تحلیل متون کوتاه در بستر پیام رسان تلگرام

مطالعه موردی : گروه دوستیابی تصادفی

اطلاعی از تعداد اعضا گروه موجود نیست.

بازه زمانی : اولین رکورد 2018-03-06 06:55:02 آخرین رکورد 2018-03-06 07:21:11

ویژگی پیام رسان (گروه ها) را میتوان محدودیت در وایرال شدن کاربر دانست، ارتباطات دو به دو بیشتر شکل میگیرد، کاربر فیک به راحتی حذف میشود. تفاوت پیام رسان و شبکه اجتماعی در تعداد یال ها و نودها است، از نظر فنی ماتریس مجاورت در پلتفورم پیام رسانها اهمیت بیشتری دارد.

اما دادهای ما در متون کوتاه :

من متونی که بین ۳ تا ۵ جمله باشد را متون کوتاه در نظر میگیرم.

داده های گروه های تلگرامی شامل موارد ذیل است

کلمات، استیکر ، ایموجی

تلگرام استیکرها معادل ایموجی دارند.

هر پست تلگرامی یک واحد معناداری در نظر گرفته شده است.

چند تا رویکرد در تحلیل این متون میتوان داشت:

۱. ایموجی ها

۲. کلماتی که بیشترین تکرار را دارند. [معمولا کار درست های در کشور عزیزمان اینرا به عنوان محصول تجاری دانش بنیان ارائه میکنند]

۳. گراف جهت دار و وزن دار روابط بین اعضای گروه

۴. تحلیل حس های منتقل شده بر اساس کلمات

۵. تحلیل حس های منتقل شده بر اساس ایموجی ها

۶.تحلیل معنا

۷. تحلیل روند

در هر ست دیتا ، فیچرها و یا ابعادی از اطلاعات وجود دارند که با توجه به هدف نهایی انتخاب میشوند، مهم است که از چه اطلاعاتی برای چه خروجی استفاده شود. اطلاعات ما شامل موارد ذیل بود :

User\_id

Group\_id

Post\_id

Post\_date

Message

Media\_id

Forward\_from\_user

Forward\_from\_channel

Reply\_to

Object

Date

Time

robot\_see\_date

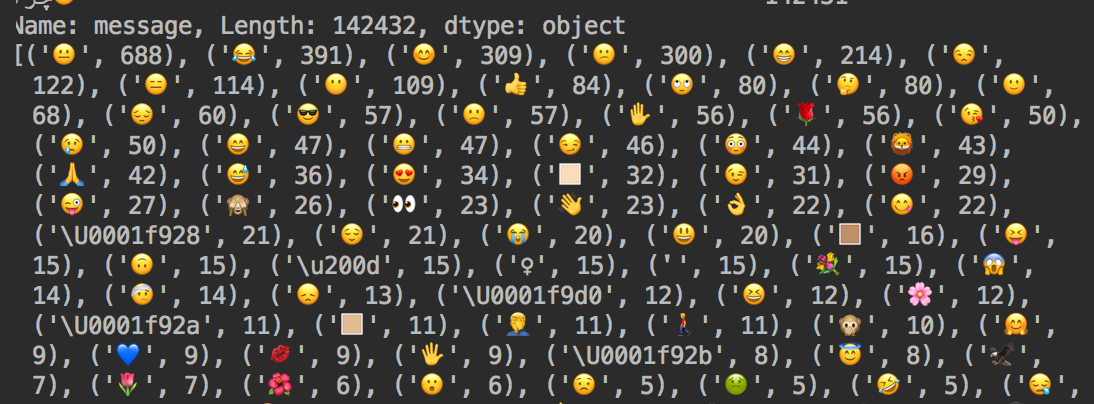
من در این گزارش به تحلیل ایموجی ها بسنده میکنم .

در داده های اولیه تقریبا ۱۵ هزار رکورد وجود دارد،

داده ها تبدیل به فریم ورک شدند، البته اینکه از دیکشنری استفاده نکردم چون نوع داده ها متفاوت بود و فریم ورک ها خروجی بهتری به من میدادند.

داده ها تبدیل به فریم ورک شدند، من با توجه به نوع خروجی مورد نیازم Message ها رو انتخاب کردم،

بعدش ایموجی ها رو اکسترکت کردم توی این کد ایموجی و استیکر با هم برابر سازی شده، این ایموجی ها رو شمارش کردم، قاعدتا و به نظر میرسید که الان باید ایموجی red heart با توجه به موضوع گروه به عنوان پر تکرارترین ایموجی داشته باشیم، اما خب نتایج بسیار جذابتر بود :



معانی هر ایموجی در این آدرس آمده است :

https://emojipedia.org/people

در تعریف پرتکرارترین ایموجی امده است:

A face featuring a straight-and-closed mouth, not giving away any particular emotion.

