

Séance 3.1: Données digitales

**Collecte par interfaces de programmation d'applications
(API)**

Visseho Adjiwanou, PhD.

UQAM

06 June 2024

Plan de présentation

- Définition
 - Application programming interface (API)
 - Type de fichiers API
- Type de fichiers
- Object de R: liste
- Exemples d'utilisation de API pour la collecte de données
 - API de crossref
 - API de twitter
 - API de DHS
 - API de IPUMS

API

Définition

- Les interfaces de programmation d'applications, ou API, sont devenues l'un des moyens les plus importants d'accéder aux données et de les transférer en ligne.

Définition

- Les interfaces de programmation d'applications, ou API, sont devenues l'un des moyens les plus importants d'accéder aux données et de les transférer en ligne.
- De plus en plus, les API peuvent même analyser vos données.

Définition

- Les interfaces de programmation d'applications, ou API, sont devenues l'un des moyens les plus importants d'accéder aux données et de les transférer en ligne.
- De plus en plus, les API peuvent même analyser vos données.
- Comparées aux données de scraping, qui sont souvent illégales, logiquement difficiles (ou les deux), les API sont un outil utile pour créer des requêtes personnalisées de données de manière bien structurée et considérablement plus facile à utiliser que les données HTML ou XML décrites précédemment.

Qu'est-ce qu'une interface de programmation d'application (applicative)?

- Les API sont des outils permettant de créer des applications ou d'autres formes de logiciels permettant aux utilisateurs d'accéder à certaines parties de bases de données volumineuses.

Qu'est-ce qu'une interface de programmation d'application (applicative)?

- Les API sont des outils permettant de créer des applications ou d'autres formes de logiciels permettant aux utilisateurs d'accéder à certaines parties de bases de données volumineuses.
- Les développeurs de logiciels peuvent combiner ces outils de différentes manières, ou avec des outils d'autres API, afin de générer des outils encore plus utiles.

Qu'est-ce qu'une interface de programmation d'application (applicative)?

- Les API sont des outils permettant de créer des applications ou d'autres formes de logiciels permettant aux utilisateurs d'accéder à certaines parties de bases de données volumineuses.
- Les développeurs de logiciels peuvent combiner ces outils de différentes manières, ou avec des outils d'autres API, afin de générer des outils encore plus utiles.
- La plupart d'entre nous utilisons ces applications chaque jour.

Qu'est-ce qu'une interface de programmation d'application (applicative)?

- Par exemple, si vous installez l'application Spotify sur votre page Facebook pour partager de la musique avec vos amis, cette application extrait des données de l'API de Spotify, puis les publie sur votre page Facebook en communiquant avec l'API de Facebook.

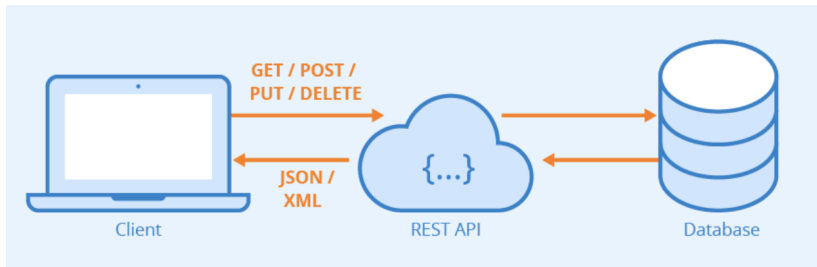
Qu'est-ce qu'une interface de programmation d'application (applicative)?

- Par exemple, si vous installez l'application Spotify sur votre page Facebook pour partager de la musique avec vos amis, cette application extrait des données de l'API de Spotify, puis les publie sur votre page Facebook en communiquant avec l'API de Facebook.
- Il existe d'innombrables exemples de ce type sur Internet à l'heure actuelle.

Qu'est-ce qu'une interface de programmation d'application (applicative)?

- Par exemple, si vous installez l'application Spotify sur votre page Facebook pour partager de la musique avec vos amis, cette application extrait des données de l'API de Spotify, puis les publie sur votre page Facebook en communiquant avec l'API de Facebook.
- Il existe d'innombrables exemples de ce type sur Internet à l'heure actuelle.
- **En résumé, Une API permet à deux applications de communiquer entre elles.**

Comment fonctionne une API?



