**1. New York Times**

**Time Spent:** 25 min

**Link:** http://www.nytimes.com/

**API:** Yes; http://developer.nytimes.com/docs

**Search Format:**

http://api.nytimes.com/svc/search/v2/articlesearch.*response-format*?[q=*search term*&fq=*filter-field*:(*filter-term*)&*additional-params*=*values*]&api-key=####

<http://developer.nytimes.com/docs/read/article_search_api_v2>

**Response Information:**

http://developer.nytimes.com/docs/read/article\_search\_api\_v2#examples

**JSON tags:**

web\_url - link to article

snippet

lead\_paragraph

abstract

source

multimedia information (url) - image link, size, type

headline (main, kicker),

keywords (name, value, rank) - name is type, value is actual data, rank of keyword

pub\_date

document\_type

section\_name

byline (person (firstname, middlename, lastname, rank, role, organization), original) - author information

**Website Meta Data:**

<meta name="hdl" content = "Article Name" />

<meta name="author" content="author name" />

<meta name="byl" content="by line" />

<meta property="article:published" itemprop="datePublished" content="2014-09-12" />

<meta property="article:section" itemprop="articleSection" content="i.e. Africa" />

<meta property="article:tag" content="Ebola Virus" />

<meta name="des" content="Ebola Virus" /> (name = des, per, geo)

<meta name="keywords" content="Ebola Virus,Epidemics,Hospitals,Johnson Sirleaf Ellen,Obama Barack,Africa" />

<meta name="news\_keywords" content="Ebola;Epidemic;Hospital;Ellen Johnson Sirleaf;Barack Obama;Africa" />

**Website HTML Tags:**

class="story-body-text story-content"

itemprop="articleBody"

2. **The Guardian**

**Time** Spent: 30 mins

**Link:** http://www.theguardian.com/us

**API:** Yes; http://explorer.content.guardianapis.com/

**Search Format:** <http://content.guardianapis.com/search?q=>

http://www.theguardian.com/open-platform/content-api-tag-search-reference-guide

**Response Information:**

http://www.theguardian.com/open-platform/content-api-item-reference-guide

In JSON and XML

There's an API explorer

showtags

keyword (id, type, webTitle)

contributor

show-fields

type

headline

byline

body

show-media

picture, video

webTitle

webURL

**Website Meta Data:**

<meta property="og:title" content="title"/>

<meta property="article:published\_time" content="2014-09-13T00:00:23+01:00" />

<meta property="article:tag" content="keyword" />

<meta property="og:url" content="article url"/>

<meta property="og:image" content="main image link"/>

<meta name="author" content="Author (and maybe position)" />

<meta name="keywords" content="keywords" />

<meta name="news\_keywords" content="keywords" />

**Website HTML Tags:**

class="flexible-content-body"

**3. The People's Daily**

**Time Spent:** 15 min

**Link:** http://english.peopledaily.com.cn/

**API:** No

**Website Meta Data:**

<meta name="keywords" content="Xi&nbsp;Jinping,SCO four-point proposal makes cooperation Dushanbe Organizat" />

<meta name="publishdate" content="2014-09-13" />

<meta name="author" content="F\_100622(?)" />

**Website HTML Tags:**

class="wb\_12 wb\_12b clear" id="p\_content"

<title>Title in head has article title and - People's Daily Online</title>

<div class="wb\_4 clear"> Followed by Editors

**4. The Daily Mail**

**Time Spent:** 15min

**Link:** http://www.dailymail.co.uk/ushome/index.html

**API**: No

**Website Meta Data:**

<meta name="keywords" content="keywords" />

<meta name="news\_keywords" content="keywords" />

<meta name="description" content="description" />

<meta name="original-source" content="url" />

<meta property="article:published\_time" content="2014-09-13T17:05:13.000Zyyyy-MM-dd HH:mm:ss z" />

<meta property="og:title" content="title" />

**Website HTML Tags:**

class="mol-para-with-font"

**5. The Washington Post**

**Time Spent**: 25 min

**Link:** http://www.washingtonpost.com/

**API**: Yes, https://apiportal.washingtonpost.com/api/trove

**Search Format**:

http://api.washingtonpost.com/trove/v1/resources/4113518?key=

**Response Information:**

resources (id, name, subtype, type) - keywords

displayName - title

published - date

authors (displayName, id, objectType) - author

relatedChannels (displayName, id, objectType) - related channels

**Website Meta Data:**

<meta name="keywords" content="keywords"/>

<meta name="description" content="description"/>

<meta name="news\_keywords" content="keywords"/>

<meta property="og:title" content="title"/>

**Website Javascript:**

author:["Author"]

