

موضوع: تشخیص اخبار جمهوریخواه از دموکرات

نام اعضا: - بابک سفیدگر

- امیر اسمعیلی

- برای جمع‌آوری داده از کتابخانه news-please استفاده می‌کنیم، به اینصورت که ابتدا سایت‌های خبری که نزدیک به یک حزب سیاسی هستند را انتخاب می‌کنیم و سپس تعدادی از لینک خبرهای آن‌ها را دستی لیبیل می‌زنیم و سپس متن خبر آن‌ها را در میاوریم .

```
264/1004 -> title: Why Emmanuel Macron is the anti-Donald Trump
265/1004 -> title: Raul Labrador's worst week in Washington
266/1004 -> title: Trump singles out historically black college funds in signing statement
267/1004 -> title: Priebus defends proposed cuts to W.H. office battling drug epidemic
268/1004 -> title: Kasich: Republican health care bill is 'inadequate'
269/1004 -> title: How House Republicans won those final health care votes
270/1004 -> title: Sally Yates told White House aides Flynn was a Russian blackmail risk. 18 days later, he was fired.
271/1004 -> title: Obama warned Trump about hiring Flynn
272/1004 -> title: WH highlights Clapper's lack of evidence on Trump-Russia collusion
273/1004 -> title: Jimmy Carter on Sanders: 'Can y'all see why I voted for him?'
274/1004 -> title: Donald Trump Jr. calls on Indiana Republicans to 'stay active'
275/1004 -> title: Rep. Evan Jenkins to challenge Sen. Joe Manchin
276/1004 -> title: Latest on Afghanistan options expected to hit Trump's desk this week
277/1004 -> title: DC festival showcases political activism, healthy living
278/1004 -> title: Sally Yates says she warned White House that Flynn was a blackmail risk
279/1004 -> title: Sally Yates explains why she won't defend travel ban
280/1004 -> title: 'Melrose Place' actor Antonio Sabato Jr. is running for Congress
281/1004 -> title: Why Trump didn't listen to all the people who told him to stay away from Mike Flynn
282/1004 -> title: Pruitt removes scientists from key EPA board
283/1004 -> title: Five things we learned from the Sally Yates hearing
284/1004 -> title: Mitch McConnell: the man who may decide health care bill
285/1004 -> title: White House meeting on Paris climate deal delayed
286/1004 -> title: Trump to unveil list of conservative judges to fill lower federal courts
287/1004 -> title: Trump campaign removes Muslim ban language from website
288/1004 -> title: Judges appear divided on Trump's travel ban
289/1004 -> title: Donald Trump is turning liberals into conspiracy theorists
290/1004 -> title: In angry letter, Page discloses 'brief interactions' with Russian official
291/1004 -> title: Kushner family in Beijing: 'Invest $500,000 and immigrate' to US
292/1004 -> title: Tiffany Trump is heading to Georgetown Law
293/1004 -> title: Dems hope Rob Quist will win over Montana's red voters
294/1004 -> title: GOP congressman: Republicans were 'irresponsible' to celebrate AHCA passage at White House
295/1004 -> title: Hill inquiries on Russia run into hurdles
296/1004 -> title: What could possibly happen before Trump's 100th day? We're keeping track
297/1004 -> title: Trump holds a 100-day rally on friendly turf, snubbing the White House Correspondents' Dinner
298/1004 -> title: Trump stumbled on healthcare and immigration, but he's been 'a wrecking ball' on the environment
299/1004 -> title: Tillerson calls for stiffer sanctions on North Korea as Pyongyang launches another missile
300/1004 -> title: Congress approves stopgap spending bill to prevent federal shutdown, but Obamacare repeal stalls
301/1004 -> title: Economic growth stumbles to worst rate in two years, underscoring Trump's challenge
302/1004 -> title: 100 days in, Trump's supporters back him, but some doubts have begun to show
303/1004 -> title: Don't stress over robots; a bright new economy is being born
304/1004 -> title: Why 100 days? Trump thinks the benchmark is lame. Is he right?
305/1004 -> title: At senators' town halls, voters take an uneven measure of Kamala Harris and Dianne Feinstein

successfully downloaded 904 out of 1004 articles.

=====
Total Natural Democrat Republican
904 12 243 649
=====
```

- در مرحله بعد بر روی متن‌های جمع‌آوری شده را که به صورت فایل json ذخیره شده‌اند، عمل پاکسازی انجام می‌دهیم و در دایرکتوری data/clean ذخیره می‌کنیم و سپس توکن‌هایشان را در می‌آوریم و در دایرکتوری data/tokens به صورت json ذخیره می‌کنیم.

- برای نحوه کار کد و روند کار به کدها و فایل Readme.md مراجعه شود.

## - نتیجه:

