

Content Analysis of Topics and Hashtags about the Coronavirus in Social Media

Masood Ghayoomi * 

Assistant Professor, Institute of Linguistics,
Institute of Humanities and Cultural Studies,
Tehran, Iran

Abstract

Coronavirus pandemic caused changes in the daily lifestyle, such as reducing social interactions and creating social distancing. In this research, we pursue two goals. One is algorithmic content analysis of comments/posts in Persian related to the Coronavirus on two social media, namely Tweeter and Instagram. To this end, topic modeling is used as a method for content analysis to cluster the data into abstract topics. The other goal is finding the correlation between topics and hashtags in the comments/posts. To this end, we developed a corpus from these two social media. We found 24 abstract topics by algorithmic content analysis of this corpus and they were manually labeled to be comprehensive. According to the corpus and the statistical information of the extracted topics, it can be speculated that about 25% of the comments/posts in this corpus focused on political and social issues of the virus. 10 fine-grained topics which contained 35% of the comments were related to the Coronavirus itself and its pandemic property. This indicates the importance of the attention that has been paid to social media for informing and disseminating information. Furthermore, the hypothesis of existing correlation between topics and hashtags was studied from statistical point of view by using the Pearson correlation coefficient. For 20 topics, a high correlation score between topics and hashtags was found; but this correlation was not found for 4 topics. The outcome of this research can be used to increase the internal coherence of a text and to make the hashtags predictable.

Keywords: social media, Coronavirus, COVID-19, hashtag, topic modeling, content analysis.

* Corresponding Author: M.Ghayoomi@ihcs.ac.ir

How to Cite: Ghayoomi, M. (2021). Content analysis of topics and hashtags about the Coronavirus in social media. *Language Science*, Corona Special Letter, Vol. 8, 87-115. Doi: 10.22054/lis.2020.53557.1356

تحلیل محتوایی موضوع‌ها و هشتک‌های کرونا در رسانه‌های اجتماعی

استادیار، پژوهشکده زبان‌شناسی، پژوهشگاه علوم انسانی و مطالعات فرهنگی، تهران، ایران

مسعود قیومی * ID

چکیده

بیماری همه‌گیر کرونا موجب تغییراتی در سبک زندگی روزمره، مانند کاهش مرادوات اجتماعی و ایجاد فاصله‌گذاری اجتماعی، شده است. در این پژوهش، دو هدف را پیگیری کرده‌ایم. هدف نخست تحلیل الگوریتمی نظرات و پیام‌های مرتبط با کرونا به زبان فارسی است که توسط افراد مختلف، در دو رسانه اجتماعی توئیتر و اینستاگرام، منتشر می‌شود. برای این هدف، با استفاده از مدل‌سازی موضوعی، به‌عنوان یک شیوه تحلیل محتوایی داده، داده‌ها را به موضوعات انتزاعی خوشه‌بندی کرده‌ایم. هدف دیگر، یافتن رابطه همبستگی بین موضوعات در نظرات و هشتک‌های به‌کاررفته در نظرات است. برای تحقق این هدف، با گردآوری نظرات و پیام‌های خزش‌شده مربوط به کرونا از این دو رسانه، یک پیکره زبانی تهیه شده است. از تحلیل محتوایی نظرات این پیکره، ۲۴ موضوع انتزاعی به دست آمد که به‌صورت دستی برچسب‌گذاری شد تا این موضوع‌ها هویت پیدا کنند. با بررسی داده‌های این پیکره و آمار به‌دست آمده از پردازش موضوعی آن‌ها، می‌توان گمانه‌زنی کرد که حدود ۲۵٪ از نظرات این پیکره بر دو موضوع «سیاسی» و «اجتماعی» متمرکز است. ۱۰ موضوع دانه‌ریز این پیکره، یعنی ۳۵٪ از حجم نظرات، به خود ویروس کرونا و ماهیت همه‌گیری آن مربوط است که بیانگر اهمیت توجه به رسانه‌های اجتماعی برای آگاه‌سازی و نشر اطلاعات است. علاوه بر آن، فرضیه وجود رابطه بین هشتک‌ها و موضوع‌ها، با استفاده از ضریب همبستگی پیرسون، از نظر آماری مورد بررسی قرار گرفت. برای ۲۰ موضوع، همبستگی بالا بین موضوع و تعدادی از هشتک‌ها یافت شد؛ ولی برای ۴ موضوع این همبستگی یافت نشد. از این پژوهش می‌توان برای افزایش انسجام درونی متن و پیش‌بینی پذیرش هشتک‌ها استفاده کرد.

کلیدواژه‌ها: رسانه‌های اجتماعی، ویروس کرونا، کووید ۱۹، هشتک، مدل‌سازی موضوع، تحلیل محتوایی.

۱. مقدمه

با شروع دهه دوم قرن ۲۱ میلادی، یک بیماری ویروسی به نام کرونا^۱ (کووید ۱۹)^۲ از شرق آسیا شروع شد و به دلیل سرعت انتشار، در مدت زمان کوتاهی، خاورمیانه، از جمله ایران را فراگرفت و سپس، به اروپا و آمریکا رسید. در تاریخ ۱۱ مارس ۲۰۲۰ (۲۱ اسفند ۱۳۹۸)، سازمان بهداشت جهانی همه‌گیری بیماری کرونا را اعلام کرد. یکی از ابتدایی‌ترین راهکارهایی که توسط دولت‌ها، در کشورهای مختلف، مورد تأکید قرار گرفت، پیشگیری از ابتلا به این بیماری، با ایجاد فاصله‌گذاری بین افراد جامعه به واسطه قرنطینه فردی و اجتماعی بود. از آنجا که ویژگی اجتماعی بودن در سرشت آدمی نهاده شده است، اعمال قرنطینه موجب شد این سرشت انسانی با محدودیت روبه‌رو شود. برای رفع این خلأ، نوع دیگری از زندگی اجتماعی ظهور پیدا کرد و آن پررنگ شدن استفاده از ابزارهای نوین ارتباطی از طریق اینترنت بود. وجود رسانه‌های اجتماعی، مانند توئیتر و اینستاگرام، موجب شد خلأهای ناشی از قرنطینه پر شود و کاربران نوع دیگری از زندگی را، در فضایی نزدیک به واقعیت، تجربه کنند. گذشت زمان و تولید حجم فراوانی از نظرات و پیام‌ها توسط افراد جامعه در این بستر موجب جلب توجه پژوهشگران به رسانه‌های اجتماعی شد. به طور کلی، به دو دلیل، تحلیل اطلاعات موجود در رسانه‌های اجتماعی از اهمیت بسزایی برخوردار است. دلیل اول این است که ماهیت این رسانه‌ها اطلاع‌رسانی است و از طریق آنها می‌توان به آخرین اخبار و تحولات دسترسی داشت؛ دلیل دوم این است که به دلیل آزادی بیان در این رسانه‌ها، امکان درج نظرات افراد مختلف جامعه، که به سطوح اجتماعی و فرهنگی متفاوت با گرایش‌های سیاسی مختلف تعلق دارند، وجود دارد. از این رو، بررسی حجمی نظرات اهمیت بسزایی دارد. در شرایط همه‌گیری بیماری کرونا، این اهمیت دوچندان می‌شود؛ زیرا می‌توان از این طریق، از نظرات افراد جامعه آگاه شد و در تصمیم‌گیری‌های کلان، از این نظرات استفاده کرد.

در این مقاله، می‌کوشیم مجموعه نظرات را به‌مثابه یک پیکره زبانی در نظر بگیریم و با رویکرد پیکره‌محور^۳ (Tognini-Bonelli, 2001: 84-85)، بدون هرگونه پیش‌داوری و فرضیه‌ای، به صورت الگوریتمی و خودکار، به تحلیل محتوایی نظرات

1. Coronavirus
2. COVID-19
3. corpus-driven

بپردازیم. در راستای اهداف این پژوهش، تلاش می‌کنیم موضوعات انتزاعی را، که در نظرات کاربران فارسی‌زبان دربارهٔ کرونا در دو رسانهٔ اجتماعی توئیتر و اینستاگرام نهفته است، بیابیم و با برجسب‌گذاری موضوعی نظرات، بتوانیم به اهم موضوعات مورد توجه کاربران این دو رسانهٔ اجتماعی دست یابیم. سپس، با فرضیهٔ وجود رابطهٔ همبستگی بین موضوع و هشتگ‌های استفاده‌شده در این نظرات، وجود این رابطه را به صورت آماری مورد بررسی قرار دهیم.

در بخش دوم این مقاله، مبانی نظری و مروری بر پیشینهٔ مطالعاتی متمرکز بر جنبهٔ نظری تحلیل داده‌های رسانه‌های اجتماعی ارائه می‌شود. در بخش ۳، دربارهٔ داده‌ها و ابزارهای مورد نظر در پژوهش توضیح داده می‌شود. در بخش ۴، پیکرهٔ گردآوری شده از رسانه‌های اجتماعی توئیتر و اینستاگرام و ابزاری که برای تحلیل نظرات فارسی به کار رفته است معرفی می‌شود. بخش ۵، به تحلیل محتوایی موضوعات انتزاعی به دست آمده از این پیکره و هشتگ‌ها می‌پردازد. در آخر، در بخش ۶، جمع‌بندی و نتیجه‌گیری از نظرات تحلیل شده ارائه می‌شود.

