

# Ulkomaankauppatilastojen jakelujärjestelmä ULJAS

-Isojen datojen ja pitkien luokitusten  
haasteet

Jaana Lehtilä, Tullihallitus, tilastoyksikkö



- Uljaksessa on yksityiskohtaiset tiedot tavaroiden viennistä ja tuonnista
- Tietokanta on ilmainen, eikä vaadi rekisteröitymistä
- Osoite: <http://uljas.tulli.fi>
- Palvelu avattiin huhtikuussa 2008
- Korvasi 10 vuotta käytössä olleen Ultika-tietokannan
- Aineistoa vuodesta 1987 lähtien vuosittaisena ja vuodesta 1996 lähtien kuukausittaisena



- Tarvekartoituksessa haastateltiin 16 käyttäjää ja Ultikan parissa työskennellyttä tullilaista
- Jo haastattelussa tuli esille aineistojen koko, aikasarjojen muodostamisen ongelmat ja laaja, vuosittain muuttuva, vaikea tavaroiden luokitus
- Näihin ja muihinkin haasteisiin lähdimme hakemaan ratkaisua



Uljias ► Ulkomaankauppatilasto ► CN

## Tilasto: CN

Valitse luokat tämän sivun valintalaatikoista tai mene [laajennetun poiminnan sivulle](#)

[Valintalaatikoiden käyttöohje](#)

[Tilaston perustiedot](#) [Alaviite](#) [Kuvaus](#)

<b>Tavaruokitus CN</b> <input checked="" type="checkbox"/> Yhteensä 19263 Valittu 0	<b>Aika</b> <input checked="" type="checkbox"/> Yhteensä 96 Valittu 0	<b>Maa</b> <input checked="" type="checkbox"/> Yhteensä 241 Valittu 0
00 - 99 (2002--) KAIKKI RYHMÄT 01 (2002--) ELÄVÄT ELÄIMET 0101 (2002--) Elävät hevoset, aasit, muilit ja r 010110 (2002--) Puhdasrotuiset siitoshevoset 01011010 (2002--) Puhdasrotuiset siitoshevos 01011090 (2002--) Puhdasrotuiset siitosaasit 0101190 (2002--) Elävät hevoset, aasit, muilit j	200912 200911 200910 200909 200908 200907 200906	AA (2002--) Kaikki maat yhteensä AD (2002--) Andorra AE (2002--) Arabiemiirikunnat AF (2002--) Afganistan AG (2002--) Antigua ja Barbuda AI (2002--) Anguilla AL (2002--) Albania
<input type="text"/> Hae <input type="checkbox"/> alusta	<input type="text"/> Hae <input type="checkbox"/> alusta	<input type="text"/> Hae <input type="checkbox"/> alusta
<b>Suunta</b> <input checked="" type="checkbox"/> Yhteensä 3 Valittu 0	<b>Indikaattorit</b> <input checked="" type="checkbox"/> Yhteensä 14 Valittu 0	
Tuonti alkuperämaittain Vienti määrämaittain Tuonti lähetyksittäin	Tilastoarvo (euro) Muutos% Paljous Paljouden yksikön koodi Tilastoarvo/Paljous Lisäpaljous Lisäpaljouden yksikön koodi	
<input type="text"/> Hae <input type="checkbox"/> alusta	<input type="text"/> Hae <input type="checkbox"/> alusta	

Valittuja rivejä: 0. Valittuja sarakkeita: 0. Valittuja soluja: 0. [Poista kaikki valinnat](#)

Näkymä ruudulla on rajoitettu 3000 riviin ja 120 sarakkeeseen

PX-Web 2007, Graph.Pages 2.4.1, StatFile 2.4.63





- Päätimme, että ratkaisu isoihin aineistoihin ja luokitusten käyttöön ei löydy kilpailutusehdoilla
- Syynä se, että palvelun kehittäminen tältä osin vaatii hyvää substanssitietämystä varsinkin luokituksista ja erilaisista luokituksen käytötavoista -> ohjelmointityö itselle
- Toinen tärkeä ratkaisu: päätimme tehdä jakelujärjestelmän emme raportointijärjestelmää (oma ajatusmallinsa kummassakin)
- Kilpailutuksen voitti VERTI



