۵	۱۲۶۷/۳۲
مهار	کرونا	۱۸۷۳	۵۵۵۹/۷۷	علامت	کرونا	۴۴۵	۱۱۹۳/۰۹
با	کرونا ویروس	۱۷۷۶	۲۸۷۵/۱۷	روزگار	کرونا	۴۳۳	۱۲۱۰/۲۸
بحث	کرونا	۱۷۵۹	۳۲۶۸/۸۸	شروع	کرونا	۴۲۶	۷۹۱/۲۰
شکست	کرونا	۱۷۴۲	۴۳۸۲/۲۴	مهار	کرونا	۴۱۳	۱۵۱۱/۴۲
واکسن	کووید ۱۹	۱۶۲۲	۵۸۹۷/۴۱	دوره	کرونا	۳۹۲	۴۴۰/۶۳
پسا	کرونا	۱۵۳۴	۷۲۹۹/۶۷	پایان	کرونا	۳۷۳	۴۷۹/۳۹
گسترش	کرونا	۱۵۰۳	۲۹۸۲/۳۳	گسترش	کرونا	۳۵۷	۹۶۸/۲۹
درگیر	کرونا	۱۴۰۴	۳۱۲۳/۷۵	بدون	کرونا	۳۵۱	۳۶۵/۰۷
آمار	کرونا	۱۳۲۱	۱۱۸۸/۵۷	شکست	کرونا	۳۴۷	۷۸۱/۰۷
درباره	کرونا	۱۲۸۶	۶۹۱/۸۹	خبر	کرونا	۳۴۷	۸۵/۲۴
تاثیر	کرونا	۱۱۰۰	۱۶۲۵/۴۱	بدلیل	کرونا	۳۴۳	۱۱۰۷/۴۱
غریبالگیری	کرونا	۱۰۸۵	۲۳۰۵/۶۲	ترس	کرونا	۳۴۳	۷۶۶/۴۷
از	کرونا ویروس	۱۰۷۵	۸۱۷/۶۸	سوارکاری	کرونا	۳۴۲	۱۴۷۰/۲۹
زنجیره	کرونا	۹۷۲	۱۹۸۳/۸۲	زمان	کرونا	۳۳۲	۱۵۱/۸۲
علت	کرونا	۸۰۰	۸۲۵/۰۷	فاجعه	کرونا	۳۱۴	۹۴۱/۷۷

براساس یافته‌ی جدول ۹ و دو واژه‌ای‌های شاهد در جدول ۱۱، تنوع عبارات ساخته‌شده با واژه «کرونا» در وبگاه‌های خبری و شبکه‌های اجتماعی با بسامد ۱۰۸ و ۱۵۳ مورد، از بیشترین میزان برخوردار است. همان‌گونه که روشن است، بیشترین تنوع ساختی عبارات در داده‌های شبکه‌های اجتماعی دیده می‌شود. عبارات ساخته‌شده با سایر واژه‌های کرونا‌یی در وبگاه‌های خبری به ترتیب عبارت‌اند از کووید ۱۹ (۴۳ مورد)، کرونا ویروس (۳۸ مورد)، کوید ۱۹ (۹ مورد) و کورونا (۲ مورد). این تنوع عبارات در داده‌های شبکه‌های اجتماعی براساس

بسامد، عبارت است از کووید ۱۹ (۲۷ مورد)، کوید ۱۹ (۸ مورد)، کرونا (۶ مورد)، کروناویروس (۵ مورد) و کرونایی (۱ مورد). از مقایسه عبارات ساخته شده در این دو پیکره، می‌توان دریافت که از واژه‌ها تخصصی مانند کروناویروس و کووید ۱۹ در وبگاه‌های خبری بیشتر از شبکه‌های اجتماعی استفاده شده و پیام‌های سلامت نگاشته شده، جنبه تخصصی دارند و این احتمال می‌رود که برای مخاطب عام، قابل درک نباشند.