**Website HTML:**

Really hard to read, no spacing L

**6. The Daily Telegraph**

**Time Spent**: 15 min

**Link:** http://www.telegraph.co.uk/

**API:** No

**Website Meta Data**:

<meta name="description" content="description />

<meta name="keywords" content="keywords" />

<meta name="news\_keywords" content="keywords">

<meta name="parsely-page" content='{"title": "title ","link": "url","image\_url": "main image url","type": "post","post\_id": "id number", "pub\_date": "2014-09-13T20:30:39Z","section": "Section","authors": ["Author"],"tags": ["tags"]}'/>

<meta property="article:tag" content="tags" />

**Website HTML:**

<div id="mainBodyArea" itemprop="articleBody">

**7. The Wall Street Journal**

**Time Spent:** 10 mins

**Link:** http://online.wsj.com/home-page

**API:** No, http://betaweb.dowjones.com/api/#article-guid - searches Wall Street Journal?

**Website Meta Data:**

<meta name="article.origheadline" content="headline" />

<meta name="article.headline" content="headline " />

<meta name="article.summary" content="summary" />

<meta name="article.published" content="2014-09-12T23:23:00.000Z" />

<meta itemprop="datePublished" content="2014-09-12T23:23:00.000Z" />

<meta name="author" content="author" />

<meta name="keywords" content="keywords" />

<meta name="news\_keywords" content="keywords" />

<meta name="description" content="same as summary?" />

**Website HTML:**

Need to pay to see Journal - can't see HTML tags that surround article

**8. USA Today**

**Time Spent:** 25 min

**Link:** http://www.usatoday.com/

**API:** Yes, http://developer.usatoday.com/

**Search Format:**

http://api.usatoday.com/open/articles/topnews/home?count=10&days=0&page=0&encoding=json&api\_key=

<http://developer.usatoday.com/docs/read/articles>

**Response Information:**

RSS/JSON

description

link

pubDate: "Wed, 21 Jul 2010 15:44:00 GMT"

title

<http://developer.usatoday.com/docs/read/articles#response_formats>

**Website Meta Data:**

<meta name="description" itemprop="description" content="description ">

<meta itemprop="datePublished" content="2014-09-13T04:53:26.9230000"/>

<meta itemprop="image" content="main image url"/>

<meta property="og:title" content="title"/>

<meta property="og:url" content="url"/>

**Website HTML Tags:**

<script type="json/data" class="ui-video-data">…"keywords":

<script type="json/pageinfo" class="pageinfo">…"byline": "author"

**9. Los Angeles Times**

**Time Spent:** 15 mins

**Link:** http://www.latimes.com/

**API:** No

**Website Meta Data:**

<meta property="og:title" content="title" data-meta-updatable/>

<meta property="og:description" content="description." data-meta-updatable/>

<meta property="og:url" content="url" data-meta-updatable/>

<meta property="og:image" content="image url" data-meta-updatable/>

<meta itemprop="datePublished" content="2014-09-13T08:59:00-0700" data-meta-updatable/>

**Website HTML Tags:**

<span class="trb\_bylines\_name\_author\_by"> Author

<div itemprop="article" data-role="panelmod\_articleBodyParagraphs"> Article

class="trb\_article\_relatedTopics" Related topics

**10. The Times of India**

**Time Spent:** 15 mins

**Link:** http://timesofindia.indiatimes.com/international-home

**API:** No, but RSS Feed? http://timesofindia.indiatimes.com/rss.cms

**Website Meta Data:**

<meta name="description" content="description.">

<meta name="twitter:url" content="url">

<meta name="twitter:title" content="title - The Times of India">

<meta name="twitter:image" content="image url"><meta name="news\_keywords" content="keywords">

<meta name="keywords" content="keywords">

<meta http-equiv="Last-Modified" content="03:20:43 AM Sunday, September 14, 2014">

