

soRvi

avoimen datan työkalupakki

Leo Lahti*, Juuso Parkkinen[†] ja Joona Lehtomäki[‡]

9. heinäkuuta 2012

1 Johdanto

Kaikille avointen yhteiskunnallisten tietovarantojen määrä on voimakkaassa kasvussa [1]. Talouteen, säähän, paikkatietoon, liikenteeseen, koulutukseen ja muihin alueisiin liittyvää dataa on alettu avaamaan julkishallinnon toimesta Suomessa ja muualla. Aineistojen pöyhintä avaa uusia näkökulmia ja avoimuus mahdollistaa yhdistämisen toisiin tietolähteisiin, jolloin voidaan vastata kysymyksiin joihin yksittäisillä aineistoilla ei päästä käsiksi. Laaja saatavuus ja käyttö voi siten merkittävästi nostaa datan arvoa.

Laskennallisten työkalujen saatavuus on osoittautunut keskeiseksi pullonkaulaksi. Tämä R-kielinen työkalupakki pyrkii paikkaamaan puutetta tarjoamalla yleiskäyttöisiä välineitä julkisten tietoaaineistojen hakuun, putsaamiseen, yhdistelyyn, louhintaan ja visualisointiin. Pakettiin on koottu kattava kokoelma ratkaisuja avoimen datan käsittelyyn. Myös uudet ehdotukset ja lisäykset ovat tervetulleita.

2 Asennus

Asennus- ja kehitysohjeet löytyvät projektin kotisivulta <http://louhos.github.com/sorvi>.

3 Esimerkit

Yksityiskohtaisia esimerkkejä paketin käytöstä Suomi-datan penkomiseen löydät lähdekoodeineen Louhos-blogista <http://louhos.wordpress.com>.

*Wageningen University, Nederland <leo.lahti@iki.fi>

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

[‡]Helsingin yliopisto

4 Yhteystiedot

Yhteystiedot löytyvät projektin kotisivulta <http://louhos.github.com/sorvi>.

5 Pakettiin viittaaminen

Paketin välineet ovat vapaasti käytettävissä, muokattavissa, ja edelleen levitettävissä. FreeBSD-lisenssillä¹. Mikäli paketista on apua, toivomme viittaamista työhön [2]. Lisää tietoa löytyy projektin kotisivulta².

6 Versiotiedot

Tämä manuaali on luotu seuraavin asetuksin:

```
> sessionInfo()
```

```
R version 2.15.0 (2012-03-30)
```

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

```
locale:
```

```
[1] LC_CTYPE=en_US.UTF-8      LC_NUMERIC=C
[3] LC_TIME=en_US.UTF-8      LC_COLLATE=en_US.UTF-8
[5] LC_MONETARY=en_US.UTF-8  LC_MESSAGES=en_US.UTF-8
[7] LC_PAPER=C               LC_NAME=C
[9] LC_ADDRESS=C            LC_TELEPHONE=C
[11] LC_MEASUREMENT=en_US.UTF-8 LC_IDENTIFICATION=C
```

```
attached base packages:
```

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

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

```
[1] tools_2.15.0
```

Viitteet

- [1] 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>
- [2] Leo Lahti, Juuso Parkkinen ja Joona Lehtomäki (2011). soRvi - suomalainen avoimen datan työkalupakki URL: <http://louhos.wordpress.com>

¹http://en.wikipedia.org/wiki/BSD_licenses

²<http://louhos.github.com/sorvi>