اگر امتیاز اطلاعات فرهنگ‌نگار بالاتر از ۱۰۰۰ را مد نظر قرار دهیم، می‌بینیم که عبارات ساخته شده در وبگاه‌های خبری، در مقایسه با شبکه‌های اجتماعی، از تنوع بیشتری (به ترتیب ۵۹ مورد در مقابل ۳۷ مورد) برخوردارند و این دسته از عبارات، با واژه «کرونا» بیشترین تنوع را ایجاد کرده‌اند. ۲۰ ترکیب ساخته شده با واژه «کرونا» که دارای بالاترین امتیاز اطلاعات فرهنگ‌نگار در میان این دو پیکره بوده‌اند و رابطه باهم‌آیی بالایی نسبت به سایر عبارات داشته‌اند، به ترتیب عبارت‌اند از: ویروس کرونا، شیوع کرونا، بیماری کرونا، تست کرونا، بحران کرونا، واکسن کرونا، ستاد کرونا، دوران کرونا، مدیریت کرونا، آزمایش کرونا، پاندمی کرونا، اپیدمی کرونا، قربانی کرونا، همه‌گیری کرونا، مهار کرونا، درمان کرونا، علامت کرونا، کنترل کرونا، وضعیت کرونا و آمار کرونا. در جدول ۱۲ توزیع آماری این عبارات، در دو پیکره کرونای خبری و اجتماعی گزارش شده است. براساس این شواهد، می‌توان دریافت که پرسامدترین پیام‌های سلامت مربوط به کرونا به اطلاع‌رسانی درمورد این بیماری، پیشگیری، درمان و همچنین مدیریت و کنترل آن معطوف بوده‌اند. هرچند در ساخت عبارات موجود در پیام‌های سلامت، علاوه بر واژه‌های لاتین تخصصی، مانند ویروس، تست، پاندمی و اپیدمی، از معادل فارسی آنها، مانند بیماری، آزمایش، همه‌گیری و شیوع، نیز استفاده شده است، میل به کاربرد از واژه‌های تخصصی بیشتر بوده است، امری که ممکن است درک پیام سلامت را برای مخاطب عام با مشکل روبه‌رو کند. شایان ذکر است که عبارت «ویروس کرونا» در مقایسه با عبارت «بیماری کرونا» جزء پرکاربردترین هشتگ‌های پیکره کرونای اجتماعی بوده و میل بیشتر به استفاده از واژه‌های تخصصی در هشتگ‌های کرونایی نیز مشهود است.

بررسی واژه‌هایی که پیش از کلیدواژه‌های کرونایی در جدول ۱۲ قرار گرفته‌اند بسیار بااهمیت است. در ادامه از دو جنبه، به بررسی این واژه‌ها می‌پردازیم. این دو جنبه عبارت‌اند از: نخست، تحلیل احساسات واژه‌های به کاررفته و سپس،

تحلیل احساسی و معنایی
عملکرد و محتوای
وبگاه‌های خبری و
شبکه‌های اجتماعی با
محوریت پیام سلامت

خوشه‌بندی این واژه‌ها براساس تشابه معنایی. همسو با این هدف، خود را به واژه‌های محتوایی محدود می‌کنیم و از واژه‌های دستوری، مانند حروف اضافه، چه ساده و چه مرکب، چشم می‌پوشیم.

جدول ۱۲. توزیع آماری ۲۰ عبارت دوواژه‌ای کرونایی دارای بالاترین امتیاز اطلاعات فرهنگ‌نگار