- Joka kuukausi tulee yrityksistä kauppaerittäin tietoa (vuoden 2009 lopun tilanteen mukaan):
  - EU kaupassa 590 000 riviä
  - ei-EU kaupassa 155 000 riviä
- Summaus tavaratasolle CN-luokituksella:
  - **3 100 000** nollista eroavaa riviä vuonna 2009 -> Lisäksi kaksi luokitusta lisää
  - Salausten vuoksi summatasojen summat viedään datassa mukana (2002 vanhemmissa datoissa on salaus laskelmissa nollana)



- Pitkä, mutta kapea data, 2002-2009 CN-luokituksen aineisto

$$19263 * 96 * 241 * 3 * 8 =$$

**10 696 050 432** solua (n. 2,1 mrd riviä)

(+ 6 laskennallista saraketta lisää)

-> datassa paljon tyhjää

-siirtotiedoston koko tällä hetkellä:

noin **3,7 Gb**



- Rakenne kaikki yhdessä vai vuodet erikseen vai luokitustasot erikseen tai muu osajoukko
- Päivitys ja ajot vievät aikaa
  - Summaukset vievät koko yön
  - Synkronointi tällä hetkellä 4,5 tuntia -> nopeutus 30 minuuttia
  - Testaus haasteellista
- Hakujen kuormitus huomioitava
  - **Blokkaus** (haku ositettu, ensin uusimpaan)
  - Hakujen rajoittaminen (3000 riviä selaimeen, 200 000 riviä suoraan tiedostoon)
- ”Tyhjän tiedon” kontrollointi
- vihreällä -merkityt ostettu erikseen Uljukseen



Ohje

- Etusivulle
- Uljas
  - Ulkomaankauppatilasto
    - CN
    - CPA
    - SITC
- Omat haut

CN

CN vuodesta 2002> - Päivitetty 1.3.2010 9:45:53

- Tavaraluokitus CN: 00 - 99 (2002-- ) KAIKKI RYHMÄT, 01 (2002-- ) ELÄVÄT ELÄIMET, ..., XX (2002-- ) 87 JA 93 RYHMÄ YHDESSÄ (
- Aika: 200912, 200911, ..., 200201 (96)
- Maa: AA (2002-- ) Kaikki maat yhteensä, AD (2002-- ) Andorra, ..., ZW (2002-- ) Zimbabwe (241)
- Suunta: Tuonti alkuperämättäin, Vienti määrämättäin, Tuonti lähetymättäin (3)
- Indikaattorit: Tilastoarvo (euro), Muutos%, ..., Kum. tilastoarvo/Kum. lisäpajous (14)

CN IHS 1989 - 2002 - Päivitetty 25.2.2010 20:31:45

- Tavaraluokitus CN vuoteen 2002: 00 - 99 (1989--2002), 01 (1989--2002), ..., 99900000 (1995--1996) (22287)
- Aika: 200212, 200211, ..., 198912 (91)
- Maa vuoteen 2002: AA (--2002) Kaikki maat yhteensä, AD (1995--2002) Andorra, ..., ZW (--2002) Zimbabwe (277)
- Suunta vuoteen 2002: Tuonti alkuperämättäin, Vienti määrämättäin (2)
- Indikaattorit: Tilastoarvo (1000 euroa, 1998-2002), Muutos% (1997-2002), ..., Kum. tilastoarvo/Kum. lisäpajous (1990-2002) (14)

PX-Web 2007, Graph.Pages 2.4.1, GraphDir 2.4.63

Suomessa

- > Tjänsteproduktionsindex (TJPI), 2005=100, efter näringsgren SNI 2002, Kvartal 2005K1-2008K4 [2009-02-17]
- > Tjänsteproduktionsindex (TJPI), 2005=100, efter näringsgren SNI 2002, Månad 2005M01-2008M12 [2009-02-17]