۲. مدل‌سازی موضوعی

۲-۱. مبانی نظری

سوسور^۱ (۱۹۱۶) دو سطح «صورت» و «معنا» را برای زبان تعریف کرده است. این دو سطح واقعیتی روان‌شناختی است که پیوندشان در یک نظام مشخص موجب دربرگرفتن تمام جملات زبانی بیان شده توسط گویشور یک زبان می‌شود. «صورت» و «معنا» در این نظام، واحدهای زبانی، یعنی همان واژه‌ها، را ایجاد می‌کنند. این واحدهای زبانی، براساس روابط همنشینی، در چارچوب قواعد آن نظام مشخص، قابلیت ترکیب با یکدیگر را دارند تا واحدی بزرگ‌تر، یعنی جمله، را بسازند. «صورت» می‌تواند از ذهن، به دو شکل امواج صوتی و نوشتار، تجلی عینی پیدا کند، به این صورت که امواج صوتی از طریق شنیدار و نوشتار از طریق دیدار در نظام ارتباطی منتقل می‌شوند. از آنجا که داده در رایانه به یکی از چهار شکل صوتی، متنی، تصویری و عددی تظاهر می‌یابد، خط، در حوزهٔ رایانه، به عنوان

1. de Saussure, F.

داده متنی تلقی می‌شود و بیانگر نوعی تجلی بصری از زبان است؛ بنابراین، با پردازش ماشینی داده، می‌توان به «معنی»، که جنبه انتزاعی دارد، دست یافت.

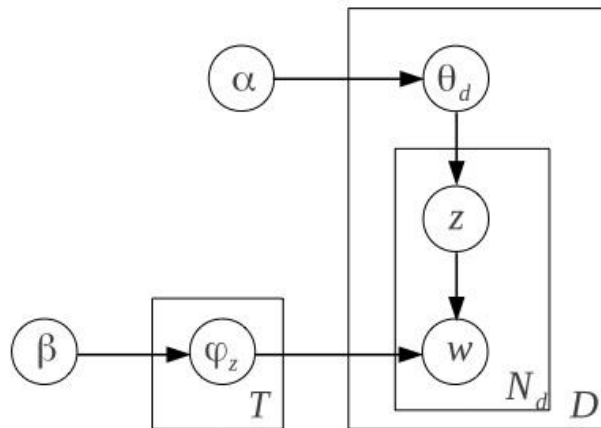
هر متن، که از ترکیب جملات در نظام مشخص زبانی شکل می‌گیرد، از مجموعه موضوعات محدودی ساخته شده است که به صورت الگوریتمی، می‌توان به این موضوعات دست یافت. یکی از الگوریتم‌های مورد استفاده، برای تحلیل محتوایی متن، «مدل‌سازی موضوع»^۱ است که توسط پاپادیمیتریو و همکاران^۲ (۲۰۰۰) معرفی شده است. این الگوریتم معمولاً در متن‌کاوی‌هایی^۳ به کار می‌رود که، همانند پژوهش حاضر، به دنبال کشف ساختار معنایی پنهان متن هستند. این الگوریتم بر این اساس شکل گرفته که هر متن از چند موضوع انتزاعی تشکیل شده است و هر مجموعه از واژه‌های خاص منجر به شکل‌گیری یک موضوع می‌شود؛ بنابراین، با یافتن واژه‌هایی که از نظر معنایی به یکدیگر شباهت دارند، می‌توان متن را به موضوع‌های انتزاعی خوشه‌بندی^۴ کرد؛ برای مثال، یک موضوع اقتصادی از واژه‌هایی مانند «چک»، «واریز»، «بانک»، «پول» و «تورم» شکل می‌گیرد و واژه‌هایی مانند «جناح»، «حزب»، «اصول‌گرا»، «میانه‌رو» و «اصلاح‌طلب» یک موضوع سیاسی را تشکیل می‌دهند.

پاپادیمیتریو و همکارانش «مدل‌سازی موضوع» را براساس جبر خطی پیشنهاد دادند، ولی هوفمن^۵ (۱۹۹۹) یک مدل احتمالاتی برای این هدف تهیه کرد. بلائی و همکاران^۶ (۲۰۰۳) مدل احتمالاتی مطرح شده را تعمیم دادند و مدلی از الگوریتم «مدل‌سازی موضوع» ارائه کردند که به «تخصیص دریشله پنهان»^۷ معروف است. این الگوریتم یک مدل آماری زایشی است که در اصل، «توزیع اولیه دریشله»^۸ بوده است و برای توزیع احتمالاتی متن-موضوع و موضوع-واژه به کار می‌رود. هر واژه در یک متن، براساس توزیع اولیه دریشله، با یک موضوع مرتبط است و هر موضوع نیز، براساس توزیع چندگانه واژه‌هایی که با یک موضوع در ارتباطند، بازنمایی می‌شود.

-
1. topic modeling
 2. Papadimitriou et al.
 3. text mining
 4. cluster
 5. Hofmann, T.
 6. Blei et al.
 7. Latent Dirichlet Allocation
 8. Dirichlet prior distribution

در شکل (۱) مدل «تخصیص دریشله پنهان» نشان داده شده است که متشکل از دو ماتریس θ و φ است. در این مدل، φ ماتریسی است که توزیع موضوع T بر روی واژه‌های W را، براساس توزیع اولیه دریشله با پارامتر β بیان می‌دارد. θ ماتریسی است که توزیع متن d بر روی موضوع‌های T را براساس توزیع اولیه دریشله با پارامتر α بیان می‌کند. برای زایش هر واحد واژگانی w در متن d ، یک موضوع z از توزیع موضوعی مربوط به متن θ_d به دست می‌آید؛ درحالی‌که خود آن واژه w از توزیع واژه‌های موضوع انتخاب شده φ_z به دست می‌آید.

شکل ۱. نمایش تصویری مدل «تخصیص دریشله پنهان»



برای استخراج موضوعات با مدل «تخصیص دریشله پنهان»، نیاز است دو توزیع θ و φ تخمین زده شود تا اطلاعات درباره توزیع متن نسبت به موضوعات و موضوعات نسبت به متن به دست آید. برای تخمین این دو، الگوریتم‌های مختلفی پیشنهاد شده است، مانند انتشار انتظار (Minka & Lafferty, 2002)، استنباط تغییرات (Blei et al., 2003) و نمونه‌گیری گیبس^۱ (Griffiths & Steyvers, 2004). از میان این الگوریتم‌ها، نمونه‌گیری گیبس، به عنوان یک رویکرد ساده و مؤثر، برای این هدف مورد استفاده قرار می‌گیرد.

1. Gibbs sampling

در نمونه گیری گیبس، احتمال انتخاب یک موضوع برای یک واژه در یک متن به واژه قبلی و دو واژه قبلی در بافت و موضوعاتی که به آن موضوعات تخصیص داده می شود مشروط شده است که با استفاده از تساوی (۱) محاسبه می شود:

$$P(z_i = t | w_i = w, z_{-i}, w_{-i}) = \frac{N_{wt,-i}^{WT} + \beta}{\sum_{w'} N_{w't,-i}^{WT} + W\beta} \times \frac{N_{td,-i}^{TD} + \alpha}{\sum_{t'} N_{t'd,-i}^{TD} + T\alpha} \quad (1)$$

در این تساوی، $w_i = w$ نشان می دهد که i امین واژه در متن، واژه w است و $z_i = t$ نشان می دهد که واژه w به موضوع t تخصیص داده شده است. w_{-i} و z_{-i} بیانگر تمام واژه ها و تمام موضوعات تخصیص داده شده به جز واژه i امین است. تعداد دفعاتی که واژه w به موضوع t تخصیص داده شده است، به جز در نظر گرفتن آمار واژه کنونی. $N_{wt,-i}^{WT}$ تعداد دفعاتی است که موضوع t به متن d تخصیص داده شده است به جز موضوع کنونی. با استفاده از نمونه گیری گیبس، برای هر یک از نمونه های این مدل، θ و ϕ در تساوی (۲) و (۳) محاسبه می شود:

$$\phi_{wt} = \frac{N_{wt}^{WT} + \beta}{\sum_{w'} N_{w't}^{WT} + W\beta} \quad (2)$$

$$\theta_{td} = \frac{N_{td}^{TD} + \alpha}{\sum_{t'} N_{t'd}^{TD} + T\alpha} \quad (3)$$

که در این تساوی، ϕ_{wt} احتمال کاربرد واژه w در موضوع t است و θ_{td} احتمال کاربرد موضوع t در متن d است.

در این پژوهش، تلاش می شود ساختار معنایی نهفته در متن، با استفاده از الگوریتم «مدل سازی موضوع»، به دست آید و پس از این تحلیل اولیه خودکار، به فراتحلیل آن توسط انسان خبره پرداخته شود.

۲-۲. مروری بر مطالعات گذشته

در پژوهش‌های متنوعی، از مدل‌سازی موضوع برای کاربردهای مختلفی استفاده شده است که بررسی همگی آنها در این مقاله ممکن نیست. آنچه به اختصار توضیح داده می‌شود متمرکز بر استفاده از مدل‌سازی موضوع در حوزه بیماری کروناست. دونگ و همکاران^۱ (۲۰۲۰) به تحلیل بیش از ۳۵ هزار مقاله علمی در حوزه کرونا پرداخته‌اند تا بتوانند موضوعات داغ مرتبط با کرونا را بیابند. آنها در این پژوهش، با استفاده از مدل‌سازی موضوع، ۸ موضوع را به‌عنوان هدف تعریف کرده‌اند. سپس، به تحلیل رابطه معنایی بین موضوعات پرداخته‌اند و توزیع موضوعات بین کووید ۱۹ و سایر بیماری‌های خانواده کرونا، مانند سارس، را بررسی کرده‌اند.