عبارت دوواژه‌ای		بیکرة کرونا (خبری)							بیکرة کرونا (اجتماعی)	
		باشگاه خبرنگاران جوان	صداسیما	خبرگزاری فارس	ایرنا	اسنا	شبکه خبر	وزارت بهداشت	اینستاگرام	توییتر
		هشتگ	نظر	هشتگ	نظر	هشتگ	نظر	هشتگ	نظر	هشتگ
ویروس کرونا	۱۳۳۹۰۹	۵۵۴۸۰	۱۲۷۰	۵۹۳۳۴	۳۷۱۵	۱۲۳۳	۷۷۶۶	۴۵	۳۰۱۰	۳۰
شیوع کرونا	۲۹۵۱۱	۱۴۵۱۵	۵۷۵	۲۳۳۷۲	۱۳۴۸	۵۹۵	۱۷۹۷	۷	۹۲۲	۲
بیماری کرونا	۴۸۶۹۶	۲۵۰۱۷	۴۴۲	۲۴۶۰۷	۱۸۱۷	۱۲۶۵	۲۲۷۰	۴	۶۹۴	۱
تست کرونا	۷۰۹۸	۲۸۳۹	۶۸	۳۶۱۰	۱۶۰	۸۰	۳۳۲	۲	۴۳۹	۱
بحران کرونا	۵۷۹۶	۳۱۲۸	۶۵	۵۲۵۵	۲۱۴	۲۱۳	۶۷۸	۳	۴۸۴	۱
واکسن کرونا	۳۵۸۳	۱۵۵۴	۱۵۶	۲۵۹۱	۷۱۴	۳۶۶	۲۷۹	۴	۸۸۶	۵
ستاد کرونا	۴۱۴۲	۱۸۰۲	۸۳	۳۰۰۸	۱۵۴	۷۲	۱۷۱	۲	۴۴۵	۶
دوران کرونا	۲۲۲۱	۸۴۶	۶۱	۲۲۰۹	۱۱۱	۱۵۳	۵۷۵	۶	۲۳۸	۰
مدیریت کرونا	۲۹۲۱	۱۴۳۳	۶۳	۱۸۶۸	۱۶۹	۳۱۹	۱۳۱	۰	۶۰	۰
آزمایش کرونا	۲۰۴۴	۱۹۰۰	۲۳	۵۲۹	۱۴۳	۱۴	۶۱	۰	۴۷	۱
پاندمی کرونا	۴۵۸	۲۳۶	۲۵	۱۳۷۹	۲۴	۸۶	۱۷۱	۱	۶۱	۰
اپیدمی کرونا	۱۰۰۰	۲۸۵	۱۰	۱۱۰۳	۳۸	۹۸	۹۰	۰	۶۰	۱
قربانی کرونا	۹۲۴	۱۱۹۳	۵	۴۶۱	۱۰۵	۳	۱۳۰	۲	۷۰۱	۳
همه‌گیری کرونا	۸۴۳	۲۷۶	۳۷	۷۵۵	۹۱	۱۹	۱۷۹	۰	۳۷	۰
مهار کرونا	۶۸۶	۵۶۵	۲۴	۴۸۴	۶۷	۴۷	۵۸	۱	۹۳	۲
درمان کرونا	۱۴۸۱	۵۷۲	۶	۸۵۲	۷۷	۲۰۸	۱۸۰	۴	۲۴۱	۳
علامت کرونا	۱۰۴۰	۴۵۸	۱۷	۵۱۷	۱۹	۱۵	۰	۰	۸	۰
کنترل کرونا	۹۶۲	۴۳۸	۱۴	۶۰۸	۳۷	۷۲	۹۱	۰	۸۰	۰
وضعیت کرونا	۱۱۸۰	۹۲۵	۳۰	۹۱۸	۵۲	۲۸	۱۵۷	۰	۲۳۸	۰
آمار کرونا	۴۵۷	۳۶۱	۲	۴۰۸	۶۱	۳۲	۱۱۷	۳	۳۳۳	۳

با حذف واژه‌های محتوایی تکراری و واژه‌های نقشی در پیکره‌های خبری و اجتماعی کرونا، به مجموعه‌ای از ۱۴۱ واژه یکتا دست می‌یابیم. برای آنکه بتوانیم به تحلیل دقیق‌تری دست پیدا کنیم، تحلیل خود را به ۱۰۰ واژه یکتای پربسامد محدود می‌کنیم. از میان واژگان صدتایی یکتا، ۵۷ واژه، در پیکره‌های خبری و اجتماعی کرونا، مشترک و بقیه، در پیکره‌ها غیرمشترک هستند. در جداول ۱۳ و ۱۴ واژه‌های مشترک و غیرمشترک به ترتیب گزارش شده‌اند. در واژه‌های غیرمشترک، تعدادی از واژه‌های تخصصی، مانند ژن، پاندمی، آزمایشگاه، غربالگری، سرولوژی و واکسیناسیون، در پیکره خبری کرونا به چشم می‌خورند اما واژه‌هایی، مانند انیمیشن، طنز و سوارکاری نیز در پیکره اجتماعی آن دیده می‌شوند که بیشتر بر مسائل فرهنگی و تفریحی متمرکزاند.