## Utrikeshandel med varor

> [Mer om denna statistik](#)

### Total och landfördelad

- > Total varuimport, varuexport och handelsnetto, År 1975-2009 [2010-02-26]
- > Total varuimport, varuexport och handelsnetto, Månad 1975M01-2010M01 [2010-02-26]
- > Varuimport och varuexport, Totala värden efter handelspartner, bortfallsjusterat, År 1998-2009 [2010-02-26]
- > Varuimport och varuexport, Totala värden efter handelspartner, bortfallsjusterat, Månad 1998M01-2010M01 [2010-02-26]

### SITC (Standard International Trade Classification)

- > Varuimport och varuexport efter varugrupp SITC, bortfallsjusterat, År 1998-2009 [2010-02-26]
- > Varuimport och varuexport efter varugrupp SITC, bortfallsjusterat, Månad 1998M01-2009M12 [2010-02-26]
- > Varuimport och varuexport efter handelspartner och varugrupp SITC, ej bortfallsjusterat, År 1995-2009 [2010-02-26]
- > Varuimport och varuexport efter handelspartner och varugrupp SITC, ej bortfallsjusterat, Månad 1995M01-2009M12 [2010-02-26]
- > Varuimport och varuexport efter varugrupp SITC, Volymindex 2000=100, År 1995-2009 [2010-03-04]
- > Varuimport och varuexport efter varugrupp SITC, Volymindex 2000=100, Kvartal 1995K1-2009K4 [2010-03-04]

### SITC, äldre serier

- > Varuimport och varuexport efter varugrupp SITCrev3 och handelspartner, total, År 1977-1994 [1999-12-09]
- > Varuimport och varuexport efter varugrupp SITCrev3, Volymindex 1980=100, År 1980-1993 [1999-12-22]
- > Varuimport och varuexport efter varugrupp SITCrev3, Volymindex 1980=100, Kvartal 1988K1-1994K2 [1999-12-22]

### SPIN (Standard för svensk produktindelning efter näringsgren)

- > Varuimport och varuexport efter produktgrupp SPIN2007, bortfallsjusterat, År 2000-2009 [2010-02-26]
- > Varuimport och varuexport efter produktgrupp SPIN2007, bortfallsjusterat, Månad 2000M01-2009M12 [2010-02-26]
- > Varuimport och varuexport efter produktgrupp SPIN2007 och handelspartner, ej bortfallsjusterat, År 2000-2009 [2010-02-26]
- > Varuimport och varuexport efter produktgrupp SPIN2007 och handelspartner, ej bortfallsjusterat, Månad 2000M01-2009M12 [2010-02-26]
- > Varuimport och varuexport efter produktgrupp SPIN 2007, Volymindex 2000=100, År 2001-2009 [2010-03-04]
- > Varuimport och varuexport efter produktgrupp SPIN 2007, Volymindex 2000=100, Kvartal 2001K1-2009K4 [2010-03-04]

### SPIN, Prod-SNI och Varu-SNI, äldre serier

- > Varuimport och varuexport efter produktgrupp Prod-SNI97 och handelspartner, ej bortfallsjusterat, År 1995-1998 [2000-01-25]
- > Varuimport och varuexport efter produktgrupp Prod-SNI97 och handelspartner, ej bortfallsjusterat, Månad 1995M01-1999M08 [2000-01-25]
- > Total varuimport och varuexport efter varugrupp Varu-SNI69 och handelspartner, År 1969-1994 [1999-12-10]
- > Varuimport och varuexport efter produktgrupp SPIN 2002, bortfallsjusterat, År 1998-2008 [2010-02-23]

Ruotsissa

PX-Web 2007, VERTITULLI 2.3.0



Uljias ► Ulkomaankauppatilasto ► CN

**Tilasto: CN**

**Valitse luokat tämän sivun valintalaatikoista tai mene laajennetun poiminnan sivulle**

[Valintalaatikoiden käyttöohje](#)

[Tilaston perustiedot](#) [Alaviite](#) [Kuvaus](#)

