Week 9 Profile Information

Group 22

May 5, 2021

Outline

1. Profile Page
   - Why need it?
   - How to get it?

2. Issues
   - Same companies don't have a profile page

3. Future Plan

Outline

Why need it?

- to security the company industry
- to identify the country/address of the companies
- example:

Apple Inc. (AAPL)
NasdaqGS - NasdaqGS Real Time Price. Currency in USD

☆ Add to watchlist

127.19 -5.35 (-4.04%)
As of 11:44AM EDT. Market open.

Summary Chart Conversations Statistics Historical Data **Profile** Financials Analysis Options Holders Sustainability

Apple Inc.

One Apple Park Way
Cupertino, CA 95014
United States
408-996-1010
http://www.apple.com

Sector(s): **Technology**
Industry: **Consumer Electronics**
Full Time Employees: **100,000**

Outline

How to get it?

- Download the profile page
  - parse the industry table `txt` to get the company names
  - download their profile pages and save as `txt`
- Parse the profile page
  - fullname; address; section & industry; official website; profile page.
  - example:
    ```
    {"CTVA": {
    "name": "Corteva, Inc.",
    "location": ["Chestnut Run Plaza 735", "PO Box 80735", "United States"],
    "website": "http://www.corteva.com",
    "section": "Basic Materials",
    "industry": "Agricultural Inputs",
    "profile": "https://finance.yahoo.com/quote/CTVA/profile?p=CTVA"}
    }
    ```

- Save as `jsonl` for each industry

Outline

Same companies don't have a profile page

- example:

  `https://finance.yahoo.com/quote/CLV.MU/profile?p=CLV.MU`

TRIP.COM.GROUP SP.ADR 1 (CLV.MU)
Munich - Munich Delayed Price. Currency in EUR

☆ Add to watchlist

31.90 -0.90 (-2.74%)
As of 8:01AM CEST. Market open.

Summary   Chart   Statistics   Historical Data   **Profile**   Financials   Analysis   Options   Holders   Sustainability

Potential solutions

- wait...
- skip this company for now
  - 0/18(0.00%) companies' profile pages inform nothing in agriculture.
  - 103/268(38.43%) companies' profile pages inform nothing in travel.
  - 14/255(5.49%) companies' profile pages inform nothing in energy.

Assemble data together

- gather ROI, stock prices, company profile as a `json`
- example:

  ```
  {"CTVA": {
  "name": "Corteva, Inc.",
  "location": ["Chestnut Run Plaza 735", "PO Box 80735", "United States"],
  "website": "http://www.corteva.com",
  "section": "Basic Materials",
  "industry": "Agricultural Inputs",
  "profile": "https://finance.yahoo.com/quote/CTVA/profile?p=CTVA"},
  "stock prices": pandas.framework,
  "ROI": pandas.framework
  }
  ```

- an industry is a `jsonl`