جدول ۱۳. واژه‌های مشترک پیش از کلیدواژه‌های کرونا در پیکره‌های خبری و اجتماعی کرونا

نابودی	قربانی	شر	جولان	بیمار	اپیدمی
ناقل	کاسب	شروع	خسارت	بیماری	اپیدمیولوژی
همه‌گیری	کنترل	شکست	خطر	پایان	اثر
واکسن	گرفتار	شیوع	درمان	پادتن	آزمایش
ورود	گسترش	ضد	دوران	پاندمی	آمار
ویروس	مبتلا	عارضه	روزگار	پسا	انتقال
وضعیت	مدیریت	علامت	زنجیره	پیک	اوج
	مهار	علیه	سایه	تست	بحبوحه
	موج	فاجعه	ستاد	تشخیص	بحث
	موضوع	فراگیری	سرایت	تهدید	بحران

تحلیل احساسی و معنایی
عملکرد و محتوای
وبگاه‌های خبری و
شبکه‌های اجتماعی با
محوریت پیام سلامت

جدول ۱۴. واژه‌های غیرمشترک پیش از کلیدواژه‌های کرونا در پیکره‌های خبری و اجتماعی کرونا

پیکره کرونا (اجتماعی)				پیکره کرونا (خبری)			
آزمایشگاه	تأثیر	طغیان	محوریت	اتمام	بودجه	سیل	گرفتن
انتشار	تکثیر	ظهور	مخصوص	آسیب	تأثیر	سنگدلی	محدودیت
اوجگیری	جریان	عقونت	مساله	اعتبار	ترس	سوارکاری	مراقب
باتلاق	چالش	علت	معرض	انفجار	تفاوت	سونامی	مریض
بروز	داستان	غربالگیری	معضل	انیمیشن	توهم	شرح	مسأله
بستری	درگیر	قضیه	منشا	اهمیت	جناب	طنز	منفعت
پیامد	ریشه‌کنی	کابوس	نشانه	اوایل	حصر	فهمیدن	موقع
پاندمیک	ژن	کشندگی	نوع	بالای	خبر	قربانگاه	نام
پیدایش	ژنوم	گزند	همزمانی	بدون	دچار	کارگروه	وضع
پدیده	سرولوژی	ماندگاری	واکسیناسیون	بیرحمی	رفع	کشتار	وقت
پیرامون	شکل	مبتلا به		بلای	زمان	کوران	

حال که مشخص شد با چه واژه‌هایی در این دو پیکره سروکار داریم، به تحلیل احساسات این واژگان می‌پردازیم. برای به‌دست‌آوردن یک نگاه کلی در تحلیل احساسات واژگان به‌کاررفته در این دو دسته، با صرف نظر از تفکیک واژه‌های مشترک یا غیرمشترک میان این دو، به تحلیل داده‌ها می‌پردازیم. برای این هدف از مدل معرفی‌شده از سوی توحیدی و همکاران^۱ (۲۰۲۵) در تحلیل احساسات به‌صورت الگوریتمی بهره می‌گیریم. شایان ذکر است این واژگان، متشکل از واژه‌های یکتایی هستند که ممکن است در پیکره‌ها با تکرار همراه باشند. در ادامه، توزیع آماری واژه‌ها را در جدول ۱۵ بررسی خواهیم کرد.