<b>Tavaraluokitus CN</b> Yhteensä <b>19263</b> Valittu <b>0</b>	<b>Aika</b> Yhteensä <b>96</b> Valittu <b>0</b>	<b>Maa</b> Yhteensä <b>241</b> Valittu <b>0</b>
00 - 99 (2002--) KAIKKI RYHMÄT 01 (2002--) ELÄVÄT ELÄIMET 0101 (2002--) Elävät hevoset, aasit, muilit ja r... 010110 (2002--) Puhdasrotuiset siitoshevoset 01011010 (2002--) Puhdasrotuiset siitoshevos... 01011090 (2002--) Puhdasrotuiset siitosaasit 010190 (2002--) Elävät hevoset, aasit, muilit j...	200912 200911 200910 200909 200908 200907 200906	AA (2002--) Kaikki maat yhteensä AD (2002--) Andorra AE (2002--) Arabiemiirikunnat AF (2002--) Afganistan AG (2002--) Antigua ja Barbuda AI (2002--) Anguilla AL (2002--) Albania
Hae <input type="text"/> alusta >	Hae <input type="text"/> alusta >	Hae <input type="text"/> alusta >
<b>Suunta</b> Yhteensä <b>3</b> Valittu <b>0</b>	<b>Indikaattorit</b> Yhteensä <b>14</b> Valittu <b>0</b>	
Tuonti alkuperämaittain Vienti määrämaittain Tuonti lähetyksittäin	Tilastoarvo (euro) Muutos% Paljous Paljouden yksikön koodi Tilastoarvo/Paljous Lisäpaljous Lisäpaljouden yksikön koodi	
Hae <input type="text"/> alusta >	Hae <input type="text"/> alusta >	

Valittuja rivejä: 0. Valittuja sarakkeita: 0. Valittuja soluja: 0. Poista kaikki valinnat  
 Näkymä ruudulla on rajoitettu 3000 riviin ja 120 sarakkeeseen

PX:Web 2007, Graph.Pages 2.4.1, StatFile 2.4.63





# Miten näkyvät käyttäjälle

Ullas ▶ Ulkomaankauppatilasto ▶ CN ▶ Muuttujat ja luokat

- Muokkaa ja laske
- Muokkaa ja laske
- Näytä uudelleenmuotoiltu taulukko
- Poista nollarivit
- Poista nolla- ja viivarivit
- Käännä myötäpäivään
- Käännä vastapäivään
- Käännä manuaalisesti
- Summaa muuttuja
- Prosenttilaskenta
- Valitse teksti/koodi taulukkoon
- Muuta desimaaleja
- Poista muuttuja
- Muuta muuttujan arvojen järjestystä
- Poista muuttujan arvo

Tallenna



	Tuonti alkuperämittain	
	Paljouden yksikön koodi	Kum. paljous
sisättyä sokeria tai muuta makeutusainetta sisältämätön		
	1	34 496 917
	1	2
	1	2 442 318
	1	1 029
EE (2002--.) Viro	1	1 458 901
FR (2002--.) Ranska	1	9 242
GB (2002--.) Iso-Britannia	1	76 736
IE (2002--.) Irlanti	1	193 996
NL (2002--.) Alankomaat	1	3 007
PL (2002--.) Puola	1	2 880
SE (2002--.) Ruotsi	1	30 308 806

**Alaviite:**  
 Paljouksien yksiköt: - = n.a., 1 = kg, 2 = ct/l, 3 = 100 kpl, 4 = c/k, 5 = gi F/S, 6 = g, 7 = GT, 8 = hl, 9 = kg k.kl., 10 = kg tot/a, 11 = kg/net e, 12 = kg/br, 13 = kg/net, 14 = 1000 l, 15 = kg m.am., 16 = kg N, 17 = kg H2O2, 18 = kg H2O2, 19 = kg KOH, 20 = kg K2O, 21 = kg P2O5, 22 = kg 90% s, 23 = kg NaOH, 24 = kg U, 25 = l a.100%, 26 = l, 27 = 1000 kpl, 28 = m2, 29 = m3, 30 = 1000 m3, 31 = m, 32 = MWh, 33 = kpl, 34 = ce/el, 35 = pari, 36 = t, 37 = TJ, 38 = t, 39 = brt, 40 = kg-V, 41 = kg-U

*This page was created by PX-Web. A member of the PC-Axis family.*

0401 – tavarakoodi: 1 litra maitoa painaa 1,03 kg.