Comment fonctionne une API? Exemple avec crossnef

- Supposer que vous voulez collecter les articles publiés sur la Covid-19

Comment fonctionne une API? Exemple avec crossnef

- Supposer que vous voulez collecter les articles publiés sur la Covid-19
- Pour cela, vous devez envoyer une requête au site, c'est-à-dire:

Comment fonctionne une API? Exemple avec crossnef

- Supposer que vous voulez collecter les articles publiés sur la Covid-19
- Pour cela, vous devez envoyer une requête au site, c'est-à-dire:
- indiquer le site

Comment fonctionne une API? Exemple avec crossnef

- Supposer que vous voulez collecter les articles publiés sur la Covid-19
- Pour cela, vous devez envoyer une requête au site, c'est-à-dire:
- indiquer le site
- indiquer la requête

Comment fonctionne une API? Exemple avec crossref

- Supposer que vous voulez collecter les articles publiés sur la Covid-19
- Pour cela, vous devez envoyer une requête au site, c'est-à-dire:
- indiquer le site
- indiquer la requête
- Pour plus d'information:
<https://www.crossref.org/education/retrieve-metadata/rest-api/a-non-technical-introduction-to-our-api/>

Comment fonctionne une API? Exemple avec crossref

- Supposer que vous voulez collecter les articles publiés sur la Covid-19
- Pour cela, vous devez envoyer une requête au site, c'est-à-dire:
- indiquer le site
- indiquer la requête
- Pour plus d'information:
<https://www.crossref.org/education/retrieve-metadata/rest-api/a-non-technical-introduction-to-our-api/>
- <https://github.com/ropensci/rcrossref>

Comment fonctionne une API? Exemple avec crossref

- Supposer que vous voulez collecter les articles publiés sur la Covid-19
- Pour cela, vous devez envoyer une requête au site, c'est-à-dire:
- indiquer le site
- indiquer la requête
- Pour plus d'information:
<https://www.crossref.org/education/retrieve-metadata/rest-api/a-non-technical-introduction-to-our-api/>
- <https://github.com/ropensci/rcrossref>
- <https://ciakovx.github.io/rcrossref.html>

Comment fonctionne une API? Exemple avec crossref

- Supposer que vous voulez collecter les articles publiés sur la Covid-19
- Pour cela, vous devez envoyer une requête au site, c'est-à-dire:
- indiquer le site
- indiquer la requête
- Pour plus d'information:
 - <https://www.crossref.org/education/retrieve-metadata/rest-api/a-non-technical-introduction-to-our-api/>
- <https://github.com/ropensci/rcrossref>
- <https://ciakovx.github.io/rcrossref.html>
- <https://www.crossref.org/education/retrieve-metadata/rest-api/a-non-technical-introduction-to-our-api/>

Comment fonctionne une API? Exemple avec crossref

- Cette requête: <https://api.crossref.org/works?query=covid-19>

Comment fonctionne une API? Exemple avec crossref

- Cette requête: <https://api.crossref.org/works?query=covid-19>
- connecte avec le site crossref pour collecter les informations sur les recherches sur la Covid-19

Comment fonctionne une API? Exemple avec crossref

- Cette requête: <https://api.crossref.org/works?query=covid-19>
- connecte avec le site crossref pour collecter les informations sur les recherches sur la Covid-19
- Regardons un des résultats:

Comment fonctionne une API? Exemple avec crossref

- Cette requête: <https://api.crossref.org/works?query=covid-19>
- connecte avec le site crossref pour collecter les informations sur les recherches sur la Covid-19
- Regardons un des résultats:
- <https://api.crossref.org/works/10.1055/a-1229-5048>

```
{“status”:“ok”,“message-type”:“work”,“message-version”:“1.0.0”,“message”:{“indexed”:{“date-parts”:[[2020,11,3]],“date-time”:“2020-11-03T00:12:53Z”,“timestamp”:1604362373737},“reference-count”:0,“publisher”:“Georg Thieme Verlag KG”,“issue”:“04”,“content-domain”:{“domain”:[],“crossmark-restriction”:false},“short-container-title”:[“Endo-Praxis”],“published-print”:{“date-parts”:[[2020,11]]},“DOI”:“10.1055/a-1229-5048”,“type”:“journal-article”,“created”:{“date-parts”:[[2020,11,2]],“date-time”:“2020-11-02T23:41:05Z”,“timestamp”:1604360465000},“page”:“173-173”,“source”:“Crossref”,“is-referenced-by-count”:0,“title”:[“COVID-19 COVID-19 COVID-19 COVID-19 COVID-19 COVID-
```

Type de fichiers

Type de fichiers

- Les requêtes renvoient à des types de fichiers que nous ne traitons pas souvent en sciences sociales.