جدول ۱۵. تحلیل احساسات واژه‌ها در پیکره‌های کرونای خبری و اجتماعی

پیکره کرونا (خبری)		پیکره کرونا (اجتماعی)		
مثبت	منفی	مثبت	منفی	
۴۲	۵۸	۳۰	۷۰	واژه یکتا
۶۹۵۱۵	۵۳۵۵۵۲	۱۷۳۶۷	۷۱۴۴۲	واژه با تکرار

با نگاهی به جدول ۱۵ مشخص می‌شود که تعداد واژه‌های به کار رفته دارای احساس منفی در پیام‌های سلامت‌محور و بگانه‌های خبری و شبکه‌های اجتماعی، نسبت به واژه‌های دارای حس مثبت، بیشتر و غالب است. تعداد واژه‌های یکتای دارای حس منفی استفاده شده در پیکره کرونای اجتماعی نیز از پیکره کرونای خبری بیشتر است؛ این یافته بیان می‌کند که در شبکه‌های اجتماعی، از تنوع واژگانی بیشتری برای انتقال پیام سلامت‌محور دارای حس منفی استفاده می‌شود. با این حال، باتوجه به توزیع گسترده خبر در وبگاه‌های خبری از نظر تکرار، واژه‌های دارای حس منفی، در پیکره کرونای خبری، بسامد تکرار بیشتری دارند و این نشان می‌دهد که واژه‌های دارای حس منفی، هر چند محدود از نظر تنوع، توزیع آماری‌شان بیشتر است. این توزیع آماری را می‌توان از جنبه دیگری نیز بررسی کرد و آن، شدت انتشار پیام‌های سلامت کرونایی دارای حس منفی است. شدت توزیع آماری واژه‌های مرتبط با کرونای دارای حس منفی ۷/۷ برابر پیام‌های سلامت مثبت خبری و ۱۱/۴ برابر پیام‌های سلامت مثبت اجتماعی بوده است. واژه‌هایی که با کلیدواژه‌های کرونا همراه بوده‌اند، از جنبه تشابه معنایی، بررسی شده‌اند و با استفاده از الگوریتم کی-مینز (مک‌کویین^۱، ۱۹۶۷)، آنها را به ۱۰ خوشه، تقسیم و هر خوشه را براساس واژگان موجود در آن، نامگذاری کرده‌ایم. نتایج به دست آمده از تفکیک و تجمیع پیکره‌ها در جدول ۱۶ گزارش شده است.

جدول ۱۶. تشابه معنایی واژه‌های به کار رفته با کلیدواژه‌های کرونایی

عنوان خوشه	توزیع آمار تجمیعی	پیکره کرونا (خبری)	پیکره کرونا (اجتماعی)
احساسات، واکنش‌های روانی و ادراکی	۸	۱	۷
انتقال، سرایت و گسترش	۱۳	۷	۶
بیماری‌ها، همه‌گیری و پدیده‌های اپیدمیک	۲۲	۱۴	۸
تهدید، خطر و پیامدهای منفی	۲۹	۱۴	۱۵
زمان، دوره و وضعیت	۲۱	۸	۱۳
علل، آثار و تحلیل مفهومی	۲۱	۱۳	۸
علوم پزشکی، زیستی و بهداشت	۲۲	۱۵	۷
مدیریت، کنترل و سیاست‌گذاری	۱۹	۸	۱۱
واژگان عمومی، توصیفی یا متفرقه	۲۲	۸	۱۴
وضعیت بیمار، درمان و پیامدهای جسمی	۲۳	۱۲	۱۱

براساس عناوین معرفی شده برای خوشه‌های معنایی واژه‌های به کاررفته با کلیدواژه‌های کرونایی، حدود یک سوم واژه‌های استفاده شده از نظر «تهدید، خطر و پیامدهای منفی» تشابه معنایی دارند و بالاترین بسامد را به خود اختصاص داده‌اند. به عبارت دیگر، در پیام‌های سلامت تلاش شده است از واژه‌های دارای تشابه معنایی برای پیامدهای منفی کرونا استفاده گردد. کوچک‌ترین خوشه‌ای که کمترین تعداد واژه‌های دارای تشابه معنایی را در خود داشته، خوشه «احساسات، واکنش‌های روانی و ادراکی» بوده و خوشه «انتقال، سرایت و گسترش» در یک پله بالاتر قرار گرفته است. این دو خوشه کم بسامد، مؤید این نکته‌اند که از واژه‌های متنوعی، برای بیان اثرات روانی بیماری کرونا و همچنین نکات مربوط به انتقال و سرایت آن استفاده نشده و بیان این مفاهیم، متنوع نبوده است.

