

nytR: An R package for the NY Times API

Shane Conway

December 23, 2009

Abstract

The **nytR** package creates a simple interface for the NY Times API. The Times API provides access to various different data sources. At present **nytR** only exposes parts of the Congress API for vote data.

1 Introduction

There is currently no easy way to pull congressional vote data into R. Given the usage of R within the political science community, and the potential wealth of data available, **nytR** aims to fill this gap.

This is still very much a beta version. I expect to expose more data in future versions.

2 Configuration

The NY Times requires each user to register for their own API key so that they can track usage. This can be done in a few quick steps on the key registration page: <http://developer.nytimes.com/apps/register>.

Once you have a key for the Congress API, the easiest way to use the package is to create an environment variable called `NYTIMES_CONGRESS` with contains the key (this is used every time you try to access the live data). You can confirm that this is working by calling `Sys.getenv("NYTIMES_CONGRESS")` from within R.

3 Congressional data

Currently, only the API that has been exposed is the Congress API.

3.1 Vote data

By default, the package gets roll-call vote data:

```
> library(nytR)
> xml <- getNYTCongress(110, "senate", 2, 194)
> x <- voteSummary(xml)
> x$description
```

```
[1] "H.R. 4137 Conference Report; College Opportunity and Affordability Act of 2008"
```

There are a number of utility functions for parsing the XML output.
The *parseVotes* function returns most of these results as a list of lists.

```
> votes <- parseVotes(xml)
> votes$total.votes
```

```
$yes
[1] "83"
```

```
$no
[1] "8"
```

```
$present
[1] "1"
```

```
$not_voting
[1] "8"
```

```
$margin
[1] "75"
```

This also includes the raw vote data.

```
> require(xtable)
> xtable(head(votes$votes), caption = "List of votes by member",
+       label = "tab:one", table.placement = "tbp", caption.placement = "top")
```

| | member_id | vote_position |
|-----------|-----------|---------------|
| position | A000069 | Yes |
| position1 | A000360 | No |
| position2 | A000109 | Yes |
| position3 | B001261 | Yes |
| position4 | B000243 | Yes |
| position5 | B001233 | Yes |

Table 1: List of votes by member

The *getNYTCongress* function currently supports 5 different vote types: roll.call, missed.votes, party.votes, nomination.votes, and members (see next section). missed.votes, party.votes, and members can all be parsed with the *em voteTypeDetail* function, while nomination.votes are processed with the *em voteNomDetail* function.

```
> xml <- getNYTCongress(type = "party.votes", 110, "senate")
> head(x <- voteTypeDetail(xml))
```

| | id | name | party | state | total_votes | votes_with_party |
|---|---------|------------------|-------|-------|-------------|------------------|
| 1 | S000663 | Olympia J. Snowe | R | ME | 657 | 413 |
| 2 | C001035 | Susan Collins | R | ME | 657 | 433 |

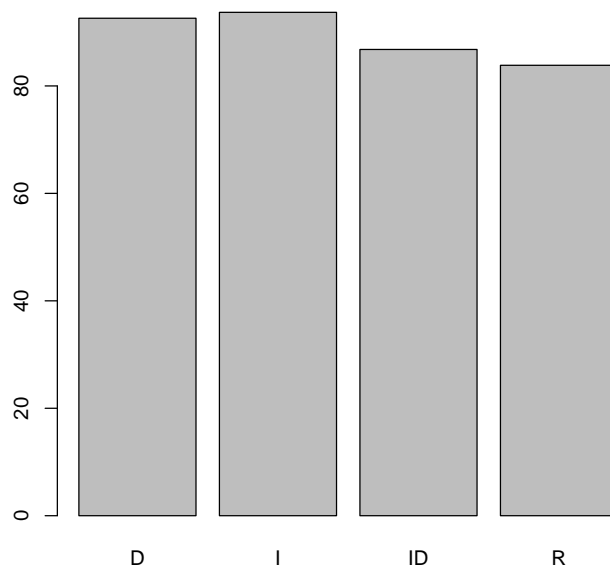
| | | | | | | |
|---|---------|---------------------|---|----|-----|-----|
| 3 | S000709 | Arlen Specter | R | PA | 657 | 447 |
| 4 | S001142 | Gordon Harold Smith | R | OR | 657 | 460 |
| 5 | V000126 | George V. Voinovich | R | OH | 657 | 474 |
| 6 | C001057 | Norm Coleman | R | MN | 657 | 488 |

| | party_votes_pct | rank | notes |
|---|-----------------|------|------------------------------|
| 1 | 62.86 | 1 | NULL |
| 2 | 65.91 | 2 | NULL |
| 3 | 69.09 | 3 | NULL |
| 4 | 70.55 | 4 | Defeated in general election |
| 5 | 73.49 | 5 | NULL |
| 6 | 76.97 | 6 | NULL |

```

> x.summary <- do.call("rbind", x.summary <- by(x$party_votes_pct,
+ x$party, summary))
> barplot(x.summary[, "Mean"])

```



3.2 Members data

Currently, the primary function for getting member data is with *getNYTCongress*:

```

> xml <- getNYTCongress(type = "members", 110, "senate")
> xtable(head(voteTypeDetail(xml))[, c(1, 2, 6, 7, 8)], caption = "List of senate members from t
+ label = "tab:one", table.placement = "tbp", caption.placement = "top")

```

| | id | name | party | state | missed_votes_pct |
|---|---------|---------------------|-------|-------|------------------|
| 1 | A000069 | Daniel K. Akaka | D | HI | 0.30 |
| 2 | A000109 | Wayne A. Allard | R | CO | 3.50 |
| 3 | A000360 | Lamar Alexander | R | TN | 4.41 |
| 4 | B000243 | Max Baucus | D | MT | 0.30 |
| 5 | B000382 | Robert F. Bennett | R | UT | 1.22 |
| 6 | B000444 | Joseph R. Biden Jr. | D | DE | 31.81 |

Table 2: List of senate members from the 110th congress

A *getMemberDetail* function also exists which retrieves XML details by member id.

4 Future plans

Future enhancements to the package will include:

- Expose all of the Congress API.
- Complete documentation and provide more meaningful examples
- Introduce other APIs (e.g. real estate, article search)