سونبهدرا و همکاران^۲ (۲۰۲۰) در پژوهش خود تلاش کرده‌اند با استفاده از الگوریتم‌های خوشه‌بندی پارامتری مانند k-means و غیرپارامتری مانند DBSCAN و HAC، موضوعات داغ را از مقالات علمی مربوط به بیماری‌های سارس، مرس و کرونا استخراج کنند. در این پژوهش، پس از خوشه‌بندی داده‌ها، از دسته‌بند^۳ برای برچسب‌گذاری سایر مقالات علمی به یک برچسب استفاده کرده‌اند. داده‌ای که در این پژوهش مورد استفاده قرار گرفته است دادگان کورد-۱۹^۴ است. این دادگان حاوی ۴۵ هزار مقاله درباره این سه بیماری است که برای ۳۳ هزار مورد، مقاله کامل موجود است^۵. برای آموزش خوشه‌بند، به بازنمایی برداری داده نیاز است که برای این هدف، بردار متن مقالات موردنظر، به وسیله الگوریتم داک ۲ و ک^۶ تهیه شده و برای تشابه‌یابی، از فاصله کسینوسی استفاده شده است.

عبدالرزاق و همکاران^۷ (۲۰۲۰) به تحلیل موضوعات مهم رسانه اجتماعی تویتر درباره بیماری کرونا پرداخته‌اند. داده‌ای که برای این پژوهش استفاده شده است نظرات افراد در رسانه اجتماعی تویتر، به زبان انگلیسی، در بازه زمانی ۲ فوریه تا ۱۵ مارس ۲۰۲۰ است.

1. Dong et al.

2. Sonbhadra et al.

3. classifier

4. CORD-19 (COVID-19 Open Research Dataset)

5. <https://pages.semanticscholar.org/corona-virusresearch>

6. doc2vec

7. Abd-Alrazaq et al.

در این بازه زمانی، حدود ۲/۵ میلیون توییت جمع آوری شده است که حجم زیادی از آنها بازنشر بوده و فقط حاوی حدود ۱۶۷ هزار توییت یکتا^۱ است. آنها علاوه بر استخراج موضوعات از توییت‌ها با کمک مدل‌سازی موضوع، به تحلیل احساسات نظرات و تحلیل‌های آماری مانند متوسط بازنشر نظرات، علاقه‌مندی^۲ نظرات و دنباله‌رو^۳ موضوعات نیز پرداخته‌اند. در این پژوهش تعداد ۱۲ موضوع برای خوشه‌بندی داده‌ها در نظر گرفته شده است. مهم‌ترین موضوع در بین کاربران که بیشترین علاقه‌مندی را به خود اختصاص داده است «ضررهای اقتصادی» بوده و کمترین علاقه‌مندی مربوط به موضوع «محدودیت سفر و هشدارها» بوده است.

اوردون و همکاران^۴ (۲۰۲۰) از سه روش «همانندیابی الگو»^۵، «تخمین و برآورد زیاد یکنواخت»^۶ و مدل‌سازی موضوع، برای استخراج موضوعات استفاده کرده‌اند. سپس، به تحلیل آماری اطلاعات موجود در نظرات، مانند علاقه‌مندی به موضوعات، ارتباط بین کاربرها و سرعت بازنشر نظرات در قالب یک نقشه تصویری در طی زمان، پرداخته‌اند. در این پژوهش، نظرات انگلیسی کاربران توییت، در بازه زمانی ۲۴ مارس تا ۹ آوریل ۲۰۲۰، جمع آوری شده که مجموعاً حاوی بیش از ۲۳ میلیون توییت و بیش از ۵ میلیون توییت یکتاست. داده‌های این پژوهش به ۲۰ موضوع خوشه‌بندی شده است. یکی از یافته‌ها این بوده است که با افزایش زمان بازنشر توییت، چگالی ارتباط‌ها نیز افزایش یافته و کاربران مشخصی کانون توجه بازنشر توییت‌های مربوط به بیماری کرونا را تحت تأثیر قرار می‌دهند.

کایلا و پراساد^۷ (۲۰۲۰)، از نگاه «جریان اطلاعات»^۸، به بررسی نظرات توییت انگلیسی و راستی‌آزمایی نظرات با استفاده از مدل‌سازی و تحلیل احساسات موضوع پرداخته‌اند. برای این هدف، ۱۸ هزار توییت بدون هرگونه بازنشر، به صورت تصادفی، انتخاب شده است. سپس، ۱۰ موضوع از این داده استخراج شده و به صورت خودکار، ۸ نوع برجسب

-
1. unique
 2. like
 3. follower
 4. Ordun et al.
 5. pattern matching
 6. uniform manifold approximation and projection
 7. Kaila, R. P., & Prasad, A. V. K.
 8. information flow

تحلیل احساسات برای این حجم داده تخصیص داده شده است. موضوعات استخراج شده منعکس‌کننده واقعیات از جامعه در بازه زمانی توییت‌های جمع‌آوری شده بوده است. «ترس» و «اعتماد به مسئولان» بالاترین بسامد تحلیل احساسات را داشته است. نتیجه‌گیری این مقاله این بوده است که نظرات ارائه‌شده در توییت‌ها اهمیت دارد و مسئولان می‌توانند از تحلیل نظرات برای تصمیم‌گیری استفاده کنند.

۳. روش پژوهش

۳-۱. داده

این پژوهش بر تحلیل محتوایی نظرات کاربران در دو رسانه اجتماعی توییت و اینستاگرام متمرکز است. از آنجا که تعداد نظراتی که روزانه از کاربران اینترنتی در این دو رسانه انتشار می‌یابد بسیار زیاد است، گردآوری این نوع داده به صورت دستی ممکن نیست. به همین دلیل، برای افزایش دقت و کاهش هزینه و تلاش نیروی انسانی در جمع‌آوری این نوع داده خاص، از روش «خزش در وب»^۱ استفاده می‌شود که به اختصار، خزش^۲ نامیده می‌شود. در فرایند خزش، ابتدا به صورت نظام‌مند، گستره جهانی اینترنت پیمایش می‌شود و سپس، با یافتن اطلاعات موردنظر از وب، داده خروجی در قالب ساختار مشخص موردنظر ذخیره می‌شود. بایلی و همکاران^۳ (۲۰۰۳: ۴) دو شیوه کلی برای خزش معرفی کرده‌اند. یکی «خزش پیوندی»^۴ است که با پیمایش گراف پیوند صفحات وب و شروع از مجموعه‌ای از نقاط آغاز، کار شناسایی و اتصال صفحات وب که معمولاً برای فهرست‌بندی استفاده می‌شود انجام می‌پذیرد. نوع دیگر «خزش محتوایی»^۵ است که با کمک خزش پیوندی، براساس محتوای اسنادی که صفحات وب را تشکیل می‌دهند، انجام می‌گیرد. چایترا و همکاران^۶ (۲۰۱۸) خزش را به سه دسته تقسیم کرده‌اند: الف) خزشگر متمرکز^۷ که به صورت انتخابی، صفحات وب مرتبط با موضوع موردنظر را

-
1. Web crawling
 2. crawl
 3. Bailey et al.
 4. link crawling
 5. content crawling
 6. Chaitra et al.
 7. focused Web crawler

می‌کاود؛ ب) خزشگر افزایشی^۱ که با بازدید مکرر صفحات وب، اطلاعات به‌روز شده در آن صفحات را پیدا و به‌روزرسانی می‌کند؛ ج) خزشگر توزیعی^۲ که در این شیوه، فرایند خزش توسط یک سرور مرکزی به سایر خزشگرها محول می‌شود.

۲-۳. مدل‌سازی موضوع در داده‌های رسانه‌های اجتماعی

برای پردازش داده و استخراج موضوعات از داده‌ها به ابزار پردازشی نیاز است. برای این هدف، از ابزار مالت^۳ (McCallum, 2002) استفاده می‌کنیم. مالت یک بسته نرم‌افزاری است که به زبان برنامه‌نویسی جاوا^۴ نوشته شده و برای پردازش‌های آماری زبان طبیعی، مانند دسته‌بندی^۵، خوشه‌بندی، مدل‌سازی موضوع و «استخراج اطلاعات»^۶، به کار می‌رود. در این ابزار، روش نمونه‌گیری گیبس در الگوریتم تخصیص دریشله پنهان، که در مدل‌سازی موضوع استفاده می‌شود، پیاده‌سازی شده است. در این پژوهش، از ابزار مالت برای استخراج موضوع‌ها از داده‌های فارسی حاصل از رسانه‌های اجتماعی استفاده می‌کنیم.

خروجی الگوریتم مدل‌سازی موضوع دو ماتریس φ و θ است. در ماتریس φ ، سطرها معرف موضوع‌ها و ستون‌ها نشان‌دهنده واژگان است. هر سطر حاوی ۲۰ واژه‌ای است که بالاترین توزیع احتمالاتی را نسبت به موضوع انتزاعی مرتبط با آن سطر دارد؛ در ماتریس θ ، سطرها معرف نظرات شبکه اجتماعی و ستون‌ها معرف موضوع‌هاست و هر سطر حاوی توزیع احتمالاتی موضوعات انتزاعی نظرات مربوط است. از آنجا که الگوریتم مدل‌سازی موضوع جزء الگوریتم‌های پارامتری محسوب می‌شود، نیاز است تعداد موضوعات از ابتدای فرایند خوشه‌بندی مشخص شود. خروجی به‌دست‌آمده از این ابزار یک «خوشه‌بندی نرم»^۷ است؛ به این مفهوم که توزیع احتمالاتی هر نظر به تعداد موضوعات انتزاعی مشخص شده محاسبه می‌شود.

