

Rýnir - Notendahandbók

1 Inngangur

Rýnir er tölfræðigreinin sem hannaður er fyrir Datamarket.com.

Datamarket er gagnagátt sem veitir aðgengi að tölfræði og skipulögðum gögnum frá ýmsum stofnunum bæði opinberum og úr einkageiranum.

Tilgangur Rýnirs er að greina tiltekið safn af tímörðum og skila af sér flöggum sem segja til um áhugaverða atburði í tíma fyrir gefið safn.

2 Kröfur

Rýnir er skrifaður í Python 2.6 og styðst við eftirfarandi aðgerðasöfn (e. libraries)

- i NumPy, útgáfa 1.5.1 eða nýrra
<http://numpy.scipy.org/>
- ii SciPy, útgáfa 0.7.0 eða nýrra
<http://www.scipy.org/>
- iii Matplotlib, útgáfa 0.99.1.1 eða nýrra
<http://matplotlib.sourceforge.net/index.html>
- iv SciKits.timeseries, útgáfa 0.91.3 eða nýrra
<http://pytseries.sourceforge.net/index.html>
- v Unittest2, útgáfa 0.5.1 eða nýrra
<http://pypi.python.org/pypi/unittest2>
- vi Simplejson, útgáfa 2.0.9 eða nýrra
<http://pypi.python.org/pypi/simplejson/>

3 Almennar leiðbeiningar

3.1 Innsetning

Engra sérstakara aðgerða er þörf ef allar kröfur eru uppfylltar. Afbjappið skráarpakkanum á þeim stað sem hentar hverju sinni.

3.2 Stillingar

Í stillingaskrá (e.config) er hægt að stilla eftirfarandi breytur;

- Url: Slóðin á þann API sem skal tengjast á
- Key: Aðgangsslykill að API
- ConsolidateFlags: Gefur þann möguleika að skila einungis áhugaverðasta flagginu af röð samliggjandi flagga
- Filter: Segir til um hversu hátt áhugaverðaleika gildi flagg þarf svo því sé skilað
- K: Ákveður efri og neðri mörk sem K sinnum N-lotu staðalfrávik þannig að efri mörk eru $K * \sigma$ og neðri mörk $-(K * \sigma)$
- FrameSize: Ákveður rammastærð eða N-lotu einfalds fljótandi meðaltals (e. simple moving average)
- MaxTimelineSize: Ef fjöldi punkta í tímalínu er stærri en MaxTimelineSize þá er sú tímalína ekki send til greiningar
- NumberOfThreads: Segir til um hversu margar þræði kerfið hefur leyfi til að nota
- TopN: Ákveður hversu mörgum flöggum er skilað. Skilar TopN áhugaverðustu flöggum
- NMostRecent: Segir til um hvort skila eigi niðurstöðu fyrir alla tímalínuna eða einungis skoða NMostRecent nýjustu punktanna

Til að fulla greiningu og óskert gögn skal stilla bæði TopN og NMostRecent á 0.

3.3 Notkun

Í Rýni er hægt að kalla á fallið `analyze()` og senda með því lista af fyrirspurnarstrengjum (e.query strings). Nánari útlistun á fyrirspurnarstrengjum Datamarket er að finna á síðu þeirra <http://datamarket.com/api/v1/>.

Kallið í `analyze()` skilar lista af flöggum. Hvert flagg inniheldur upplýsingar um úr hvaða gagnasafni og tímalínu viðkomandi flagg er ásamt staðsetningu innan tímalínu.