با بررسی دقیق‌تر خوشه‌های معنایی در دو پیکره کرونا، این نتیجه به دست آمد که تشابه معنایی واژه‌ها با مفهوم «علوم پزشکی، زیستی و بهداشت» در پیکره کرونای خبری پربسامدتر است. استفاده از واژه‌های تخصصی پزشکی که در واژگان غیرمشترک در پیکره‌های خبری و اجتماعی در جدول ۱۴ گزارش شده بود، مؤید این یافته است. با وجود این، تشابه معنایی واژه‌ها، با مفهوم «تهدید، خطر و پیامدهای منفی» در پیکره کرونای اجتماعی پربسامدترین است که این نتیجه، همراستای تحلیل وجود تنوع واژگانی در واژگان دارای احساس منفی در پیکره کرونای اجتماعی در جدول ۱۵ است.

بحث و نتیجه‌گیری

در این پژوهش، به تحلیل پیام‌های سلامت در محتوای وبگاه‌های خبری و شبکه‌های اجتماعی، همچنین، عملکرد دو منبع اطلاع‌رسانی در حوزه سلامت پرداخته شده است. برای محدود کردن دامنه پژوهش به همه‌گیری کرونا، ابتدا، یک پیکره زبانی را در مورد کرونا، از این دو منبع اطلاع‌رسانی تهیه کرده‌ایم تا سپس، به صورت الگوریتمی، پردازش‌های متنوعی را روی آنها اعمال کنیم. یافته‌های حاصل از پژوهش عبارت‌اند از:

اطلاع‌رسانی و انتشار خبر، جزء رسالت خبرگزاری‌هاست. انفعال این منابع بر سلامت عموم تأثیرگذار است. انفعال در افزایش آگاهی، می‌تواند به افزایش مرگ و میر در زمان همه‌گیری منجر شود.

شبکه‌های اجتماعی در غیاب عملکرد منابع رسمی، به فعالیت بیشتری می‌پردازند تا خلاء موجود را پر کنند اما به‌طور معمول، نسبت به منابع رسمی، چندان اثرگذار نیستند.

ترکیبات و عبارات ساخته‌شده با کلیدواژه‌های کرونایی واژه‌های عمومی و تخصصی هستند. درک نشدن واژه‌های تخصصی به‌کاررفته در پیام‌های سلامت، رسالت آگاهی‌سازی و اطلاع‌رسانی را برای عموم افراد جامعه غیرمتخصص مختل می‌کند و ممکن است ارتباط زبانی، آن‌طور که انتظار می‌رود، برقرار نشود.

با بررسی واژه‌های محتوایی پیام‌های سلامت کرونا در وبگاه‌های خبری و شبکه‌های اجتماعی، این نتیجه به‌دست آمد که حس غالب منتقل‌شده در پیام‌ها، منفی است؛ اگر انتشار پیام‌های سلامت را مد نظر قرار دهیم، می‌بینیم که شدت انتشار خبرهای سلامت منفی در وبگاه‌های خبری، بیشتر از شبکه‌های اجتماعی است.

با بررسی معنایی واژه‌های به‌کاررفته همراه با کلیدواژه‌های کرونا، مشخص شد که پرسامدترین واژه‌های دارای تشابه معنایی در وبگاه‌های خبری، متعلق به حوزه «علوم پزشکی، زیستی و بهداشت» بوده‌اند، امری که نشان‌دهنده تخصصی عمل کردن وبگاه‌های خبری برای انتقال پیام سلامت است و ممکن است موجب شود که ارتباط با مخاطب به‌خوبی برقرار نشود درحالی‌که واژه‌های مشابه در شبکه‌های اجتماعی، ذیل عنوان «تهدید، خطر و پیامدهای منفی» مطرح می‌شوند که با یافته انتقال حس منفی با واژگان متنوع همسو است.

