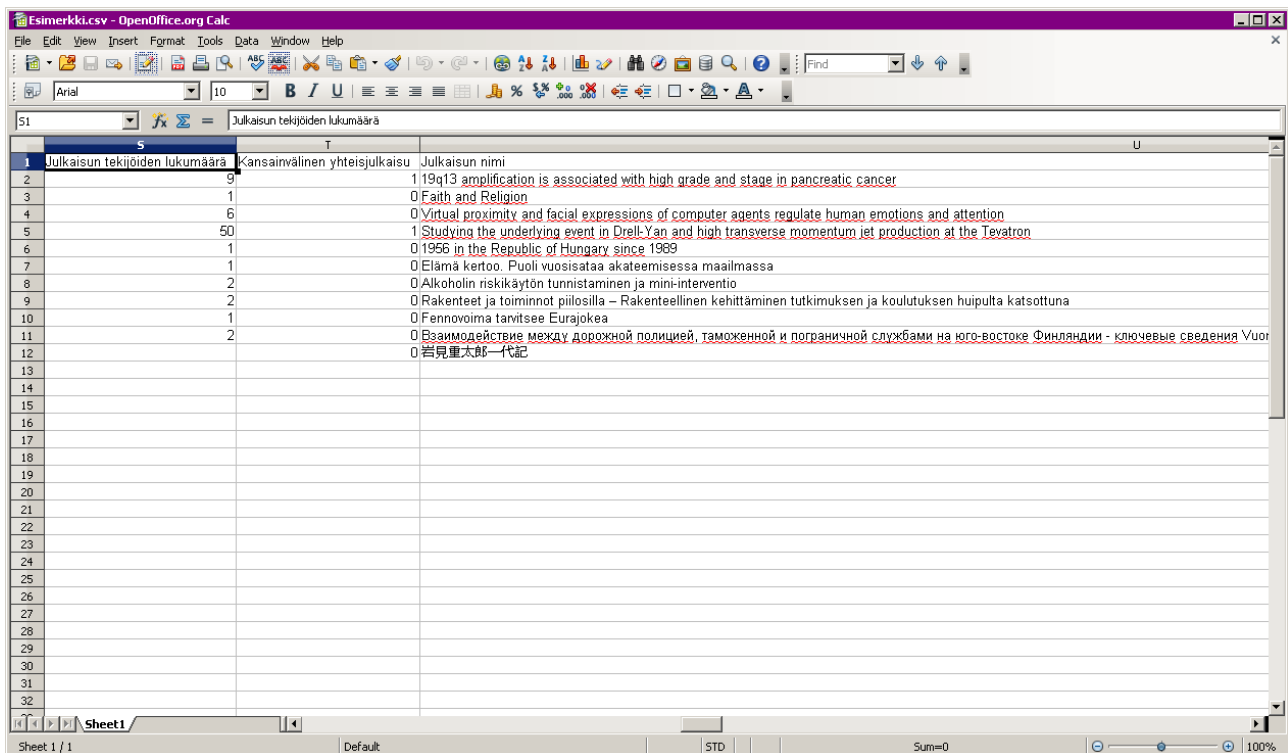


## Ohje UTF-8-merkistökoodatun CSV-tiedoston tuottamiseksi Calc-ohjelmalla

Suomalaisten korkeakoulujen julkaisu toiminta on kansainvälistä. Julkaisujen ja tekijöiden nimissä esiintyy merkkejä, jotka eivät mahdu yhteen perinteiseen maakohtaiseen merkistöön. Kerättyjen tietojen automaattinen käsittely johtaa virheellisiin tuloksiin, jos osa merkeistä rikkoontuu tiedonsiirrossa. Globalisoituneen maailman merkistö- ja tiedonsiirto standardit on luotu jo yli 20 vuotta sitten, mutta kaikki ohjelmit (mm. Excel) eivät edelleenkään tue näitä standardeja.