-
1. incremental Web crawler
 2. distributional Web crawler
 3. Mallet
 4. Java
 5. classification
 6. information extraction
 7. soft clustering

۳-۳. کاربرد هشتگ در داده‌های رسانه‌های اجتماعی

می‌دانیم نظرات منتشر شده در رسانه‌های اجتماعی، علاوه بر خود نظر، حاوی یک برچسب محتوایی است که با علامت # شروع و به اصطلاح، «هشتگ»^۱ نامیده می‌شود. معمولاً بین محتوای متن و هشتگ‌های استفاده شده ارتباط محتوایی وجود دارد. از آنجا که نظرات در رسانه‌های اجتماعی کوتاه است، بافت کامل برای جملات نوشته شده وجود ندارد تا یک متن منسجم شکل بگیرد. این هشتگ‌ها کار فراداده^۲ را انجام می‌دهند و در تکمیل اطلاعات زبانی اثر دارند. از ترکیب نظرات و این هشتگ‌ها تا حدودی انسجام درونی متن و محتوا شکل می‌گیرد. تسور و راپوپورت^۳ (۲۰۱۲) از بررسی محتوایی هشتگ‌ها برای پیش‌بینی توزیع نظرات در جامعه استفاده کرده‌اند. گوراجالا و متیوس^۴ (۲۰۱۸) تحلیل هشتگ‌های تویتر را برای فهمیدن عکس‌العمل جامعه به کیفیت هوا به کار برده‌اند. در این پژوهش، از نظرات تویتر در یک بازه زمانی دوساله استفاده شده است تا همبستگی بسامد اعلام نظرات و کیفیت هوا، در سه شهر پاریس، لندن و دهلی‌نو، سنجیده شود.

با نگاهی به مجموعه نظرات رسانه‌های اجتماعی، می‌توان به این فرضیه رسید که بین نظرات و هشتگ‌ها نوعی همبستگی وجود دارد. یکی از روش‌های آماری برای محاسبه همبستگی، «ضریب همبستگی پیرسون»^۵ است. همبستگی یک شیوه آماری است که برای مقایسه دو متغیر کمی پیوسته استفاده می‌شود. در همبستگی، شدت پیوستگی دو متغیر سنجیده می‌شود. این ضریب عددی بین ۱ تا -۱ است. اگر ضریب همبستگی پیرسون عدد ۱ باشد، بیانگر رابطه مستقیم بین دو متغیر است؛ به این مفهوم که با هرگونه تغییر در یک متغیر، متغیر دیگر نیز تغییر می‌کند. چنانچه این ضریب عدد ۱- باشد، رابطه معکوس بین دو متغیر وجود دارد که با افزایش یک متغیر، متغیر دیگر کاهش می‌یابد. اگر این ضریب صفر باشد بیانگر عدم وجود رابطه خطی بین این دو متغیر است (Boddy & Smith, 2009: 92-94). از تساوی (۴) برای محاسبه ضریب همبستگی پیرسون استفاده می‌شود:

-
1. hashtag
 2. meta-data
 3. Tsur, O., & Rappoport, A.
 4. Gurajala, S., & Matthews, J. N.
 5. Pearson correlation coefficient

$$r = \frac{N \sum xy - (\sum x)(\sum y)}{\sqrt{[N \sum x^2 - (\sum x)^2][N \sum y^2 - (\sum y)^2]}} \quad (۴)$$

که در این تساوی، N تعداد جفت‌های موردنظر برای مقایسه است، $\sum xy$ مجموع تعداد توالی جفت‌های موردنظر است، $\sum x$ مجموع تعداد متغیر X و $\sum y$ مجموع تعداد متغیر Y است، $\sum x^2$ مجموع مجذور تعداد متغیر X و $\sum y^2$ مجموع مجذور تعداد متغیر Y است. در این پژوهش، می‌کوشیم بر زبان فارسی متمرکز شویم و ضمن استخراج موضوع‌ها از میان نظرات و پیام‌های مرتبط با کرونا به زبان فارسی در رسانه‌های اجتماعی، به بررسی همبستگی رابطه موضوع‌ها و هشتک‌ها نیز پردازیم. برای تحلیل هشتک‌ها، از ضریب همبستگی پیرسون میان موضوع‌ها و هشتک‌ها استفاده می‌کنیم. در این راستا، هر موضوع با برداری به ابعاد تعداد نظرات نمایش داده می‌شود و همچنین، هر هشتک نیز با برداری با همین ابعاد نشان داده می‌شود. برای این منظور، تعداد ۲۴ موضوع و ۲۰۱ هشتک در نظر گرفته شده است. معیار انتخاب تعداد هشتک‌ها بسامدشان است؛ به این صورت که هشتک‌های دارای بسامد بالای ۱۰۰ برای این منظور انتخاب شده‌اند. با داشتن این بردار برای هریک از موضوعات و هشتک‌ها، ضریب همبستگی پیرسون استخراج می‌شود و از این میان، زوج موضوع-هشتک‌هایی که همبستگی بالا دارد گزارش و تحلیل می‌شود.

۴. گردآوری داده‌های پژوهش

برای انجام هر پژوهش زبان‌شناختی، به پیکره زبانی نیاز است. قیومی (۱۳۹۳) فهرستی از پیکره‌های زبانی موجود برای زبان فارسی را معرفی کرده و به مشکل واژگان خارج و چالش سامانه‌های پردازش زبان پرداخته است. این مشکل، در شرایطی که واژه‌ها و اصطلاحات جدید مربوط به کرونا وارد زبان شده، حادث‌تر است؛ چراکه معمولاً سامانه‌های پردازش زبان با پیکره‌های موجود آموزش دیده‌اند و به دلیل عدم به‌روزر بودن پیکره‌ها، واژه‌ها و اصطلاحات جدید در این پیکره‌های موجود یافت نمی‌شود. از آنجا که مقاله حاضر بر بیماری کرونا متمرکز است، پیکره‌های عمومی معرفی شده کارکرد لازم را ندارد؛ بنابراین، نیاز است یک پیکره تخصصی در این حوزه تهیه شود.

این پیکره زبانی از طریق خزش در نظرات و پیام‌های کاربران رسانه‌های اجتماعی اینستاگرام و تلگرام تهیه شده است. شیوه‌ای که برای خزش این دو رسانه اجتماعی استفاده شده از نوع خزش افزایشی است. داده‌های به‌دست آمده از طریق خزش، بدون هرگونه ساختارمندی و به‌منظور داشتن قابلیت کاربرد در یک پژوهش، نیاز است ضمن ساختارمندی، نوفه داده کاهش یابد. از این‌رو، در مرحله پیش‌پردازش، چندین اقدام انجام شده است که در ادامه، مراحل آن توضیح داده می‌شود. لازم به ذکر است از مجموعه‌نظراتی که در رسانه‌های اجتماعی منتشر می‌شود، از نظر زبانی و محتوایی، بسیار متنوع است. با توجه به اهداف این پژوهش، نظرات و پیام‌های خزش شده مرتبط با موضوع کرونا که به زبان فارسی بوده‌اند، به‌عنوان پیکره زبانی مورد نیاز در این پژوهش، استفاده می‌شود. شایان ذکر است وجود واژه‌هایی که غلط‌املائی داشته باشند و یا به گونه زبانی محاوره‌ای یا گویشی نوشته شده باشند دور از انتظار نیست و از این نظر، تغییری در داده اصلی ایجاد نکرده‌ایم تا داده طبیعی به نظر برسد.

۴-۱. مراحل پیش‌پردازش

۴-۱-۱. ساختارمندی داده

به‌منظور ساختارمندی داده زبانی، قیومی (۱۳۹۸) چندین ساختار را معرفی کرده و با توجه به نیاز، کار ساختارمندی داده را انجام داده است (قیومی، ۱۳۹۸). داده گردآوری شده از نظرات کاربران در شبکه‌های اجتماعی حاوی ۶ ستون است. ستون اول حاوی نمایه پیام است؛ ستون دوم و سوم حاوی منبع و تاریخ انتشار نظر است؛ ستون چهارم حاوی اطلاعات درمورد بازنشر نظر است و در ستون پنجم، نام حساب کاربری فرد در رسانه اجتماعی که نظری را بازنشر کرده ذکر شده است؛ و در نهایت، ستون ششم حاوی محتوای نظر است.

۴-۱-۲. بهنجارسازی

نظرات خزش شده از رسانه‌های اجتماعی چند مشکل اساسی دارد که کاربردی شدن آن را با چالش مواجه می‌کند. اول اینکه تنوع در تلفن‌های هوشمند سبب شده است کدگذاری‌های متون یکدست نباشد؛ بنابراین، نیاز است نظرات، از نظر کد، یکدست‌سازی شوند و همگی به کد استاندارد UTF8 تبدیل گردند. ناگفته نماند این نوع یکدست‌سازی

موجب می شود تنوع در حروف، مانند «ک»، «ی» یا «ے» عربی، از بین برود. مشکل دیگر این نوع داده استفاده از شکلک^۱ در متن نظرات و گاهی به عنوان خود نظر است. در این پژوهش، به صورت دستی، کد هریک از این شکلک ها شناسایی شد و از محتوای نظرات حذف شد.