Type de fichiers

- Les requêtes renvoient à des types de fichiers que nous ne traitons pas souvent en sciences sociales.
- Les principaux sont:

Type de fichiers

- Les requêtes renvoient à des types de fichiers que nous ne traitons pas souvent en sciences sociales.
- Les principaux sont:
- XML: XML signifie eXtensible Markup Language

Type de fichiers

- Les requêtes renvoient à des types de fichiers que nous ne traitons pas souvent en sciences sociales.
- Les principaux sont:
- XML: XML signifie eXtensible Markup Language
- html

Type de fichiers

- Les requêtes renvoient à des types de fichiers que nous ne traitons pas souvent en sciences sociales.
- Les principaux sont:
 - XML: XML signifie eXtensible Markup Language
 - html
 - json

Fichier xml

- XML est un langage de balisage extensible.

Fichier xml

- XML est un langage de balisage extensible.
- C'est un langage : Cela signifie que ce format de fichier est conçu pour transmettre des informations.

Fichier xml

- XML est un langage de balisage extensible.
- C'est un langage : Cela signifie que ce format de fichier est conçu pour transmettre des informations.
- On utilise des balises pour délimiter les informations, par exemple:

Fichier xml

- XML est un langage de balisage extensible.
- C'est un langage : Cela signifie que ce format de fichier est conçu pour transmettre des informations.
- On utilise des balises pour délimiter les informations, par exemple:
- XML permet donc - en théorie - de stocker des informations indépendamment de leur représentation.

Fichier xml

```
<BIBLIOTHEQUE>

  <ROMAN>
    <TITRE>Imajica</TITRE>
    <AUTEUR>Clive Barker</AUTEUR>
    <PRIX>6</PRIX>
  </ROMAN>

  <ROMAN>
    <TITRE>Dune</TITRE>
    <AUTEUR>Frank Herbert</AUTEUR>
    <PRIX>7</PRIX>
  </ROMAN>

  <MAGAZINE>
    <TITRE>Science et Vie</TITRE>
    <DATEPARUTION>2005-02-01</DATEPARUTION>
  </MAGAZINE>

  <ROMAN>
    <TITRE>Christine</TITRE>
    <AUTEUR>Stephen King</AUTEUR>
    <PRIX>5</PRIX>
  </ROMAN>
```

Fichier xml: comment accéder aux données

- XQuery et XPath permettent d'extraire les informations qui vous intéressent d'un document XML

Fichier xml: comment accéder aux données

- XQuery et XPath permettent d'extraire les informations qui vous intéressent d'un document XML
- Avec un package comme **XML** (voir exemple SOC8655_S10_xml.RMD)

Fichier xml: comment accéder aux données

- XQuery et XPath permettent d'extraire les informations qui vous intéressent d'un document XML
- Avec un package comme **XML** (voir exemple SOC8655_S10_xml.RMD)
- https://www.tutorialspoint.com/r/r_xml_files.htm

Fichiers html

Fichier JSON : Syntaxe

- `{...}` : les accolades définissent un objet.

Fichier JSON : Syntaxe

- `{...}` : les accolades définissent un objet.
- `"language": "Java"` : Les guillemets (double-quotes) et les double-points définissent un couple clé/valeur (on parle de membre).

Fichier JSON : Syntaxe

- `{...}` : les accolades définissent un objet.
- `"language": "Java"` : Les guillemets (double-quotes) et les double-points définissent un couple clé/valeur (on parle de membre).
- `[...]` : Les crochets définissent un tableau (ou array en anglais).