Tässä ohjeessa kuvataan yksi tapa miten OKM:n julkaisutiedonkeruun edellyttämä UTF-8-merkistökoodattu CSV-siirtotiedosto voidaan tuottaa. Ohje perustuu ilmaisen Open Office -ohjelmistoon kuuluvaan Calc-aulukkolaskentaohjelmaan. Ohjelma on ladattavissa osoitteesta <http://www.openoffice.org/fi/>. Korkeakoulut ylläpitävät julkaisutietoja erilaisissa tietojärjestelmissä, joten tässä ohjeessa on mahdoton kuvata kaikille korkeakouluille yleispätevällä tavalla tietojen toimittaminen alusta loppuun.

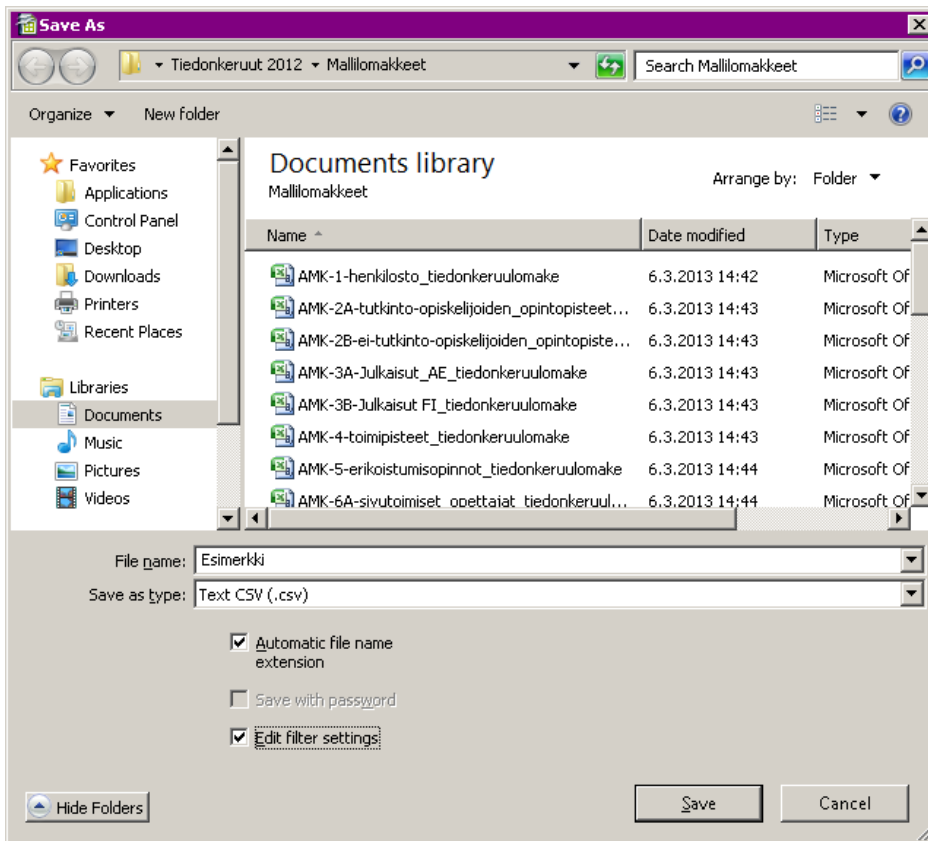
Tämän ohjeen lähtökohta on, että OKM:n julkaisutiedonkeruun edellyttämät tiedot luodaan Calc-ohjelmalla tai ladataan siihen (File-valikon Open-toiminnolla) olemassa olevasta tiedostosta. Ensimmäiseksi on syytä varmistaa silmämääräisesti, että tiedot näyttävät kaikin puolin oikeellisilta (esim. erikoismerkit näkyvät oikein ja tiedot ovat oikeissa sarakkeissa). Calc-ohjelma huolehtii lainausmerkkien lisäämisestä automaattisesti puolipisteitä sisältäviin tekstikenttiin, joten lainausmerkkejä ei pidä lisätä käsin kenttiin!



