****

**ProntoNLP Functional API**

ProntoNLP is quickly becoming the favorite natural language processing framework for the investment world. Empowering the entire spectrum of financial professionals from portfolio managers, economists, and risk monitors to researchers, analysts, and quants; ProntoNLP delivers cutting-edge tools allowing users to research and develop complete NLP models for any text-based requirements. Efficiently move from theory to model, giving your strategy and clients total confidence. These data-driven strategies and analytics greatly outperform benchmark measures and can be further customized for maximal value.

Our model pipeline is a hybrid system using the best of our linguistic engine coupled with our deep learning sentiment model. This *Deep Linguistics* hybrid system takes full advantage of precise grammatical structures as well as the robust abstractions provided by deep learning models to deliver unparalleled performance. In a single system, users can create and then apply customized NLP models for full information extraction, event classification, sentiment analysis, and data structuring. By using the functional API, users can easily access three main endpoints in a programmatic way, providing for an even smoother integration into existing workflows. These are the main functions: Signal Generation, Extraction Analytics, Process User Documents.

Best practice recommends first creating a virtual environment and then installing the ProntoSDK via Git.

**Signal Generation**

The ProntoNLP Signal is a CSV file containing the related event and sentiment counts aggregated at the document level. In that way, users can digest document level metrics derived from the NLP models. Additionally, overall sentiment and theme-related scores are included. The signal file is typically used as a direct input into investment making decisions.

To generate the Signal CSV use the following command line prompt:

pronto\_nlp fief generate\_signal\_csv -u USER -p PASSWORD -r RULESET -d DB [-s STARTDATE] [-e ENDDATE] [-g TAGS] outputCSV

* USER: your ProntoNLP username (the same one used to access the web platform)
* PASSWORD: your ProntoNLP password (the same one used to access the web platform)
* RULESET: name of a ruleset that exists in your account (e.g. Alpha, ESG, etc.)
* DB: name of a corpus (e.g. SnP\_ProntoStageDB\_First\_ParseCache.db3)
  + currently only the transcript corpus is relevant to producing signals
* STARTDATE: (optional) limit the signal to start from this date (YYYY-MM-DD)
* ENDDATE: (optional) limit the signal to end at this date (YYYY-MM-DD)
* TAGS: (optional) limit the signal to specific document characteristics (e.g. the Answer section of Earning Calls, companies in a specific sector, only text spoken by the CEO, etc.)
  + For a complete list of applicable tags, select a corpus in the web platform and select filter by tags
* outputCSV: path to store the generated Signal CSV

Example Usage:

pronto\_nlp fief generate\_signal\_csv -u “org:user@example.com” -p “password” -r Alpha -d SnP\_Transcripts\_ParseCache.db3 -s 2018-01-01 -e 2020-01-01 -g "#DocType\_EarningsCalls #Sector\_Utilities" mySignal.csv

This will generate the Signal CSV with the following characteristics:

* Values based on extractions from the Alpha model applied to the Transcript Corpus
* For the date range 2018 to 2020
* Limited to Earning Call transcripts and specifically the Utilities sector
* The signal will be saved to the local directory in a file called mySignal.csv

Example Usage:

pronto\_nlp fief generate\_signal\_csv -u “org:user@example.com” -p “password”-r Alpha -d SnP\_Transcripts\_ParseCache.db3 -g "#DocType\_EarningsCalls #SpeakerType\_Executives\_CEO #SpeakerType\_Executives\_CFO" mySignal.csv

Similar to above, but no date limitation and aggregating events appearing exclusively in what the CEO or CFO said.

**Extraction Analytics**

The ProntoNLP Extraction Analytics is a CSV file containing the actual extractions on an individual level. Each sentence and related event with all of the various properties produced by the NLP model are returned. In that way, users can easily transform raw unstructured text data into organized events and properties. The Extraction Analytics file is typically used to feed a BI tool or as the basis of downstream analysis.

To generate the Extraction CSV use the following command line prompt:

pronto\_nlp fief generate\_find\_matches\_csv -u USER -p PASSWORD -r RULESET -v EVENTS -d DB [-s STARTDATE] [-e ENDDATE] [-g TAGS] [-m] outputCSV

* USER: your ProntoNLP username (the same one used to access the web platform)
* PASSWORD: your ProntoNLP password (the same one used to access the web platform)
* RULESET: name of a ruleset that exists in your account (e.g. Alpha, ESG, etc.)
* EVENTS: Regex specifying events from Ruleset to extract (e.g. ".\*" or "Dividend|Margin")
* DB: name of a corpus (e.g. SnP\_ProntoStageDB\_First\_ParseCache.db3)
* STARTDATE: (optional) limit the query to start from this date (YYYY-MM-DD)
* ENDDATE: (optional) limit the query to end at this date (YYYY-MM-DD)
* TAGS: (optional) limit the query to specific document characteristics (e.g. the Answer section of Earning Calls, companies in a specific sector, only text spoken by the CEO, etc.)
  + For a complete list of applicable tags, select a corpus in the web platform and select filter by tags
* M: Flag to download extra metadata
* outputCSV: path to store the generated Extraction CSV

Example Usage:

pronto\_nlp fief generate\_find\_matches\_csv -u "org:user@example.com" -p "password" -r Alpha -v "Margin|Dividend" -d SnP\_Transcripts\_ParseCache.db3 -s 2018-01-01 -e 2020-01-01 -g "#DocType\_EarningsCalls #Sector\_Utilities" myExtractions.csv

This will generate the Extraction CSV with the following characteristics:

* Extractions and their properties which matched on either a Margin or Dividend event from the Alpha model in the Transcript Corpus
* For the date range 2018 to 2020
* Limited to Earning Call transcripts and specifically the Utilities sector

Example Usage:

pronto\_nlp fief generate\_find\_matches\_csv -u "org:user@example.com" -p "password" -r Alpha -v "Excited|Disappointed" -d SnP\_ProntoStageDB\_First\_ParseCache.db3 -g "#SpeakerType\_Executives\_CEO #SpeakerType\_Executives\_CFO" myExtractions.csv

Similar to above, but no date limitation and showing extractions for Excited and Disappointed events exclusively in what the CEO or CFO said.