نوآوری و محدودیت پژوهش

هم‌افزایی میان وبگاه‌های خبری و شبکه‌های اجتماعی می‌تواند به استفاده همزمان از ظرفیت سرعت و گستردگی انتشار در شبکه‌های اجتماعی و دقت در خبرگزاری‌ها منجر شود. همچنین بهره‌گیری از تحلیل زبانی و معنایی هنگام پایش پیام‌های سلامت‌محور می‌تواند کیفیت تولید محتوای پیام سلامت را بهبود داده و اثرگذاری پیام سلامت‌محور بر مخاطب را باتوجه به دستاوردهای پژوهش حاضر افزایش دهد. ازجمله نوآوری‌های پژوهش حاضر تحلیل آماری پیکره‌بنیان پیام‌های سلامت‌محور در حوزه کرونا از دیدگاه زبان‌شناسی است. در این تحلیل دو ویژگی انتقال احساسی مفاهیم واژگانی و همچنین دسته‌بندی معنایی واژه‌های مرتبط به لحاظ همزمانی رخداد با کلیدواژه کرونا مد نظر قرار

داده است. توجه به تنوع داده‌ها در شبکه‌های اجتماعی و وبگاه‌های خبری در یک بازه زمانی مشخص، ویژگی دیگر پژوهش حاضر است که سبب می‌شود داده‌های پژوهش مقایسه‌پذیر شوند. ازجمله مسائلی که به‌عنوان چالش و محدودیت در انجام پژوهش حاضر مطرح می‌شود عبارت است از تنوع در نگارش واژه‌های مربوط به بیماری کرونا، وجود تنوع در استفاده از واژه‌های تخصصی حوزه پزشکی و همچنین دشواری و اعمال محدودیت در گردآوری داده‌های زبانی از وبگاه‌های خبری و شبکه‌های اجتماعی. به‌طورقطع افزایش حجم داده‌ها و تنوع‌بخشی به آنها می‌تواند به تحلیل دقیق‌تری از پیام‌های سلامت منجر گردد. دستیابی به این اهداف بیانگر نیاز به وجود یک سامانه پایشگر خبری به‌طور پیوسته و مداوم در موضوعات متنوع حوزه سلامت است؛ چراکه افزایش آگاهی به افزایش سواد سلامت می‌انجامد.

سلامتی، نتیجه انتخاب‌های کوچک اما مداوم هر روز است. انتشار روزانه اخبار سلامت‌محور می‌تواند در تداوم سلامتی اثرگذار باشد. بنابراین، رسانه و ارتباط با مخاطب، از جایگاه بااهمیتی برخوردار است. براساس نتایج حاصل از بررسی داده‌های این پژوهش، می‌توان نتیجه گرفت که پیام‌های سلامت باید به زبانی ساده و غیرتخصصی منتشر شوند تا ضمن کمک به حفظ آرامش در جامعه و پرهیز از بزرگ‌نمایی، بر حس منفی پیام، غلبه کنند و برای عامه مردم جامعه قابل درک باشند؛ چراکه رسالت رسانه اطلاع‌رسانی است و در حوزه سلامت، مسئولیت‌پذیری اجتماعی دارد؛ افزایش آگاهی به خودمراقبتی می‌انجامد و تا حدودی مانع از شکل‌گیری بحران در جامعه می‌شود.

منابع

احمدی، ثریا و المظهری، فریبا. (۱۳۹۷). بازنمایی پیام‌های سلامت‌محور در مطبوعات: تحلیل مضمونی مطالب روزنامه‌های ایران، همشهری و جام جم. رسانه، ۲۹ (۱)، ۱۲۷-۱۴۴.

Doi: 20.1001.1.10227180.1397.29.1.6.6

خانیکی، هادی و راثی‌تهرانی، حبیب. (۱۳۹۰). بررسی نگرش مخاطبان درباره تأثیر پیام‌های سلامت تلویزیون. مطالعات فرهنگ-ارتباطات، ۱۲ (۱۵)، ۱۸۷-۱۵۱.