	S	T	U
1	Julkaisun tekijöiden lukumäärä	Kansainvälinen yhteisjulkaisu	Julkaisun nimi
2	9		1 19q13 amplification is associated with high grade and stage in pancreatic cancer
3	1		0 Faith and Religion
4	6		0 Virtual proximity and facial expressions of computer agents regulate human emotions and attention
5	50		1 Studying the underlying event in Drell-Yan and high transverse momentum jet production at the Tevatron
6	1		0 1956 in the Republic of Hungary since 1989
7	1		0 Elämä kertoo. Puoli vuosisataa akateemisessa maailmassa
8	2		0 Alkoholin riskikäytön tunnistaminen ja mini-interventio
9	2		0 Rakenteet ja toiminnot piilosilla – Rakenteellinen kehittäminen tutkimuksen ja koulutuksen huipulta katsottuna
10	1		0 Fennovoima tarvitsee Eurajokea
11	2		0 Взаимодействие между дорожной полицией, таможенной и пограничной службами на юго-востоке Финляндии - ключевые сведения
12			0 岩見重太郎一代記
13			
14			
15			
16			
17			
18			
19			
20			
21			
22			
23			
24			
25			
26			
27			
28			
29			
30			
31			
32			

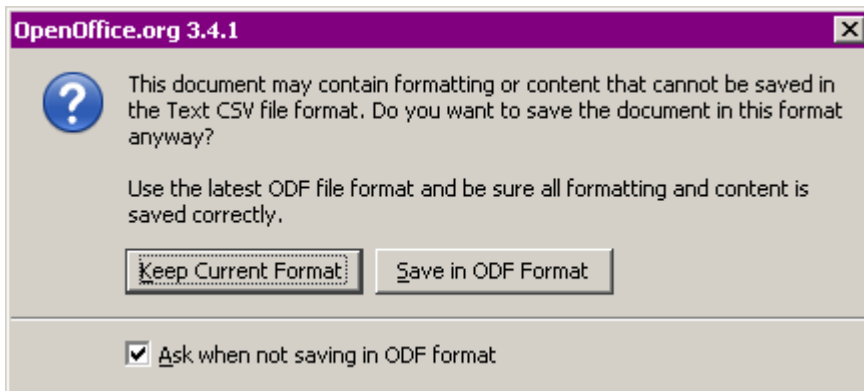
Kuva 1. Tiedoston tarkastelu Calc-ohjelmassa.

Jos tiedosto näyttää olevan kunnossa, niin seuraavaksi valitaan File-valikosta Save As ja annetaan tiedostolle nimi, tiedoston tyyppiä valitaan "Text CSV (.csv)" ja laitetaan rasti ruutuun Edit filter settings kuten kuvassa 2.



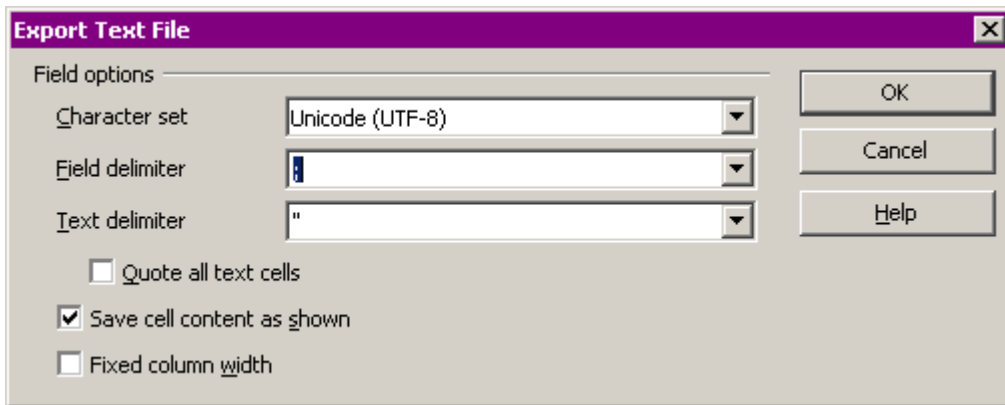
**Kuva 2. Save As -valikon oikeat valinnat.**

Save-näppäimen painamisen jälkeen näytölle tulee varoitus, ettei kaikki muotoilu säily CSV-muodossa. Tätä ei tarvitse hätkähtää, vaan ilmoitus sivuutetaan tyyneästi painamalla Keep Current Format -näppäintä.