Fichier JSON : Syntaxe

- `{...}` : les accolades définissent un objet.
- `"language": "Java"` : Les guillemets (double-quotes) et les double-points définissent un couple clé/valeur (on parle de membre).
- `[...]` : Les crochets définissent un tableau (ou array en anglais).
- `,` : Les virgules permettent de séparer les membres d'un tableau ou, comme ici, d'un objet

Fichier JSON : Syntaxe

- `{...}` : les accolades définissent un objet.
- `"language":"Java"` : Les guillemets (double-quotes) et les double-points définissent un couple clé/valeur (on parle de membre).
- `[...]` : Les crochets définissent un tableau (ou array en anglais).
- `,` : Les virgules permettent de séparer les membres d'un tableau ou, comme ici, d'un objet
- Exemple: `{"id":1, "language":"json", "author":"Douglas Crockford"}`

Fichier JSON

```
{  
  "titre_album": "Abacab",  
  "groupe": "Genesis",  
  "annee": 1981,  
  "genre": "Rock"  
}
```

- Pour en savoir un peu plus:

<https://sebsauvage.net/comprendre/xml/>

Fichier JSON

```
{
  - data: [
    - {
      acronym: null,
      - badges: [
        - {
          kind: "public-service"
        },
        - {
          kind: "certified"
        }
      ],
      created_at: "2014-04-17T18:21:57.325000",
      deleted: null,
      description: "
        Aux côtés de son cabinet, l'administration du Premier ministre comprend de nombreux services qui l'assistent et prennent part à l'élaboration de la politique du
        Gouvernement. Sont rattachés au Premier ministre les organismes chargés de missions de coordination interministérielle qui ne peuvent être attribuées à un seul
        ministère.",
      id: "534fffa5a3a7292c64a7809e",
      last_modified: "2018-09-03T20:47:50.820000",
      logo: "https://static.data.gouv.fr/avatars/08/98902e29244685862bccd3198cef7b-original.png",
      logo_thumbnail: "https://static.data.gouv.fr/avatars/08/98902e29244685862bccd3198cef7b-100.png",
      - members: [
        - {
          role: "editor",
          - user: {
            avatar: "https://static.data.gouv.fr/avatars/aa/733490184c4c2684e793ea074d57d6-original.jpg",
            avatar_thumbnail: "https://static.data.gouv.fr/avatars/aa/733490184c4c2684e793ea074d57d6-100.jpg",
            class: "User",
            first_name: "Anne-Gaelle",
            id: "534fff3ca3a7292c64a773ab",
            last_name: "Javelle",
            page: "https://www.data.gouv.fr/fr/users/anne-gaelle-javelle/",
            slug: "anne-gaelle-javelle",
            uri: "https://www.data.gouv.fr/api/1/users/anne-gaelle-javelle/"
          }
        }
      ]
    }
  ]
}
```


Fichier JSON: Visualisation et validation

- Quand on débute dans un langage, le fait d'avoir à disposition des outils permettant de visualiser et de valider le code est d'une grande aide.

Fichier JSON: Visualisation et validation

- Quand on débute dans un langage, le fait d'avoir à disposition des outils permettant de visualiser et de valider le code est d'une grande aide.
- Là encore json est doté de tout ce dont vous avez besoin.

Fichier JSON: Visualisation et validation

- Quand on débute dans un langage, le fait d'avoir à disposition des outils permettant de visualiser et de valider le code est d'une grande aide.
- Là encore json est doté de tout ce dont vous avez besoin.
- Plusieurs visualiseurs/validateurs sont disponibles en ligne.

Fichier JSON: Visualisation et validation

- Quand on débute dans un langage, le fait d'avoir à disposition des outils permettant de visualiser et de valider le code est d'une grande aide.
- Là encore json est doté de tout ce dont vous avez besoin.
- Plusieurs visualiseurs/validateurs sont disponibles en ligne.
- Un des visualiseurs les plus connu est jsonviewer.stack.hu.

