

sorvi

avoimen datan työkalupaketti

Leo Lahti* ja Juuso Parkkinen†

9. lokakuuta 2011

1 Johdanto

Vapaasti saatavilla olevien tietoaineistojen määrä on ollut voimakkaassa kasvussa [2]. Talouteen, säähän, liikenteeseen, koulutukseen ja muihin alueisiin liittyvää dataa on alettu avaamaan julkishallinnon toimesta Suomessa ja muualla. Aineistojen pöyhintä voi avata uusia näkökulmia ja avoimuus mahdollistaa yhdistämisen toisiin tietolähteisiin, jolloin voidaan vastata kysymyksiin joihin yksittäisillä aineistoilla ei päästä käsiksi. Siten laaja saatavuus ja käyttö voi merkittävästi nostaa datan arvoa.

Laskennallisten työkalujen saatavuus on osoittautunut pullonkaulaksi avointen datojen laajemmalle käytölle. Tämä R-paketti pyrkii paikkaamaan puuttava tarjoamalla yleiskäyttöisiä välineitä avointen rajapintojen kautta saatavilla olevien julkisten tietoaineistojen automatisoituun lataamiseen, siistimiseen, yhdistelyyn, louhintaan ja visualisointiin. Pakettiin pyritään kokoamaan kattava kokoelma eri lähteistä koottuja ratkaisuja avoimen datan käsittelyyn. Uusia ehdotuksia ja lisäyksiä otetaan mieluummin vastaan.

Yksityiskohtaisempia esimerkkejä paketin käytöstä Suomi-datan penkomiseen löydät Louhos-blogista¹.

2 Paketin asennus

Paketin saa asennetuksi suoraan R:stä käsin:

```
> install.packages("dmt", repos = "http://R-Forge.R-project.org")
```

*Helsingin yliopisto <leo.lahti@iki.fi>

†Aalto-yliopisto <juuso.parkkinen@gmail.com>

¹<http://louhos.wordpress.com>

3 Pakettiin viittaaminen

Paketin tarjoamat välineet ovat vapaasti käytettävissä FreeBSD-lisenssillä². Mikäli paketista on apua, toivomme viittausta työhön [1]. Lisää tietoa löytyy projektin kotisivulta³.

4 Versiotiedot

```
> sessionInfo()
```

```
R version 2.13.0 (2011-04-13)
```

```
Platform: i686-pc-linux-gnu (32-bit)
```

```
locale:
```

```
[1] LC_CTYPE=fi_FI.utf8      LC_NUMERIC=C
[3] LC_TIME=fi_FI.utf8      LC_COLLATE=fi_FI.utf8
[5] LC_MONETARY=C           LC_MESSAGES=fi_FI.utf8
[7] LC_PAPER=fi_FI.utf8     LC_NAME=C
[9] LC_ADDRESS=C            LC_TELEPHONE=C
[11] LC_MEASUREMENT=fi_FI.utf8 LC_IDENTIFICATION=C
```

```
attached base packages:
```

```
[1] stats      graphics  grDevices  utils      datasets  methods    base
```

```
loaded via a namespace (and not attached):
```

```
[1] tools_2.13.0
```

Viitteet

- [1] Leo Lahti ja Juuso Parkkinen (2011). sorvi - avoimen datan työkalupaketti
URL: <http://louhos.wordpress.com>
- [2] Antti Poikola, Petri Kola ja Kari A. Hintikka (2010). Julkinen data – johdatus tietovarantojen avaamiseen Liikenne- ja Viestintäministeriö. Edita Prima Oy, Helsinki 2010. URL: <http://www.julkinendata.fi>

²http://en.wikipedia.org/wiki/BSD_licenses

³<http://sorvi.r-forge.r-project.org/>