**Process User Documents**

This function allows users to upload documents to be processed by a ProntoNLP model of choice. The outputted format is chosen by the user and can be similar to either option above (i.e. actual extractions or aggregated event counts). The input documents must be a CSV file where each row contains a document. The system will run the selected model on the documents and add the result as an additional column to the input. This allows users to run ProntoNLP models on private documents to create a trading signal or generate insights.

To Process User Docs use the following command line prompt:

pronto\_nlp fief process\_corpus -u USER -p PASSWORD -r RULESET -o OUTPUTTYPE -n NUMTHREADS inputCSV outputCSV

* USER: your ProntoNLP username (the same one used to access the web platform)
* PASSWORD: your ProntoNLP password (the same one used to access the web platform)
* RULESET: name of a ruleset that exists in your account (e.g. Alpha, ESG, etc.)
* OUTPUTTYPE: type of the processing output (‘XML’, ‘JSON’, ‘events’)
* NUMTHREADS: number of concurrent processing threads
* inputCSV: path to CSV containing the user documents
* outputCSV: path to store the generated CSV

Example Usage:

pronto\_nlp fief process\_corpus -u "org:user@example.com" -p "password"-r Alpha -o events customDocs.csv customDocs\_analyzed.csv

**Common Tags for Filtering**

Transcript Data: SnP\_Transcripts\_ParseCache.db3

* Doc Type
  + Analyst Investor Day, Company Conference Presentations, Earnings Calls, Guidance Update Calls, Shareholder Analyst Calls, Special Calls
    - Answer, Presentation Opr Msg, Presentation Section, Presenter Speech, QA Opr Msg, Question, Question And Answer Section, Unknown QA Message
* Sector
  + Communication Services
    - Diversified Telecommunication Services
    - Entertainment
    - Interactive Media Services
    - Media
    - Wireless Telecommunication Services
  + Consumer Discretionary
    - Automobile Components
    - Automobiles
    - Broadline Retail
    - Distributors
    - Diversified Consumer Services
    - Hotels Restaurants Leisure
    - Household Durables
    - Leisure Products
    - Specialty Retail
    - Textiles Apparel Luxury Goods
  + Consumer Staples
    - Beverages
    - Consumer Staples Distribution Retail
    - Food Products
    - Household Products
    - Personal Care Products
    - Tobacco
  + Energy
    - Energy Equipment Services
    - Oil Gas Consumable Fuels
  + Financials
    - Banks
    - Capital Markets
    - Consumer Finance
    - Financial Services
    - Insurance
    - Mortgage Real Estate Investment Trusts REI Ts
  + Health Care
    - Biotechnology
    - Health Care Equipment Supplies
    - Health Care Providers Services
    - Health Care Technology
    - Life Sciences Tools Services
    - Pharmaceuticals
  + Industrials
    - Aerospace Defense
    - Air Freight Logistics
    - Building Products
    - Commercial Services Supplies
    - Construction Engineering
    - Electrical Equipment
    - Ground Transportation
    - Industrial Conglomerates
    - Machinery
    - Marine Transportation
    - Passenger Airlines
    - Professional Services
    - Trading Companies Distributors
    - Transportation Infrastructure
  + Information Technology
    - Communications Equipment
    - Electronic Equipment Instruments Components
    - IT Services
    - Semiconductors Semiconductor Equipment
    - Software
    - Technology Hardware Storage Peripherals
  + Materials
    - Chemicals
    - Construction Materials
    - Containers Packaging
    - Metals Mining
    - Paper Forest Products
  + Real Estate
    - Diversified REI Ts
    - Health Care REI Ts
    - Hotel Resort REI Ts
    - Industrial REI Ts
    - Office REI Ts
    - Real Estate Management Development
    - Residential REI Ts
    - Retail REI Ts
    - Specialized REI Ts
  + Utilities
    - Electric Utilities
    - Gas Utilities
    - Independent Power Renewable Electricity Producers
    - Multi Utilities
    - Water Utilities
* Speaker Type
  + Analysts
  + Attendees
  + Executives
    - CEO
    - CFO
    - CIO
    - COO
    - CTO
    - CMO
    - Chief Officer
    - Director
    - President
    - Board
  + Operator
  + Shareholders

SEC Filing Data: SnP\_SECFilings\_Parsecache.db3

* Doc Type
  + 10K
    - P1
      * I1, I1A, I1B, I2, I3, I4
    - P2
      * I5, I6, I7, I7A, I8, I9
    - P3
      * I10, I11, I12, I13, I14
    - P4
      * I15, I16
  + 10Q
    - P1
      * I1, I2, I3, I4
    - P2
      * I1, I1A, I2, I3, I4, I5
* Sector filters same as Transcript Data