Fichier JSON: Visualisation et validation

- Quand on débute dans un langage, le fait d'avoir à disposition des outils permettant de visualiser et de valider le code est d'une grande aide.
- Là encore json est doté de tout ce dont vous avez besoin.
- Plusieurs visualiseurs/validateurs sont disponibles en ligne.
- Un des visualiseurs les plus connu est jsonviewer.stack.hu.
- Le grand avantage de cet outil est qu'il intègre toutes les fonctionnalités utiles pour l'analyse d'un fragment de code .json :

Fichier JSON: Visualisation et validation

- Quand on débute dans un langage, le fait d'avoir à disposition des outils permettant de visualiser et de valider le code est d'une grande aide.
- Là encore json est doté de tout ce dont vous avez besoin.
- Plusieurs visualiseurs/validateurs sont disponibles en ligne.
- Un des visualiseurs les plus connu est jsonviewer.stack.hu.
- Le grand avantage de cet outil est qu'il intègre toutes les fonctionnalités utiles pour l'analyse d'un fragment de code .json :
- un validateur (très basique puisqu'il ne fait qu'indiquer si le code contient des erreurs),

Fichier JSON: Visualisation et validation

- Quand on débute dans un langage, le fait d'avoir à disposition des outils permettant de visualiser et de valider le code est d'une grande aide.
- Là encore json est doté de tout ce dont vous avez besoin.
- Plusieurs visualiseurs/validateurs sont disponibles en ligne.
- Un des visualiseurs les plus connu est jsonviewer.stack.hu.
- Le grand avantage de cet outil est qu'il intègre toutes les fonctionnalités utiles pour l'analyse d'un fragment de code .json :
 - un validateur (très basique puisqu'il ne fait qu'indiquer si le code contient des erreurs),
 - un visualiseur,

Fichier JSON: Visualisation et validation

- Quand on débute dans un langage, le fait d'avoir à disposition des outils permettant de visualiser et de valider le code est d'une grande aide.
- Là encore json est doté de tout ce dont vous avez besoin.
- Plusieurs visualiseurs/validateurs sont disponibles en ligne.
- Un des visualiseurs les plus connu est jsonviewer.stack.hu.
- Le grand avantage de cet outil est qu'il intègre toutes les fonctionnalités utiles pour l'analyse d'un fragment de code .json :
 - un validateur (très basique puisqu'il ne fait qu'indiquer si le code contient des erreurs),
 - un visualiseur,
 - un minificateur.

Fichier JSON: Visualisation et validation

- Le principe est simple, vous collez le code dans la fenêtre “Text” puis vous cliquez sur l’onglet “Viewer”. S’il n’y a pas d’erreur dans votre code, l’arborescence de votre **.json** sera affichée.

Fichier JSON: Visualisation et validation

- Le principe est simple, vous collez le code dans la fenêtre “Text” puis vous cliquez sur l’onglet “Viewer”. S’il n’y a pas d’erreur dans votre code, l’arborescence de votre **.json** sera affichée.
- Copie d’écran pour le dernier exemple traité :