۳-۱-۴. واحدسازی

رعایت فاصله گذاری بیرونی بین واژه ها و درج نیم فاصله درونی در یک واژه یکی از چالش های بسیار مهم در تشخیص واژه است که بر کیفیت شیوه های پردازش زبان طبیعی مبتنی بر آمار تأثیر می گذارد. برای مثال، عدم رعایت فاصله گذاری صحیح سبب می شود رایانه زنجیره «ویابهتر است» را به عنوان یک واژه تلقی کند، زیرا برای رایانه، فاصله بین واژه ها به عنوان معیار تشخیص یک واژه تعریف شده است. عکس این موضوع نیز صادق است، مانند «دانش آموز» که به دلیل وجود فاصله کامل به جای نیم فاصله، به عنوان دو واژه شمارش می شود. برای کاهش این مشکلات در انجام این پژوهش، از الگوریتم معرفی شده توسط قیومی (۱۳۹۷) برای واحدسازی^۲ متن استفاده کرده ایم. این الگوریتم سه مرحله ای است. در مرحله اول، کار تفکیک چندواحدی ها به واژه های مستقل یا ترکیب وندها با پایه ها در متن صورت می پذیرد. در مرحله دوم، به واحدسازی فعل پرداخته می شود و صورت های صرفی فعل با یکدیگر ترکیب می شوند. در مرحله سوم، امکان ساخت واحدهای واژگانی از زنجیره های چندواحدی میسر می شود؛ این واحدها در مرحله اول از یکدیگر تفکیک شده اند و ترکیب آنها با یکدیگر می تواند به ساخت یک واحد واژگانی منجر شود. نتایج عملی گزارش شده که روی واحدسازی پیکره آزمونی، با نمونه گیری تصادفی از پایگاه داده زبان فارسی (عاصی، ۱۳۸۴)، تهیه شده حاکی از صحت^۳ ۹۷/۸۰ درصد در واحدسازی و میزان ۰/۰۲ درصد تولید خطا توسط الگوریتم معرفی شده است.

1. emoji
2. tokenization
3. accuracy

۲-۴. پیکره نظرات کرونا

برای تهیه «پیکره نظرات کرونا» از نظرات کاربران فارسی‌زبان دو رسانه اجتماعی اینستاگرام و توییتر، پس از جمع‌آوری تمامی نظرات در بازه زمانی ۳۰ بهمن ۱۳۹۸ تا ۳۱ خرداد ۱۳۹۹، واژه «کرونا» به زبان فارسی در این مجموعه داده جستجو شد. نظراتی که حاوی این واژه بود استخراج شد و با ساختارمندسازی و ذخیره‌سازی آن، به‌عنوان پیکره زبانی هدف در این پژوهش مورد استفاده قرار گرفت. در جدول (۱)، اطلاعات آماری استخراج‌شده از این مجموعه داده گزارش شده است.

جدول ۱. اطلاعات آماری استخراج‌شده از «پیکره نظرات کرونا»

| مقدار | اطلاعات آماری |
|-----------|----------------------|
| ۰/۵۸ | نسبت نظرات بازنشرشده |
| ۱۳۴,۹۳۵ | نظرات بازنشرشده |
| ۰/۱۹ | تنوع هشتک‌ها |
| ۵۴,۳۸۳ | هشتک‌های بدون تکرار |
| ۲۹۳,۱۵۲ | هشتک‌های با تکرار |
| ۰/۳۲ | تنوع واژگانی |
| ۲۱۶,۱۸۲ | واژه‌های بدون تکرار |
| ۶,۶۱۶,۲۴۴ | واژه‌های با تکرار |
| ۲۳۳,۵۵۷ | تعداد نظرات |

همان‌گونه که در جدول (۱) مشاهده می‌شود، حجم زیادی از نظرات در حقیقت بازنشر نظرات دیگران است. این ویژگی موجب کاهش تنوع واژگانی شده است؛ بنابراین، نیاز است نظرات بازنشرشده حذف شوند. در جدول (۲)، اطلاعات آماری «پیکره نظرات کرونا»، بدون در نظر گرفتن نظرات بازنشرشده، گزارش شده است. با حذف نظرات بازنشرشده، حدود نیمی از نظرات کنار گذاشته شد و در این پژوهش مورد استفاده قرار نگرفت. همچنین، حذف این نظرات به افزایش تنوع واژگانی و تنوع هشتک‌ها منجر شد.

جدول ۲. اطلاعات آماری استخراج شده از «پیکره نظرات کرونا» بدون نظرات بازنشر شده

| تعداد نظرات | واژه‌های با تکرار | واژه‌های بدون تکرار | تنوع واژگانی | هشتگ‌های با تکرار | هشتگ‌های بدون تکرار | تنوع هشتگ‌ها |
|-------------|-------------------|---------------------|--------------|-------------------|---------------------|--------------|
| ۹۹,۸۱۳ | ۳,۳۸۴,۹۳۸ | ۱۷۹,۰۰۴ | ۰/۰۵۳ | ۱۹۰,۵۱۱ | ۳۰,۵۶۱ | ۰/۱۶ |

۵. نتایج به دست آمده

۵-۱. تنظیمات آزمایش‌ها

هدف این پژوهش، تحلیل محتوایی نظرات کاربران رسانه‌های اجتماعی و یافتن اهم موضوعاتی است که در نظرات آنها انعکاس داده شده است. علاوه بر یافتن موضوعات، وجود رابطه همبستگی بین موضوع نظرات و هشتگ‌های تخصیص داده شده نیز بررسی خواهد شد. برای ارائه آمار نظرات از تحلیل موضوعات این پژوهش، که در بخش ۵-۲ توضیح داده می‌شود، نیاز است «خوشه‌بندی نرم» ارائه شده توسط الگوریتم مدل‌سازی موضوعی را به «خوشه‌بندی سخت»^۱ تبدیل کنیم؛ به این مفهوم که برای هر متن، فقط یک موضوع انتزاعی تعیین شود. معیار انتخاب موضوع انتزاعی کاندید، موضوعی است که بالاترین احتمال را برای آن متن دارد. در مرحله بعد، نیاز است موضوع‌های انتزاعی، که به واسطه مجموعه ۲۰ واژه‌ای در فایل φ الگوریتم مدل‌سازی موضوع مشخص و معرفی شده‌اند، توسط یک انسان خبره، به صورت دستی، فراتحلیل شود و هویت آن موضوع، با یک برچسب محتوایی، مشخص گردد. شایان ذکر است در بخش ۵-۳، جایی که همبستگی موضوعات و هشتگ‌ها ارائه می‌شود، خوشه‌بندی نرم در بررسی انجام شده مورد استفاده قرار می‌گیرد.

۵-۲. نتایج تحلیل موضوعی

در این پژوهش، داده ورودی ابزار مالت، پیکره نظرات کروناست. تعداد موضوعاتی که نیاز است برای ابزار مالت تنظیم شود، به صورت تجربی، تعداد ۲۴ موضوع براساس پیکره موجود تعیین شد تا موضوع‌های انتزاعی استخراج شوند. سپس، براساس ۲۰ واژه هر موضوع، به صورت دستی، برچسب هر موضوع مشخص شد. در جدول (۳)، مجموعه برچسب موضوع‌ها، توزیع آماری هر موضوع و ۵ واژه‌ای که بیانگر آن موضوع است گزارش شده‌اند.

1. hard clustering

جدول ۳. موضوعات استخراج شده از پیکره کرونا

| ردیف | بسامد نظرات | درصد نسبی نظرات | برچسب موضوعی | واژه‌های موضوع |
|------|-------------|-----------------|------------------------------|---|
| ۱ | ۱۴۳۳۴ | ۱۴/۳۶ | سیاسی داخلی | حکومت، ملت، انتخابات، اتحاد، دشمن |
| ۲ | ۱۰۱۷۶ | ۱۰/۲۰ | اجتماعی | خودکشی، زندگی، عادی، مسافرت، رعایت |
| ۳ | ۶۵۰۱ | ۶/۵۱ | خانواده | خونه، کار، مامانم، بیدار، دوست، زنگ |
| ۴ | ۵۹۷۴ | ۵/۹۹ | مدیریت بحران | ستاد، وضعیت، مدیریت، بحران، شرایط |
| ۵ | ۵۹۴۹ | ۵/۹۶ | مذهبی | دعا، حرم، نماز، دین، اسلام |
| ۶ | ۵۱۰۱ | ۵/۱۱ | اخبار داخلی کرونا | مبتلایان، تعداد، ابتلا، افزایش، قربانیان |
| ۷ | ۴۸۱۳ | ۴/۸۲ | توهین آمیز | سگ، خر، الاغ، نفهم، بی شعور |
| ۸ | ۳۸۳۷ | ۳/۸۴ | اخبار بین الملل کرونا | آمریکا، اروپا، تجهیزات، غرب، تحریم |
| ۹ | ۳۷۳۰ | ۳/۷۴ | فاصله گذاری اجتماعی | قرنطینه، رعایت، جلوگیری، انتقال، قطع |
| ۱۰ | ۳۷۲۵ | ۳/۷۳ | بهداشتی | ماسک، ویروس، الکل، ضد عفونی، دستکش |
| ۱۱ | ۳۶۹۸ | ۳/۷۰ | اقتصادی | دلار، پول، قیمت، نفت، بازار |
| ۱۲ | ۳۴۷۲ | ۳/۴۸ | سیاسی خارجی | ترامپ، جهان، جنگ، رژیم، حمله |
| ۱۳ | ۳۴۱۱ | ۳/۴۲ | نوروز | سال، عید، مبارک، تبریک، شادی |
| ۱۴ | ۳۰۹۱ | ۳/۱۰ | تشخیص کرونا | آزمایش، مشکوک، مثبت، بستری، علائم |
| ۱۵ | ۳۰۴۵ | ۳/۰۵ | درمان | واکسن، داروی، کشف، تست، کیت |
| ۱۶ | ۲۸۰۰ | ۲/۸۱ | پزشکی | بیمارستان، کادر، مبارزه، پزشکان، سلامت |
| ۱۷ | ۲۷۹۶ | ۲/۸۰ | داستانک | حرف، دروغ، توییت، جواب، گوش |
| ۱۸ | ۲۵۳۸ | ۲/۵۴ | افغانستان | هرات، ولایت، صحت، کابل، #هرات_در_بحران |
| ۱۹ | ۲۲۶۶ | ۲/۲۷ | پوش و هشتگ های کرونا | #کرونا_را_شکست_می_دهیم، #ویروس، #کرونادرایران، #در_خانه_بمانیم، #قرنطینه |
| ۲۰ | ۲۰۴۳ | ۲/۰۵ | سرگرمی | فیلم، برنامه، نمایش، تیم، #موسیقی |
| ۲۱ | ۱۸۹۸ | ۱/۹۰ | کرونا و زندان | زندان، زندانیان، فرار، آزاد، مرخصی |
| ۲۲ | ۱۷۵۸ | ۱/۷۶ | رسانه | سایت، پست، آنلاین، ارسال، پیج |
| ۲۳ | ۱۶۴۲ | ۱/۶۵ | رویدادهای غم انگیز سال ۹۸ | دی، شکست، هواپیما، بنزین، زلزله |
| ۲۴ | ۱۲۱۵ | ۱/۲۲ | تبلیغات | #بوتیک، #لباس، #تحویل، #اسپرت، دایرکت |

در ادامه، نمونه‌هایی از پنج موضوع پربسامد ارائه می‌شود تا ضمن نمایش عملکرد الگوریتم مدل‌سازی موضوع، امکان فراتحلیل داده‌ها فراهم شود.