Doi: 20.1001.1.20088760.1390.12.15.5.4

رضوی طوسی، سیدمجتبی، آذرهمایون، رمضان، یاهک، سجاد و غلامی‌پور، اسماعیل. (۱۳۹۲). برساخت پیام‌های سلامت و پزشکی در رسانه و تأثیر آن بر آگاهی مردم. *فصلنامه اخلاق زیستی*، ۳(۸)، ۴۳-۷۷.

Doi: 10.22037/bioethv3i8.13967

عظیمی‌فرد، فاطمه. (۱۳۹۹). پیام‌های سلامت در رسانه؛ مطالعه موردی کرونا. سخنرانی ارائه‌شده در همایش ملی ابعاد انسانی- اجتماعی کرونا در ایران، پژوهشگاه علوم انسانی و مطالعات فرهنگی، تهران.

Bulut, O., & Poth, C. N. (2022). Rapid assessment of communication consistency: sentiment analysis of public health briefings during the COVID-19 pandemic. *AIMS Public Health*, 9(2), 293-306.

Chakraborty, K., Bhatia, S., Bhattacharyya, S., Platos, J., Bag, R., & Hassanien, A. E. (2020). Sentiment analysis of COVID-19 tweets by deep learning classifiers—A study to show how popularity is affecting accuracy in social media. *Applied Soft Computing Journal*, 97, 106754. Doi: 10.1016/j.asoc.2020.106754

Church, K., Gale, W., Hanks, P., & Hindle, D. (1991). Using statistics in lexical analysis. In U. Zernic (Ed), *Lexical Acquisition* (pp. 115-164). Englewood Cliff, NJ: Erlbaum. Doi:10.4324/9781315785387-8

de Saussure, F. (1916). *Cours de Linguistique Generale*. Lausanne, Paris: Payot.

Evert, S. (2005). *The Statistics of Word Co-occurrences Word Pairs and Collocations*. [Doctoral dissertation, Universität Stuttgart].

Kilgariff, A., Rychlý, P., Smrz, P., & Tugwell, D. (2004). The Sketch Engine. In G. Williams & S. Vessier (Eds), *Proceedings of the Eleventh Euralex Congress*, pp. 105-116. Lorient, France: UBS.

Li, Z., Zhou, O., & Zhang, R. (2025). The persuasion effects of vaccination message: The interactive effects of language style, source type, and the provision of choice. *Science Communication*, 48(2), 248-279. Doi: 10.1177/10755470251314725

MacQueen, J. (1967). Some methods for classification and analysis of multivariate observations. In *Proceedings of the 5th Berkeley Symposium on Mathematical Statistics and Probability*, volume 1, pp. 281-297.

Morrison, F. P., Kukafka, R., & Johnson, S. B. (2005). Analyzing the structure and content of public health messages. In *Proceedings of the AMIA Symposium*, pp. 540-544.

Nutbeam, D. (2000). Health literacy as a public health goal: A challenge for contemporary health education and communication strategies into the 21st century. *Health Promotion International*, 15(3), 259-267. Doi: 10.1093/heapro/15.3.259

Salahshouri, A., Araban, M., Heydarabadi, A., Hatamzadeh, N., Jasemzadeh, M., & Mohamadian, H. (2024). Assessing the quality of COVID-19 messages on Ahvaz Jundishapur University of Medical Sciences Website: A content analysis. *The Open Public Health Journal*, 17: e18749445340225.

Doi: 10.2174/0118749445340225240904061656

Szkody, E., & McKinney, C. (2021). Evaluating comforting messages with the sentiment analysis and social cognition engine. *Applied Psychology: Health and Well-Being*, 14, 2: 645-662.

Doi: 10.1111/aphw.12324

Tohidi, K., Dashtipour, K., Rebora, S., & Pourfaramarz, S. (2025). A comparative evaluation of large language models for Persian sentiment analysis and emotion detection in social media texts. arXiv:2509.14922. <https://arxiv.org/abs/2509.14922>

Udvardi, A. (2019). The role of linguistics in improving the evidence base of healthcare communication. *Patient Education and Counseling*, 102(2), 388-393. Doi: 10.1016/j.pec.2018.09.012