Fichier JSON: Visualisation et validation

Viewer

Text

Paste

Copy

Format

Remove white space

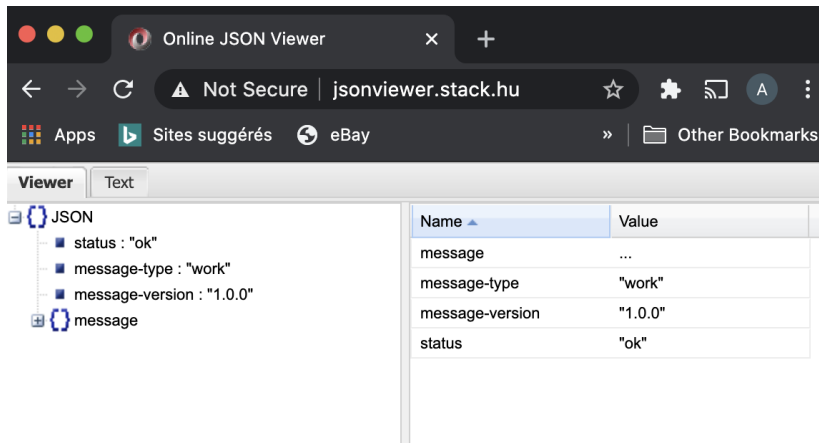
Clear

Load JSON data

About

```
{
  "status": "ok",
  "message-type": "work",
  "message-version": "1.0.0",
  "message": {
    "indexed": {
      "date-parts": [[2020, 11, 3]],
      "date-time": "2020-11-03T00:12:53Z",
      "timestamp": 1604362373737,
      "reference-count": 0,
      "publisher": "Georg Thieme Verlag KG",
      "issue": "04",
      "content-domain": {
        "domain": [],
        "crossmark-restriction": false
      },
      "short-container-title": ["Endo-Praxis"],
      "published-print": {
        "date-parts": [[2020, 11]]
      },
      "DOI": "10.1055/a-1229-5048",
      "type": "journal-article",
      "created": {
        "date-parts": [[2020, 11, 2]],
        "date-time": "2020-11-02T23:41:05Z",
        "timestamp": 1604360465000
      },
      "page": "173-173",
      "source": "Crossref",
      "is-referenced-by-count": 0,
      "title": ["COVID-19 COVID-19 COVID-19 COVID-19 COVID-19"],
      "prefix": "10.1055",
      "volume": "36",
      "author": [
        {
          "given": "Rainer",
          "family": "Duchmann",
          "sequence": "additional",
          "affiliation": [
            {
              "name": "Medizinischen Klinik, Hospital zum heiligen Geist GmbH, Frankfurt/Main"
            }
          ]
        }
      ],
      "member": "194",
      "published-online": {
        "date-parts": [[2020, 11, 2]]
      },
      "container-title": ["Endo-Praxis"],
      "original-title": [],
      "language": "de",
      "link": {
        "URL": "http://www.thieme-connect.de/products/ejournals/pdf/10.1055/a-1229-5048.pdf"
      },
      "content-type": "unspecified",
      "content-version": "vor",
      "intended-application": "similarity-checking"
    },
    "deposited": {
      "date-parts": [[2020, 11, 2]],
      "date-time": "2020-11-02T23:41:07Z",
      "timestamp": 1604360467000,
      "score": 1.0,
      "subtitle": [],
      "short-title": [],
      "issued": {
        "date-parts": [[2020, 11]]
      },
      "references-count": 0,
      "journal-issue": {
        "published-online": {
          "date-parts": [[2020, 11, 2]]
        },
        "published-print": {
          "date-parts": [[2020, 11]]
        },
        "issue": "04"
      },
      "URL": "http://dx.doi.org/10.1055/a-1229-5048",
      "relation": {},
      "ISSN": ["0177-4077", "1611-6429"],
      "issn-type": [
        {
          "value": "0177-4077",
          "type": "print"
        },
        {
          "value": "1611-6429",
          "type": "electronic"
        }
      ]
    }
  }
}
```

Fichier JSON: Visualisation et validation



The screenshot shows a web browser window titled "Online JSON Viewer" at the URL `jsonviewer.stack.hu`. The browser's address bar indicates "Not Secure". The page has a dark theme. Below the browser window, the JSON viewer interface is visible. It has a "Viewer" tab selected, and a "Text" tab is also present. On the left, a tree view shows the JSON structure: a root object with fields `status`, `message-type`, `message-version`, and `message`. The `message` field is expanded, showing a nested object. On the right, a table displays the key-value pairs of the JSON object.

Name	Value
message	...
message-type	"work"
message-version	"1.0.0"
status	"ok"

Fichier JSON

- Pour plus d'information :
<https://www.alsacreations.com/article/lire/1675-json-stockage-leger-pratique-donnees-multitypes.html>

wrapper API

Wrapper API

- Spécifier la syntaxe d'un appel API nécessite de maîtriser la documentation de chaque API, ce qui peut prendre un temps très long.

Wrapper API

- Spécifier la syntaxe d'un appel API nécessite de maîtriser la documentation de chaque API, ce qui peut prendre un temps très long.
- Cet outil essaie de vous aider, mais il ne fonctionne pas parfaitement et la construction du bon appel peut nécessiter beaucoup de temps pour lire la documentation de l'API.

Wrapper API

- Spécifier la syntaxe d'un appel API nécessite de maîtriser la documentation de chaque API, ce qui peut prendre un temps très long.
- Cet outil essaie de vous aider, mais il ne fonctionne pas parfaitement et la construction du bon appel peut nécessiter beaucoup de temps pour lire la documentation de l'API.
- Plus important encore, comme vous le verrez bientôt, un certain nombre de packages R contiennent des fonctions permettant de faire des appels d'API, ce qui vous fera économiser du temps et de l'énergie nécessaires pour apprendre la syntaxe propre à chaque API.

