



مقایسه سخنان رهبری در دوره‌های مختلف ریاست جمهوری

مصطفی محمد علی ابراهیم



جمع آوری داده

- مرجع
 - `farsi.khamenei.ir`
- مشکلات
 - یکسان نبودن قالب صفحات

داده‌ها از لحاظ آماری

دکتر روحانی	دکتر احمدی نژاد	حجة الاسلام خاتمی	آیت الله هاشمی	/
257	442	385	615	تعداد سخنرانی‌ها
151502	265426	253442	314402	تعداد کلمات
34466	43292	42962	57219	تعداد کلمات یکسان



پیش پردازش ها

- ابزار

- ماژول هضم

- مراحل

- درست کردن فاصله و نیم فاصله و حذف whitespace ها (Normalizer)

- جدا کردن کلمات (word_tokenize)

- برچسب گذاری کلمات (tagger)

- حذف کلمات ربط و جدا کننده ها و حروف اضافه



نتایج

آیت اللہ ہاشمی

- مردم
- اسلام
- کار
- دنیا
- کشور
- ایران
- انقلاب
- خدا

حجة الاسلام خاتمی

- مردم
- کشور
- اسلام
- کار
- ایران
- دنیا



دکتر احمدی نژاد

- مردم
- کشور
- کار
- ایران
- اسلام
- انسان



-

کلمات با بیشترین اختلاف
نسبی در گروه دکتر روحانی
نسبت به دکتر احمدی نژاد





با تشکر از توجه شما