بله بی حسی !

جالب است همه ما به عنوان اعضای جامعه این بی حسی را هر روز تجربه میکنیم … حتی در گروه هایی که به دنبال معشوق هستیم این برای من بسیار تلخ بود . این گزارش ایسنا را بخوانید :

[https://www.isna.ir/news/95022918468/حال-جوان-امروز-چطور-است](https://www.isna.ir/news/95022918468/%D8%AD%D8%A7%D9%84-%D8%AC%D9%88%D8%A7%D9%86-%D8%A7%D9%85%D8%B1%D9%88%D8%B2-%DA%86%D8%B7%D9%88%D8%B1-%D8%A7%D8%B3%D8%AA)

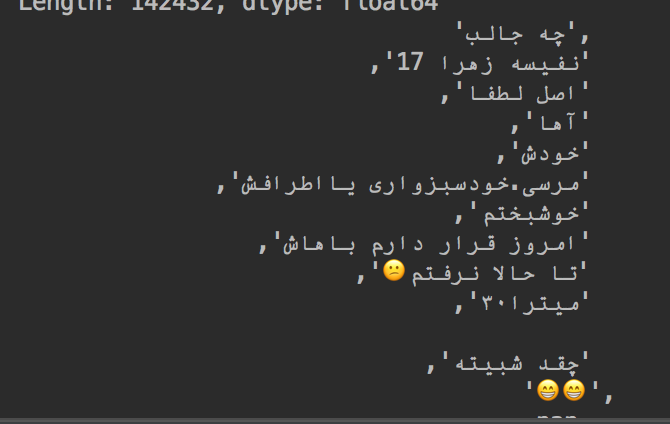
[بی تفاوتی اجتماعی؛علل شکل گیری،آسیب ها،آثارو راهکارخروج از آنwww.iribnews.ir/.../بی-تفاوتی-اجتماعیعلل-شکل-گیریآسیب-هاآثا...](http://www.iribnews.ir/fa/news/1670657/%D8%A8%DB%8C-%D8%AA%D9%81%D8%A7%D9%88%D8%AA%DB%8C-%D8%A7%D8%AC%D8%AA%D9%85%D8%A7%D8%B9%DB%8C%D8%B9%D9%84%D9%84-%D8%B4%DA%A9%D9%84-%DA%AF%DB%8C%D8%B1%DB%8C%D8%A2%D8%B3%DB%8C%D8%A8-%D9%87%D8%A7%D8%A2%D8%AB%D8%A7%D8%B1%D9%88-%D8%B1%D8%A7%D9%87%DA%A9%D8%A7%D8%B1%D8%AE%D8%B1%D9%88%D8%AC-%D8%A7%D8%B2-%D8%A2%D9%86)

قاعده این است که مکالمه دو طرف دارم یک ایموجی ارسال شده سه حالت ممکن است رخ دهد ۱. بی پاسخ بماند ۲. استیکر و یا ایموجی دریافت کند. ۳. کلمه دریافت کند، برای فاز دوم امدم و توالی ایموجی ها را محاسبه کردم.

از آنجا که در کتابخانه ایموجی چیزی پیدا نکردم، تمام اطلاعات را ایندکس کردم و فاصله ایموجی ها را محاسبه کردم !

من از distance بین دو ایموجی استفاده کردم، نتایج به شرح ذیل است :

به نظر میرسد اغلب در ادامه ایموجی کلمه رد و بدل میشود ، این نمایی از کلمات ایندکس شده است.



به نظر میرسید بد از ایموجی (پوکر فیس) معمولا واکنش به صورت کلمه است: این یعنی در جواب بی حسی کلمه ای برای برقراری ارتباط به کار برده میشود.

این یک نمونه کوچک و دم دستی از اندیشیدن به شیوه یک text minner است. هنوز هم معتقدم باید بدانی خروجی کجا خلق ارزش میکند و کجا سوداور خواهد بود .

من اگر باشم برای طراحی خدمت به این دسته از کاربران این گروه، پر تکرار ترین کلمات، هم رخدادی و هم معنایی کلمات استفاده میکنم، این یک راهنمای اولیه است برای تعیین نقطه خلق ارزش از داده های که به ظاهر بی ارزش هستند

این گزارش را من سمیرا کرانی نوشتم و اگر خواستید با استفاده از داده اولیه هر یک از انواع گزارش های فوق را استخراج کنید، من با ایمیل زیر در خدمت تان خواهم بود :

samira.korani@gmail.com