**Website HTML Tags:**

id="authortext" Author

<div id="artext1" class="Normal showpage1"><div class="Normal"> Article

**11. The Examiner**

**Time Spent:** 20 min

**Link:** http://www.sfexaminer.com/

**API:** No

**Website Meta Data:**

<meta name="title" content="title" />

<meta name="description" content="description" />

<meta name="keywords" content="keywords" />

<meta property=”og;url” content=“url” />

<meta property=”og;image” content = “imgurl” />

**Website HTML Tags:**

class=”l1 byline” Author

class=”date” Published date

id=”storyBody” Article

**12. The Independent**

**Time Spent:** 10 mins

**Link:** http://www.independent.co.uk/

**API:** No

**Website Meta Data:**

<meta property="og:title" content="title" />

<meta property="og:image" content="img url"/>

<meta property="og:description" content="description." />

<meta name="description" content="description" />

<meta property="article:author" content="url to author’s page" />

<meta property="article:published\_time" content="2014-09-14" />

<meta property="article:section" content="News"/>

<meta property="og:url" content="article url" />

**Website HTML Tags:**

<div class="body "> Article

**13. Daily News**

**Time Spent:** 10 mins

**Link:** http://www.nydailynews.com/

**API:** No

**Website Meta Data:**

<meta name="news\_keywords" content="keywords"/>

<meta id="og\_title" property="og:title" content="title"/>

<meta id="og\_image" property="og:image" content="img url”>

<meta name="description" content="description"/>

<meta name="nydn\_byline" content="Author"/>

**Website HTML Tags:**

<div id="a-date-published" content="2014-09-13T16:42:42"> Date published

**14. El Mundo**

**Link:** http://www.elmundo.es/america.html

Couldn’t find an english version. There is a Spain version and an American (in Spanish) version.

**15. Financial Times**

**Time Spent:** 15 mins

**Link:** http://www.ft.com/home/us

**API:** No

**Website Meta Data:**

<meta name="description" content="description" />

<meta name="keywords" content="keywords" />

<meta property="og:title" content="title"/>

<meta property="og:url" content="article url"/>

<meta property="og:image" content="image url"/>

**Website HTML Tags:**

class="byline” Author

<p class="lastUpdated" id="publicationDate"> <span class="time"> date published

class="fullstory fullstoryBody" -> id="storyContent” Article body

**16. Bild**

**Link:** http://www.bild.de/

Could not find English version. (Original in German)

**17. El Pais (in English)**

**Time Spent:**  15 mins

**Link:** http://elpais.com/elpais/inenglish.html

**API:** No

**Website Meta Data:**

<meta name="DC.date.issue" content="2014-09-12" />

<meta name="DC.title" lang="en" content="Title" />

<meta name="DC.description" lang="en" content="Description" />

<meta name="DC.subject" lang="en" content="keywords" />

<meta property="og:url" content="Article url"/>

<meta property="og:image" content="image url"/>

**Website HTML Tags:**

rel="author" Author Link

id="cuerpo\_noticia" class="cuerpo\_noticia" Article Body

**18. Chicago Tribune**

**Time Spent:**  15 mins

**Link:** http://www.chicagotribune.com/

**API**: No, but there is an external crime API: <http://blog.apps.chicagotribune.com/2013/03/22/announcing-the-chicago-crime-api/>

**Website Meta Data:**

<meta name="Description" content="description" />

<meta property="og:title" content="title" data-meta-updatable/>

<meta property="og:url" content="url" data-meta-updatable/>

<meta itemprop="datePublished" content="2014-09-13T18:25:00-0700" data-meta-updatable/>

**Website HTML Tags:**

itemprop="author" class="trb\_bylines\_name\_author"><span class="trb\_bylines\_name\_author\_by" Author

<section class="trb\_mainContent" data-role="panelmod\_articleBody"> Article

**19. The Sydney Morning Herald**

**Time Spent:**  15 mins

**Link:** http://www.smh.com.au/

**API:** No

**Website Meta Data:**

<meta name="description" content="An eight-year-old girl flung to her death from a ride at the Royal Adelaide Show was on holiday from Malaysia with her family." />

<meta name="keywords" content="keywords; might be empty" />

<meta name="news\_keywords" content="keywords" />

<meta property="og:title" content="title"/>

<meta property="og:url" content="url"/>

<meta property="og:image" content="image url"/>

**Website Javascript:**

var digitalData -> page -> page information -> author

var digitalData -> page -> page information -> issueDate

**Website HTML Tags:**

class="articleBody" itemprop="articleBody"

**20. La repubblica**

**Link:** http://www.repubblica.it/

Could not find english version. (Original in Italian.)

There is a style/fashion section in English and one random article by involving the pope in English.

**Google Doc Link:**

<https://docs.google.com/a/columbia.edu/document/d/1QUH9JwII3fBpGLttRGl9_7x61FOHfyTnS-XiQr8Kd4M/edit>