۵-۲-۱. موضوع سیاسی-داخلی

الف) اعمال سیاست‌های جهانی کنترل جمعیت در ایران بارها مورد تذکر رهبری قرار گرفته است. یکی از دلایل تلفات بالای #کرونا در اتحادیه اروپا افزایش کهولت و کاهش نسبت جوانان در هرم سنی است. دولت‌ها باید به این چالش جهانی #کنترل_جمعیت توجه کنند.

ب) روش جدیدی که دشمن علیه ایران در جریان معضل ویروس کرونا پیش گرفته فشار بر روی مردم و انتساب آن به رهبری است. برخی اصلاح‌طلبان نیز دشمن را همراهی می‌کنند. خوب است آگاهی‌های لازم در این مورد داده شود و راه‌های مقابله با این حربه باید طراحی و اجرایی بشوند.

ج) افراد و شبکه‌هایی که معتقدند تحریم‌ها فقط حکومت ایران را در بر گرفته نه مردم را و هدف براندازی نظام این کشور نیست، نمونه تحریمی ایران را در مورد ایتالیا، که ششمین اقتصاد برتر دنیاست اعمال کنند! ایتالیا در بحران کرونا مجبور به دزدیدن ماسک شده، تحریم مالی، دارو و... هم بماند.

د) مبارزه با کرونا اتحاد مسئولین سیاسی و مردم را می‌خواهد. باید که #متحدانه در این راستا حرکت کنیم که خدای نخواستہ از فاجعه انسانی #جلوگیری شود. همه می‌دانیم که #کرونا وضعیت سیاسی، اقتصادی و اجتماعی را در جهان تغییر می‌دهد.

۵-۲-۲. موضوع اجتماعی

الف) پیام جکی چان به مردم ایران: قوی باش چین! قوی باش ایران! جکی چان بازیگر مطرح و بین‌المللی اهل چین، همراه با جمعی از بازیگران چینی، با ارسال پیامی، از مردم ایران و چین خواست قوی بمانند. این پیام به پاس کمک‌های بشردوستانه و پیام‌های روحیه‌بخش ایران برای کشور چین در هنگام مبارزه با #کرونا ارسال شده، چون آنها معتقدند دوست واقعی هر کشوری در هنگام سختی‌ها مشخص خواهد شد.

ب) بعد از کرونا، چند روز همه جوگیرن، به این صورت که از کافه رفتن و قدم زدن تو خیابون و... لذت می‌برن و قدر همدیگه رو می‌دونن ولی بعدش بازم همه چیز برمی‌گرده به روال عادی و روزمرگی همه مون رو دیوونه می‌کنه.

ج) اولین خودکشی به دلیل کرونا؛ وزیر دارایی ایالت آلمان خودکشی کرد، زیرا نگران بود نتواند با تأثیرات منفی کرونا مقابله و انتظارات مردم در این باره را برآورده کند. او خود را در ریل قطار انداخت و قبل از آن، با دل‌نوشته‌ای دلیل خودکشی را اعلام کرده بود.

د) من می‌تونم بگم بارها به این فکر می‌کردم که این روش خودکشی آسون‌ترین و اگر روزی بخوام انجام بدم این کارو می‌کنم. بعد این ویدیو و اون صدای تهش کاملاً منصرف شدم - یه میلیاردر ایتالیایی چون کل خانواده‌ش بر اثر کرونا مردن، خودشو می‌کشه.

۳-۲-۵. موضوع خانواده

الف) حتی اگه زنگ بزنی که لطفاً بیا قرارداد ببندیم برای ادامه همکاری مون نمی‌رم، تا آخر کرونا هم خونه می‌مونم. از جون مامان بابام مهم‌تر نیست.

ب) حس می‌کنم در دنیای پساکرونا! آدم‌ها و فامیل از هم دورتر می‌شن. دیگه مثل قبل مسافرت خونه هم نمی‌رن و شاید مهمانی‌ها یه دورهمی ساده‌ی سه‌ساعته تو رستوران و کافه بشه.

ج) همسایه روبه‌روی مون خانوادگی کرونا گرفتن. مامانم دم‌به‌ساعت زنگ می‌زنه می‌گه پنجره خونه تو ببند کرونا میاد تو.

د) نمی‌دونم روز چندم قرنطینه است. اگر تقویم چک نکنم، حتی نمی‌دونم چندشنبه است. فقط اینکه دیگه دل و دماغ ندارم. اگر هم از این ور اون ور روحیه می‌گیرم، سه‌چهار ساعت بیشتر دووم نداره. دوباره رفتم تو مود تنهایی. همه ناتیفیکیشن‌های معاشرتی تعطیله. فقط اینجا میام یه چیزی شر می‌کنم و بعد پیش خودم می‌گم چه کار مسخره‌ای بود. فکر کنم اگر بر اثر کرونا نمیریم، افسردگی، دیرشن‌های شدید خواهیم گرفت. مخصوصاً برای ماهایی که تنهایییم، تنها زندگی می‌کنیم، یه کشور دیگه‌ایم. انقدر حساس شدم که با کوچک‌ترین انگولکی، کنترل خودم رو از دست می‌دم. تنها دلخوشی این روزام شده آشپزی. که اونم نمی‌دونم تا کی ادامه داشته باشه. اصلاً تا کی قراره تو خونه بشینیم، من

اصلاً آدم تنهایی و خونه‌نشینی نیستم و خودم می‌دونم دارم دپرشن می‌گیرم. این روند ادامه داشته باشه، نمی‌دونم تهش قراره با خودم چجوری رفتار کنم.

۴-۲-۵. موضوع مدیریت بحران

الف) همسفران عزیز؛ با توجه به نگرانی‌های موجود از شیوع کرونا، تور این هفته رو کنسل کردیم. با توجه به اخبار ضدونقیضی که بیرون اومده و آمار کشته‌شدگان که بیشتر شده، نسبت به کنسل کردن تورهای پیش رو حساسیت بیشتری به خرج می‌دیم و اگر لازم باشه، تا زمان بهبود شرایط، تمام تورها رو کنسل می‌کنیم. لطفاً مراقب خودتون باشید و توصیه‌های پزشکی رو جدی بگیرید.

ب) فردا سر کار رفتن ممنوع است. کیانوش جهانپور، سخنگوی وزارت بهداشت، امروز درباره‌ی آغاز به کار برخی مشاغل از روز شنبه (۱۶ فروردین) گفت تمامی مشاغل باید تابع مصوبات ستاد مقابله با کرونا باشن.

ج) اطلاع‌رسانی کرونا. وزیر بهداشت، در نامه‌ای، به رئیس‌جمهوری اعلام کرد که نامه‌ی وزارت صمت، مبنی بر بازگشایی کلیه‌ی مشاغل، با نقش، مسئولیت و جایگاه ستاد ملی مقابله با کرونا مغایرت دارد.

د) گویا نحسی و ویروس کرونا بیشتر دامن مدعیان توسعه‌یافتگی و پیشرفت را گرفته است. شیوع بیماری کرونا، با این سرعت وحشتناک، نشان از وضعیت اسفناک بهداشت در کشورهای به‌ظاهر تمیز دارد.

۵-۲-۵. موضوع مذهبی

الف) دعا کنیم برای دل‌های شکسته... برای کسانی که مشکل افتاده تو زندگی شون... برای صبر بازماندگانی که عزیزانشون رو از دست دادن... خدا یا از تو برای همه آرامش می‌خوام، دنیای خالی از کرونا ویروس قسمت کن.

ب) در دورانی که حرم‌ها و مساجد و نماز جمعه تعطیل است و مؤمنین اجتماعات دینی را به‌خاطر پیشگیری از شیوع بیماری کرونا لغو کرده‌اند، پلیس ۱۱۱ مرد و زن را در یک پارتنی مختلط شبانه در شهریار دستگیر کرد!

ج) #کرونا ثابت کرد کسانی که می‌گفتن وقت نداریم نماز قضا و روزه و خواندن قرآن و نهج‌البلاغه و کتاب آقای مطهری و... بخونم... بهانه بود، چون تو این یک ماه فقط لم دادن تو مجازی از ترس دارن پیام ارسال می‌کنند.

د) از بیانات این عالم دین؛ ویروس کرونا برای کفار عذاب است، برای فاسق تنبیه است و برای مسلمان رحمت!