Limitation de taux

- Avant de passer d'autres appels aux API, nous devons nous familiariser avec un concept important appelé «Limitation du débit».

Limitation de taux

- Avant de passer d'autres appels aux API, nous devons nous familiariser avec un concept important appelé «Limitation du débit».
- Les informations d'identification de la section précédente non seulement définissent le type d'informations auquel nous sommes autorisés à accéder, mais également la fréquence à laquelle nous sommes autorisés. Celles-ci sont appelées «limites de débit».

Limitation de taux

- Avant de passer d'autres appels aux API, nous devons nous familiariser avec un concept important appelé «Limitation du débit».
- Les informations d'identification de la section précédente non seulement définissent le type d'informations auquel nous sommes autorisés à accéder, mais également la fréquence à laquelle nous sommes autorisés. Celles-ci sont appelées «limites de débit».
- Si nous faisons trop de demandes de données dans un délai trop court, une API nous bloque temporairement des données collectées pendant une période pouvant aller de 15 minutes à 24 heures ou plus en fonction de l'API.

Limitation de taux

- Avant de passer d'autres appels aux API, nous devons nous familiariser avec un concept important appelé «Limitation du débit».
- Les informations d'identification de la section précédente non seulement définissent le type d'informations auquel nous sommes autorisés à accéder, mais également la fréquence à laquelle nous sommes autorisés. Celles-ci sont appelées «limites de débit».
- Si nous faisons trop de demandes de données dans un délai trop court, une API nous bloque temporairement des données collectées pendant une période pouvant aller de 15 minutes à 24 heures ou plus en fonction de l'API.
- La limitation de débit est nécessaire pour que les API ne soient pas submergées par un trop grand nombre de demandes

Limitation de taux

- La limitation des débits permet également aux grandes entreprises telles que Google, Facebook ou Twitter d'empêcher les développeurs de collecter de grandes quantités de données susceptibles de compromettre la confidentialité de leurs utilisateurs ou de menacer leur modèle commercial (car les données ont une si grande valeur dans l'économie actuelle).

Limitation de taux

- La limitation des débits permet également aux grandes entreprises telles que Google, Facebook ou Twitter d'empêcher les développeurs de collecter de grandes quantités de données susceptibles de compromettre la confidentialité de leurs utilisateurs ou de menacer leur modèle commercial (car les données ont une si grande valeur dans l'économie actuelle).
- Le moment exact de la limitation de débit n'est pas toujours public, car connaître ces incréments de temps pourrait permettre aux développeurs de «jouer» au système et de faire des demandes rapides dès que la limitation de débit est terminée.

Limitation de taux

- Cependant, certaines API vous permettent de passer un appel ou une requête à l'API afin de connaître le nombre de demandes supplémentaires que vous pouvez effectuer au cours d'une période donnée avant que votre taux ne soit limité.

Un exemple avec l'API de Twitter

- 1 Obtenir les information d'identification : voir document partager sur Slack

Un exemple avec l'API de Twitter

- 1 Obtenir les information d'identification : voir document partager sur Slack
- 2 Utiliser le package **rtweet** pour interagir avec le site

Twitter API wrapper

- **rtweet** est un package de R qui permet de travailler avec le API de Twitter.

Twitter API wrapper

- **rtweet** est un package de R qui permet de travailler avec le API de Twitter.
- Si vous n'avez pas ce package, vous devez apprendre à interagir directement avec le API en lisant toute la documentation le concernant:
<https://developer.twitter.com/en/docs/twitter-api>

Twitter API wrapper

- **rtweet** est un package de R qui permet de travailler avec le API de Twitter.
- Si vous n'avez pas ce package, vous devez apprendre à interagir directement avec le API en lisant toute la documentation le concernant:
<https://developer.twitter.com/en/docs/twitter-api>
- Heureusement, pour la plupart des sites aujourd'hui, il y a un package qui interagit avec le site.