- Tarkin tavaraluokitus on ns. CN8-luokitus. Esim.

48044290 Päälystämätön voimapaperi, -kartonki ja -pahvi, valkaistu tasaisesti kauttaaltaan massana, rullina, joiden leveys on  $> 36$  cm, tai suorakaiteen tai neliön muotoisina arkkeina, joiden yksi sivu on pituudeltaan  $> 36$  cm ja toinen  $> 15$  cm kun arkki on aukilevitetty, kokonaiskuitusisällöstä  $> 95$  painoprosenttia kemiallisella menetelmällä saatuja puukuituja, paino  $> 150$  g/m<sup>2</sup> mutta  $< 225$  g/m<sup>2</sup> (ei kuitenkaan kokonaiskuitusisällöstä  $\geq 80$  painoprosenttia kemiallisella sulfaatti- tai soodamenetelmällä saatuja havupuukuituja sisältävä)



- CN-koodeja ylätasoinen 19 263 kappaletta ajanjaksolla 2002-2009, kaikki samassa listassa, koska nimikkeistössä ei koskaan oteta uudelleen samaa nimikenumeroa käyttöön eri tarkoituksessa
- Luokitusten ryhmittely ja samalla porautumisilmiö hoidettu luokituksen koodistolla "tasonvaihtoina"
- Esim. 04 ryhmästä



- **04 (2002--.) MAITO JA MEIJERITUOTTEET; LINNUNMUNAT; LUONNONHUNAJA; MUUALLE KUULUMATTOMAT ELÄINPERÄISET SYÖTÄVÄT TUOTTEET**
- **0401 (2002--.) Maito ja kerma, tiivistämätön lisättyä sokeria tai muuta makeutusainetta sisältämätön**
- **040110 (2002--.) Maito ja kerma, rasvapitoisuus  $\leq 1$  painoprosentti, tiivistämätön, lisättyä sokeria ja muuta makeutusainetta sisältämätön**
  - 04011010 (2002--.) Maito ja kerma, rasvapitoisuus  $\leq 1$  painoprosentti, tuotetta lähinnä olevan pakkauksen nettosisältö  $\leq 2$  l, tiivistämätön, lisättyä sokeria tai muuta makeutusainetta sisältämätön
  - 04011090 (2002--.) Maito ja kerma, rasvapitoisuus  $\leq 1$  painoprosenttia, tiivistämätön, lisättyä sokeria tai muuta makeutusainetta sisältämätön (paitsi tuotetta lähinnä olevan pakkauksen nettosisältö  $\leq 2$  l)
- **040120 (2002--.) Maito ja kerma, rasvapitoisuus  $> 1$  painoprosentti mutta  $\leq 6$  painoprosenttia, tiivistämätön, lisättyä sokeria tai muuta makeutusainetta sisältämätön**
  - 04012011 (2002--.) Maito ja kerma, rasvapitoisuus  $> 1$  painoprosentti mutta  $\leq 3$  painoprosenttia, tuotetta lähinnä olevan pakkauksen nettosisältö  $\leq 2$  l, tiivistämätön, lisättyä sokeria tai muuta makeutusainetta sisältämätön
  - 04012019 (2002--.) Maito ja kerma, rasvapitoisuus  $> 1$  painoprosentti mutta  $\leq 3$  painoprosenttia, tiivistämätön, lisättyä sokeria tai muuta makeutusainetta sisältämätön (paitsi tuotetta lähinnä olevan pakkauksen nettosisältö  $\leq 2$  l)



- Joka vuosi usea sata nimikettä vaihtuu, poistuu tai tulee uusia, laitoimme voimassaolon vuosiluvut nimen eteen
- Tekstien pituus 1000 merkkiä, valintalaatikot venyvät -> tulosteissa 256 merkkiä, kuvioissa edelleen ongelma
- Nimikkeillä myös vuosittain määritelty lisäpaljous, jos on sellaista vaadittu
- Joissakin nimikkeissä on salauksia (merkitty ...)