همان‌طور که در جدول (۳) مشخص است، دو موضوع «سیاسی-داخلی» و «اجتماعی» محور حدود ۲۵٪ از نظرات در پیکره نظرات کرونای بوده است. موضوع «تبلیغات» کمترین حجم از نظرات (۱/۲٪ نظرات) را در پیکره مذکور به خود اختصاص داده است. از میان ۲۴ موضوعی که در مورد بیماری کرونا مطرح شده است، تعداد ۱۰ موضوع مستقیماً با خود ویروس کرونا و ماهیت همه‌گیری آن مربوط است: «مدیریت بحران»، «اخبار داخلی کرونا»، «اخبار بین‌الملل کرونا»، «فاصله‌گذاری اجتماعی»، «بهداشتی»، «تشخیص کرونا»، «درمان»، «پزشکی»، «پوشش و هشتگ‌های کرونایی» و «کرونا و زندان». این موضوعات بیش از ۳۵٪ از حجم نظرات مطرح شده در پیکره نظرات را به خود اختصاص داده است. می‌دانیم زبان فارسی بین کشورهای ایران، افغانستان و تاجیکستان مشترک است و ممکن است علاوه بر ایران، افرادی از دو کشور دیگر نیز نظرات خود را در رسانه‌های اجتماعی انعکاس دهند. با استخراج موضوعات، مشخص شد موضوع «افغانستان» ۲/۵۴٪ از نظرات را به خود اختصاص داده است که معرف نظرات به زبان فارسی در افغانستان در کنار نظرات به زبان فارسی در ایران است.

۳-۵. نتایج تحلیل هشتگ‌ها

در پیکره نظرات کرونای بررسی شده، ۱۹۰،۳۵۲ هشتگ در نظرات یافت شد که بدون در نظر گرفتن تکرار، تعداد ۳۰،۹۸۵ هشتگ باقی ماند. از این تعداد، ۱۹۰،۱۹۶ هشتگ در نظرات (به‌طور نسبی ۱۰/۹۸٪) فقط یک بار به کار رفته است و تعداد ۲۰۱ هشتگ بیش از ۱۰۰ بار تکرار شده بودند. در جدول (۴)، فهرست ۱۰ هشتگ پرکاربرد در نظرات گزارش شده است. ۸ هشتگ اول بر موضوع کرونا متمرکز است و خود هشتگ کرونا، یعنی #کرونا، بیشترین کاربرد را داشته است.

جدول ۴. فهرست ۱۰ هشتگ پرکاربرد در نظرات

| هشتگ | تعداد | درصد نسبی | هشتگ | تعداد | درصد نسبی |
|--------------|-------|-----------|------------------------|-------|-----------|
| #کرونا | ۲۱۸۰۵ | ۱۱/۴۶ | #کرونا_ویروس | ۱۴۳۷ | ۰/۷۶ |
| #کروناویروس | ۲۸۱۹ | ۱/۴۸ | #کرونا_را_شکست_می_دهیم | ۱۳۱۳ | ۰/۶۹ |
| #ایران | ۲۷۳۴ | ۱/۴۴ | #در_خانه_بمانیم | ۱۲۲۵ | ۰/۶۴ |
| #ویروس_کرونا | ۲۵۳۹ | ۱/۳۳ | #سراب_غرب | ۱۱۷۰ | ۰/۶۱ |
| #قرنطینه | ۱۵۸۸ | ۰/۸۳ | #تهران | ۱۱۵۶ | ۰/۶۱ |

تحلیل دیگری که در این پژوهش ارائه می‌شود به بررسی وجود رابطه بین موضوع و هشتگ‌های مرتبط با آن موضوع می‌پردازد. برای این هدف، از ضریب همبستگی پیرسون استفاده می‌شود. همان‌طور که در بخش ۳-۳ توضیح داده شد، امتیاز ضریب همبستگی پیرسون عددی بین ۱ تا ۱- است. در این پژوهش، اگر امتیاز ضریب همبستگی پیرسون صفر باشد، هیچ‌گونه ارتباطی بین موضوع و هشتگ وجود ندارد؛ هرچقدر امتیاز به‌دست آمده به سمت ۱ میل کند، این ارتباط بیشتر است و اگر به سمت ۱- میل کند، رابطه معکوس میان آن هشتگ و آن موضوع برقرار است که بیانگر عدم نمایش آن هشتگ در آن موضوع است. در جدول (۵)، هشتگ‌هایی که بالاترین ضریب همبستگی پیرسون را با موضوع داشته‌اند گزارش شده‌اند.

براساس نتایج گزارش شده در جدول (۵)، همبستگی‌های کشف شده بین اکثر موضوعات استخراج شده از پیکره کرونا و هشتگ‌های انتخاب شده توسط کاربران منطقی و معنادار است. به‌عنوان مثال، بین موضوع «پزشکی» و هشتگ «#جهادگران_سلامت» همبستگی بالایی وجود دارد؛ به این معنا که این هشتگ، در اکثر مواقع، در نظرات، مرتبط با موضوع «پزشکی» به کار رفته و در سایر موضوعات، مانند «سیاسی» و «اجتماعی»، به کار نرفته است.

شایان ذکر است این ضریب همبستگی برای تمام موضوعات استخراج نشده است؛ به این معنا که بین برخی موضوعات، مانند «تشخیص کرونا»، «رسانه»، «داستانک» و «توهین آمیز»، و هشتگ‌های به کاررفته برای آنها همبستگی وجود ندارد و ضریب همبستگی بین این موضوعات و هشتگ‌ها در حدود صفر است.

جدول ۵. همبستگی میان هشتگ و موضوع نظرات

| موضوع | هشتگ | موضوع | هشتگ |
|-----------------------|---|---------------------|--|
| اجتماعی | #قرنطینه، #قرنطینه_خانگی، #مرگ | درمان | #ایران_قوی، #وزارت_بهداشت |
| اخبار بین‌الملل کرونا | #سراب_غرب، #ایران_قوی، #پزشکان_بدون_مرز، #آمریکا، #ایتالیا، تحریم | رویدادهای پیرامون | #رأی، #عید_نوروز، #انتخابات |
| اخبار داخلی کرونا | وزارت_بهداشت، #کووید، #ویروس_کرونا | سرگرمی | #فوتبال، #استقلال، #فیلم، #موسیقی |
| افغانستان | #Afghanistan، #افغانستان | سیاسی-داخلی | #آزادی، #رأی، #تحریم، #امید، #جمهوری_اسلامی، #سپاه، #ایران، #اعتراضات_سراسری |
| اقتصادی | #جهش_تولید، #تحریم، #دولت | سیاسی-خارجی | #عراق، #سراب_غرب، #آمریکا، #ترامپ، #چین، #آمریکا |
| بهداشتی | #الکل، #ویروس، #پیشگیری، #ماسک، #کرونادرایران | فاصله‌گذاری اجتماعی | #در_خانه_بمانیم، #قرنطینه |
| پزشکی | #جهادگران_سلامت، #پرستاران، #مدافعان_سلامت، #پرستار، #پاهم_علیه_کرونا | کرونا و زندان | #تبریز |
| پوشش کرونایی | #کرونادرایران، #کروناویروس | مدیریت بحران | #روحانی، #مجلس، #وزارت_بهداشت، #کرونا، #ویروس_کرونا، #شماره_۴۰۳۰، #هبری، #دولت، #حسن_روحانی، #همه_با_هم_علیه_کرونا |
| تبلیغات | #تیشرت، #شلوار، #بوتیک_محب، #boutiquemoheb، #کاپشن | مذهبی | #امام_زمان، #ظهور، #کربلا، #خدا |
| خانوادگی | #قرنطینه، #قرنطینه_خانگی | نوروز | #نوروز، #عید، #سال_نو، #در_خانه_بمانیم، #در_خانه_میمانیم |

۶. نتیجه‌گیری و جمع‌بندی

در این پژوهش، تلاش شد با جمع‌آوری نظرات و پیام‌ها از رسانه‌های اجتماعی تویتر و اینستاگرام و انتخاب موارد مربوط به کرونا، یک پیکره‌زبانی تخصصی به نام «پیکره‌نظرات کرونا» از نظرات و پیام‌های کاربران این دو رسانه اجتماعی تهیه شود. پس از پیش‌پردازش این پیکره، با بهره‌گیری از یک شیوه‌الگوریتمی در تحلیل‌گفتمان، به تحلیل محتوایی این پیکره و خوشه‌بندی نظرات، براساس موضوع‌های نهان در آنها، پرداخته شد. ۲۴ موضوع انتزاعی از پیکره کرونا استخراج شد و هر موضوع، براساس واژه‌های با احتمال بالا، توسط ناظر انسانی برچسب‌زنی شد. این پیکره و روش‌شناسی پردازشی معرفی شده بستری را برای تحلیل‌های متفاوت متخصصان حوزه‌های مختلف علوم انسانی فراهم آورده است که در این پژوهش، فقط به دو نمونه از تحلیل‌های ممکن در بررسی محتوای نظرات و هشتک‌ها پرداخته شده است.

