

بسمه تعالی

# WordMap مناظرات انتخاباتی سال ۱۳۹۶

<https://github.com/amirmohammadkz/wordMap>

امیر محمد کاظمینی زاده

## چکیده

مناظرات انتخاباتی سال ۹۶ که بین ۶ نامزد برگزار شده بود، نقش بسزایی در انتخاب ملت داشت. در این ماراتن چهار ساعته که در سه نوبت برگزار شده بود، نامزد ها به پرسش های مجری برنامه پاسخ می دادند و در ادامه هریک به نقد فرد پاسخ دهنده می پرداختند. در آخر نیز فرصت دفاعی برای فرد پاسخ دهنده در نظر گرفته می شد و روال سوالات بدین صورت ادامه پیدا می کرد.

## مراحل انجام کار

### • جمع آوری داده

داده های جمع آوری شده از سایت [www.mashreghnews.ir](http://www.mashreghnews.ir) جمع آوری شدند؛ تنها سایتی که به طور تقریباً کامل صحبت های این مناظرات را به صورت کتبی در سایت انتشار داده بود.

### • خواندن داده

خواندن داده و ذخیره سازی آن به چند صورت انجام گرفت. برای هر نامزد سه فایل text برای هر سه بخش مناظرات در نظر گرفته شد تا با تفکیک پذیری بالا امکان مقایسه وجود داشته باشد. همچنین برای خواندن داده ها نیز چند دسته بندی اتخاذ شد:

۱. All: متودی که تمامی مناظرات را با هم ترکیب می کند

۲. First: متن های مناظره اول

۳. Second: متن های مناظره دوم

۴. Third: متن های مناظره سوم

۵. و متود دیگر برای ترکیب کردن حرف های هر کدام از نامزدها در سه مناظره.

### • Tokenize کلمات

در این مرحله کل فایل متن باز می شود و کلمه به کلمه تفکیک می شود و و به همراه تکرار هر کلمه، در دیکشنری مربوطه ذخیره می شود. لازم به ذکر است که قبل از این کار کلمه هایی که بینشان فاصله وجود دارد مانند "می رود"، "رفته ای" و ... نیز اصلاح می شوند.

### • پاکسازی داده

در این مرحله کلمه‌های جدا شده در صورت نداشتن ساختار مناسب و یا همراه داشتن علائمی مانند نقطه، ویرگول و ... به وسیله regex تشخیص داده می‌شود و پاکسازی می‌شود. لازم به ذکر است که در قسمت ابتدایی مرحله‌ی ریشه‌یابی کلمات نیز انجام می‌شد. با توجه به تغییر لحن صحبت‌ها این مرحله حذف شد.

- هرس کردن داده
  - در این مرحله به هرس کردن حروف ربط و حروف اضافی می‌پردازیم
  - مرتب‌سازی داده
  - در این مرحله داده‌ها به ترتیب تکرار مرتب می‌شوند
  - ذخیره‌سازی داده
  - در این مرحله کلمات پاکسازی شده به همراه تکرار به صورت مرتب شده ذخیره می‌شوند
  - ذخیره‌سازی به فرمت wordCloud
  - برای ساختن word map از ماژول WordCloud استفاده شده. برا استفاده از این ماژول نیاز به ذخیره‌سازی به فرمتی خاص است که بدان صورت ذخیره می‌کنیم
  - استفاده از wordCloud به همراه تعیین stopword های نهایی
- در این مرحله که مرحله آخر می‌باشد، با توجه به وردمپ تولید شده، فعل‌ها و یا کلماتی که معنی ارزش افزوده خاصی در روند انتقال مفهوم به مخاطب را ندارند و تعداد قابل توجهی از آن‌ها موجود است، حذف می‌کنیم. سپس با تنظیمات دلخواه، وردمپ را ذخیره می‌کنیم.

## نتایج



کل مناظرات





مناظره‌ی دوم



مناظره‌ی سوم





روحانی



رئیس



جهانگیری



میر سلیم



هاشمی طباطبائی