- Tavoite ei erillisiä käsikirjoja
- Mietittiin asiakkaan käyttötapoja ja luokiteltiin eri asiakaskunnat (käyttötapakuvauksia)
  - Satsattiin koulutukseen asiakaskunnittain, 60% tilastojen käyttäjistä yrityksiä
  - Tulossa ohjeistusta lisää
- Miten asiakas löytää oikean luokituksen ja saa valittua useamman yli ryhmärajojen?
  - Sanahaut
  - Laajennettu poiminta



# Esitys laajennetusta poiminnasta



## Laajennettu poiminta

**Luokitus: Tavaraluokitus CN2**

Edellinen

Seuraava luokitus

Valmis

Peruuta

Hae:  Hakukriteeri:  Haun kohdeluokitus:

### Merkitse luokkia poimintaa varten:

Löytyi 14 luokkaa (Tavaraluokitus CN)

- 640212 (2002--.) Laskettelu-, pujottelu- ja muut hiihtokengät sekä lumilautailukengät, joiden ulkopohjat
- 64021210 (2002--.) Laskettelu-, pujottelu- ja muut hiihtokengät, joiden ulkopohjat ja päälliset ovat kumia
- 64021290 (2002--.) Lumilautailukengät, joiden ulkopohjat ja päälliset ovat kumia tai muovia (paitsi nimikkeet
- 640219 (2002--.) Urheilujalkineet, joiden ulkopohjat ja päälliset ovat kumia tai muovia (paitsi nimikkeet
- 64021900 (2002--.) Urheilujalkineet, joiden ulkopohjat ja päälliset ovat kumia tai muovia (paitsi nimikkeet
- 640312 (2002--.) Laskettelu-, pujottelu- ja muut hiihtokengät sekä lumilautailukengät, joiden ulkopohjat
- 64031200 (2002--.) Laskettelu-, pujottelu- ja muut hiihtokengät sekä lumilautailukengät, joiden ulkopohjat
- 640319 (2002--.) Urheilujalkineet, joiden ulkopohjat ovat kumia, muovia, nahkaa tai tekonahkaa ja päälliset

Tee poiminta luokitukselta:

Poimittu 2 luokkaa

- 0401 (2002--.) Maito ja kerma, tiivistämätön lisättyä sokeria tai muuta makeutusainetta sisältämätön
- 640212 (2002--.) Laskettelu-, pujottelu- ja muut hiihtokengät sekä lumilautailukengät, joiden ulkopohjat ja

Poista merkityt rivit listasta

PX-Web 2007, Graph.Pages 2.4.1, StatFile 2.4.63



# Miten asiakkaat selviävät luokitusmuutoksista

---

- Ulrikassa valittiin datat vuosi kerrallaan, Uljaksessa se ominaisuus saatu aineistoon ”sisäänrakennettua”
- Jälleen apuna, että kaikki luokitusryhmät samaan aikaan haettavissa
- Vuosiluvut nimiketektien ja maiden nimien edessä
- Etusivulla varoitusteksti vuoden vaihtuessa sekä mahdollisesti linkki luokitusavaimiin
- Myös paljoudet datassa muuttujana, muutettiin numeeriseksi (koodien selitykset alaviitteessä)



- Asiakastyytyväisyyskyselyssä Uljaksen kouluarvosana oli 8
- Tulossa käytettävyyden parantamiseksi näyttöjen, ohjeiden, linkkien, yms. selkeyttämistä

## Yhteystiedot:

- Tullihallitus, tilastoyksikkö  
[tilastot@tulli.fi](mailto:tilastot@tulli.fi) tai puh. 020 690 603