rtweet

- rtweet permet de collecter : -1. des anciens **tweet** (jusqu'à 18000 - non retweet - tweets)

rtweet

- rtweet permet de collecter : -1. des anciens **tweet** (jusqu'à 18000 - non retweet - tweets)
- **search_tweets** avec ses options -2. et des tweet qui se passent en direct (échantillon aléatoire (environ 1%) du flux en direct de tous les tweets)

rtweet

- rtweet permet de collecter : -1. des anciens **tweet** (jusqu'à 18000 - non retweet - tweets)
- **search_tweets** avec ses options -2. et des tweet qui se passent en direct (échantillon aléatoire (environ 1%) du flux en direct de tous les tweets)
- **stream_tweets** et ses options -3. et un ensemble d'autres informations

rtweet : les autres principales fonctions de rtweet

- **get_friends** : récupérez une liste de tous les comptes **suivis** par un utilisateur.

rtweet : les autres principales fonctions de rtweet

- **get_friends** : récupérez une liste de tous les comptes **suivis** par un utilisateur.
- **get_followers** : récupérez une liste des comptes **qui suivent** un utilisateur.

rtweet : les autres principales fonctions de rtweet

- **get_friends** : récupérez une liste de tous les comptes **suivis** par un utilisateur.
- **get_followers** : récupérez une liste des comptes **qui suivent** un utilisateur.
- **get_timelines** : recevez les 3200 tweets les plus récents de cnn, BBC World et fox news.

rtweet : les autres principales fonctions de rtweet

- **get_friends** : récupérez une liste de tous les comptes **suivis** par un utilisateur.
- **get_followers** : récupérez une liste des comptes **qui suivent** un utilisateur.
- **get_timelines** : recevez les 3200 tweets les plus récents de cnn, BBC World et fox news.
- **get_favorites** : obtenez les 3000 derniers statuts favoris de quelqu'un.

rtweet : les autres principales fonctions de rtweet

- **get_friends** : récupérez une liste de tous les comptes **suivis** par un utilisateur.
- **get_followers** : récupérez une liste des comptes **qui suivent** un utilisateur.
- **get_timelines** : recevez les 3200 tweets les plus récents de cnn, BBC World et fox news.
- **get_favorites** : obtenez les 3000 derniers statuts favoris de quelqu'un.
- **search_users** : recherchez 1000 utilisateurs avec un spécifique **hashtag** dans leur profil.

rtweet : les autres principales fonctions de rtweet

- **get_friends** : récupérez une liste de tous les comptes **suivis** par un utilisateur.
- **get_followers** : récupérez une liste des comptes **qui suivent** un utilisateur.
- **get_timelines** : recevez les 3200 tweets les plus récents de cnn, BBC World et fox news.
- **get_favorites** : obtenez les 3000 derniers statuts favoris de quelqu'un.
- **search_users** : recherchez 1000 utilisateurs avec un spécifique **hashtag** dans leur profil.
- **get_trends** : découvrez les tendances actuelles dans un lieu.

Questions d'éthique

Conclusion

Il est à présent espéré que les API sont une ressource inestimable pour la collecte de données à partir d'Internet. Dans le même temps, il peut également être clair que le processus d'obtention d'informations d'identification, d'éviter les limitations de taux et de comprendre le jargon unique employé par les concepteurs de chaque API peut signifier de nombreuses heures à passer au crible la documentation d'une API, en particulier là où Les packages R ne fonctionnent pas correctement pour l'interfaçage avec l'API en question. Si vous devez développer votre propre code personnalisé pour fonctionner avec une API - ou si vous avez besoin d'informations qui ne peuvent pas être obtenues à l'aide de fonctions dans un package R, il peut être utile de parcourir le code source des fonctions R décrites ci-dessus. afin de voir où ils passent le langage de requête API nécessaire pour produire les résultats avec lesquels nous avons travaillé ci-dessus.

Conclusion

Une liste des API d'intérêt Il existe de nombreuses bases de données décrivant les API populaires sur le Web, y compris le site Web programmable susmentionné, mais également une variété de listes générées par des utilisateurs ou par des utilisateurs:

<https://www.programmableweb.com/>

<https://github.com/toddmotto/public-apis>

<https://apilist.fun/>

Le site R OpenSci contient également une liste de packages R fonctionnant avec des API:

<https://ropensci.org/packages/>

Conclusion

- Vous avez d'autres applications pour collecter les données pour vous comme Facepager
 - <https://github.com/strohne/Facepager>