با بررسی داده‌های «پیکره‌نظرات کرونا» و آمار به‌دست‌آمده از پردازش موضوعی آن می‌توان گمانه‌زنی کرد که اگرچه رسالت رسانه‌های اجتماعی اطلاع‌رسانی، آگاه‌سازی و نشر اطلاعات است، موضوعات «سیاسی» و «اجتماعی» و تبادل نظر درباره آنها حدود ۲۵٪ از نظرات در این پیکره را در بر گرفته است. این به آن مفهوم است که در جامعه ایران، دغدغه‌های سیاسی و اجتماعی بیش از هر موضوعی، مانند شیوع همه‌گیری بیماری کرونا که سلامتی آحاد جامعه را تحت‌الشعاع قرار می‌دهد، مورد توجه است و رسانه‌های اجتماعی زمینه‌ای را برای ابراز این دسته از نظرات فراهم آورده است. موضوعاتی که در این پیکره مستقیماً به خود ویروس کرونا و ویژگی‌های یک همه‌گیری مربوط بود طیف متنوعی داشت که در لابه‌لای نظرات به چشم خورد. از میان این موضوعات، «مدیریت بحران» از بیشترین اهمیت نزد کاربران برخوردار بود. امکاناتی مانند رسانه‌های اجتماعی می‌تواند، در شکل یک رسانه، به رسالت اولیه خود که همانا نشر اطلاعات و آگاهی است بپردازد. از این رو، می‌توان از ظرفیت رسانه‌های اجتماعی برای اطلاع‌رسانی استفاده و به افزایش سطح دانش عمومی افراد جامعه در شیوع یک همه‌گیری کمک کرد.

از رابطه همبستگی بین موضوعات و هشتک‌ها چنین نتیجه گرفته شد که این هشتک‌ها تصادفی نبوده‌اند و کاربر، آگاهانه، از محتوای معنایی فشرده آن برای توصیف یک متن

کوتاه استفاده می‌کند. از این رو، استفاده صحیح هشتگ‌ها می‌تواند به وضوح مطلب کمک شایانی کند.

دستاورد های مهم این پژوهش یافتن راهکاری رایانشی برای بیان وجود رابطه معنایی بین محتوای یک متن کوتاه (موضوع متن) و هشتگ‌های تعریف شده برای آن است. این دستاورد می‌تواند زمینه‌ساز پژوهش‌های آتی باشد، به این صورت که امکان خوشه‌بندی هشتگ‌ها، پیش‌بینی خودکار هشتگ براساس مطلب نوشته شده و همچنین، پیشنهاد خودکار هشتگ مناسب براساس متن به کاربر فراهم شود. وجود چنین امکاناتی به افزایش انسجام درونی بیشتر در متن کوتاه می‌انجامد.

تعارض منافع

تعارض منافع ندارم.

ORCID

Masood Ghayoomi  <http://orcid.org/0000-0001-6685-1332>

منابع

- عاصی، مصطفی (۱۳۸۴). پایگاه داده زبان فارسی در اینترنت. *خبرنامه پژوهشگاه علوم انسانی و مطالعات فرهنگی*، ۲، ۱۳-۱۶.
- قیومی، مسعود (۱۳۹۳). پردازش زبان طبیعی و زبان فارسی: نیازها و چالش‌ها. ارائه شده در *میزگرد تخصصی پردازش زبان طبیعی و علوم شناختی*، پژوهشگاه علوم و فناوری اطلاعات (ایرانداک)، تهران.
- قیومی، مسعود (۱۳۹۷). ارائه یک روش مبتنی بر مدل زبانی برای واحدسازی پیکره فارسی. *زبان و زبان‌شناسی*، ۱۴(۲۷)، ۲۱-۵۰.
- قیومی، مسعود (۱۳۹۸). ساماندهی تحلیل‌های چندلایه‌ای زبان‌شناختی در پیکره‌های فارسی. *واژه ویژه زندگی: جشن‌نامه استاد ویدا شتقاقی*. ۲۸۷-۳۱۲. تهران: نشر نویسه پارس.
- قیومی، مسعود (۱۳۹۹). چالش سامانه‌های پردازش زبان طبیعی در مواجهه با نوآژه‌های «کرونایی». ارائه شده در *ششمین پیش‌نشست تخصصی ابعاد انسانی-اجتماعی کرونا در ایران*. پژوهشگاه علوم انسانی و مطالعات فرهنگی.

References

- Abd-Alrazaq, A., Alhuwail, D., Househ, M., Hamdi, M., & Shah, Z. (2020). Top concerns of Tweeters during the COVID-19 pandemic: Infoveillance study. *Journal of Medical Internet Research*, 22(4), e19016. <https://doi.org/10.2196/19016>
<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7175788/>
- Bailey, P., Craswell, N., & Hawking, D. (2003). Engineering a multi-purpose test collection for Web retrieval experiments. *Information Processing and Management*, 39(6), 853–871. [https://doi.org/10.1016/S0306-4573\(02\)00084-5](https://doi.org/10.1016/S0306-4573(02)00084-5)
- Blei, D. M., Ng, A. Y., Jordan, M. I., & Lafferty, J. (2003). Latent Dirichlet allocation. *Journal of Machine Learning Research*. 3, 993–1022. <https://doi.org/10.1162/jmlr.2003.3.4-5.993>
- Boddy, R., & Smith, G. (2009). *Statistical Methods in Practice: For scientists and Technologists*. Chichester, U.K.: Wiley. <https://doi.org/10.1002/9780470749296.ch10>
- Chaitra, P.G., Deepthi, V., Vidyashree, K. P., & Rajini, S. (2018). A study on different types of Web crawlers. In S. Choudhury et al. (eds.), *Intelligent Communication, Control, and Devices, Advances in Intelligent System and Computing*, Vol. 989, 781-789. https://doi.org/10.1007/978-981-13-8618-3_80
- Cothey, V. (2004). Web-crawling reliability. *Journal of the American Society for Information Science and Technology*. 55(14), 1228–1238. <https://doi.org/10.1002/asi.20078>
- de Saussure, F. (1916). *Cours de linguistique générale*, Lausanne, Paris: Payot.
- Dong, M., Cao, X., Liang, M., Li, L., Liu, G., & Liang, H. (2020). Understand research hotspots surrounding COVID-19 and other Coronavirus infections using topic modeling, <https://www.medrxiv.org/content/10.1101/2020.03.26.20044164v2>
<https://doi.org/10.1101/2020.03.26.20044164>
- Griffiths, T. L., & Steyvers, M. (2004). Finding Scientific topics. *Proceedings of the National Academy of Sciences*, Vol. 101, 5228–5235. <https://doi.org/10.1073/pnas.0307752101>
- Gurajala, S., & Matthews, J. N. (2018). Twitter data analysis to understand societal response to air quality. *Proceedings of the 9th International Conference on Social Media and Society*, 82–90. <https://doi.org/10.1145/3217804.3217900>

- Hofmann, T. (1999). Probabilistic latent semantic indexing. *Proceedings of the Twenty-Second Annual International SIGIR Conference on Research and Development in Information Retrieval*.
<https://doi.org/10.1145/312624.312649>
- Kaila, R. P., & Prasad, A. V. K. (2020). Informational flow on Twitter – Coronavirus outbreak – Topic modeling approach. *International Journal of Advanced Research in Engineering and Technology*, 11(3), 128-134.
- McCallum, A. K. (2002). Mallet: A machine learning for language toolkit. <http://mallet.cs.umass.edu>.
- Minka, T., & Lafferty, J. (2002). Expectation-propagation for the generative aspect model. *Proceedings of the 18th Conference on Uncertainty in Artificial Intelligence*, 352–359, San Francisco, CA, USA.
- Ordun, C., Purushotham, S., & Raff, E. (2020). Exploratory analysis of Covid-19 tweets using topic modeling, UMAP, and DiGraphs <https://arxiv.org/pdf/2005.03082.pdf>
- Papadimitriou, C., Raghavan, P., Tamaki, H., & Vempala, S. (2000). Latent semantic indexing: A probabilistic analysis. *Journal of Computer and System Sciences*, 61(2), 217-235. <https://doi.org/10.1006/jcss.2000.1711>
- Sonbhadra, S. S., Agarwal, S., & Nagabhushan, P. (2020). Target specific mining of COVID-19 scholarly articles using one-class approach. *Chaos, Solitons & Fractals*, 140, 110155. <https://doi.org/10.1016/j.chaos.2020.110155>
- Tognini-Bonelli, E. (2001), *Corpus Linguistics at Work, Studies in Corpus Linguistics*, Amsterdam, The Netherlands: John Benjamins Publishing Company.
- Tsur, O., & Rappoport, A. (2012). What's in a Hashtag? Content based prediction of the spread of ideas in microblogging communities. *Proceedings of the 5th ACM international conference on Web search and data mining*, 643-652. <https://doi.org/10.1145/2124295.2124320>
- Assi, M. (2005). Persian linguistic database in Internet. *Newsletter of the Institute for Humanities and Cultural Studies*, 2, 13-16. [In Persian].
- Ghayoomi, M. (2014). Natural language processing and the Persian language: Requirements and challenges. *Expert Meeting on Natural Language Processing and Cognitive Science*, Institute for Information science and Technology, Tehran. [In Persian].
- Ghayoomi, M. (2018). A tentative method of tokenizing Persian corpus based on language modeling. *Language and Linguistics*, 14(27), 21-50. [In Persian].

Ghayoomi, M. (2020). Structuring multilayer linguistic analyses in linguistic corpora. *Word by Word Life: Festschrift for Professor Vida Shaghghi*, pp:278-312, Tehran: Neveeseh Parsi Publications. [In Persian].

Ghayoomi, M. (2020). Challenge of Natural language processing systems by facing coined “Corona” words. *The 6th Expert Meeting on Human-social Dimensions of the Coronavirus in Iran*, Institute for Humanities and Cultural Studies, Tehran. [In Persian].

استناد به این مقاله: قیومی، مسعود. (۱۴۰۰). تحلیل محتوایی موضوع‌ها و هشتگ‌های کرونا در رسانه‌های اجتماعی. *علم*

زبان، ویژه‌نامه کرونا، سال ۸، ۸۷-۱۱۵. Doi: 10.22054/ls.2020.53557.1356



Language Science is licensed under a Creative Commons Attribution-Noncommercial 4.0 International License.