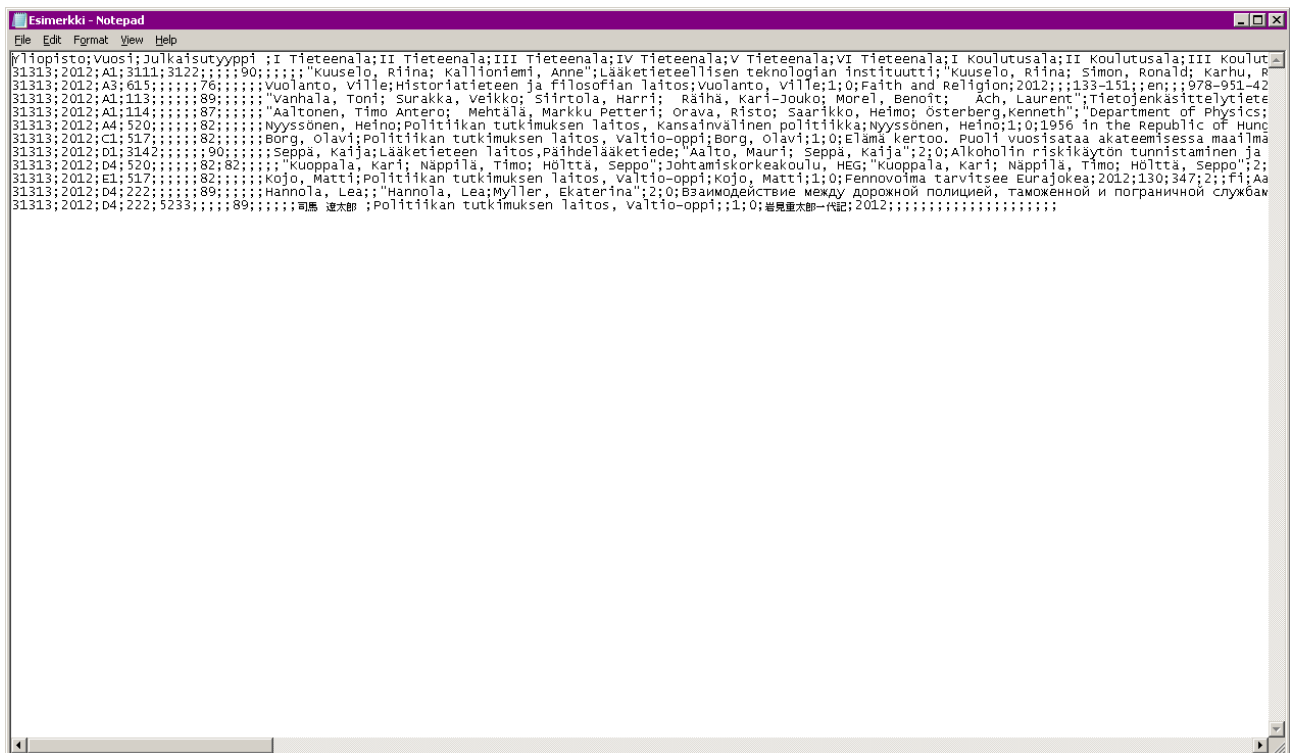
**Kuva 3. Varoitus CSV-muodon muotoilurajoituksista.**

Tämän jälkeen näytölle ilmestyy Export Text File -ikkuna (kuva 4), jossa vetovalikoista valitaan merkistökksi Unicode (UTF-8), kentän erottimeksi puolipiste (;) ja tekstierottimeksi lainausmerkki ("). Näiden valintojen jälkeen painetaan OK-näppäintä.



Kuva 4. Export Text File -valintaikkunan oikeat valinnat.

HUOM! Excel ei osa näyttää syntynyttä CSV-tiedostoa enää oikein (erikoismerkit korruptoituvat), joten jos haluaa varmistua lopputuloksen onnistumisesta syntynyt CSV-tiedosto pitää avata esim. Notepad-ohjelmalla, joka osaa näyttää UTF-8-merkistöä.



Kuva 5. Syntyneen CSV-tiedoston tarkastelu Notepad-ohjelmalla.

Tiedosto on nyt valmis ladattavaksi KOTAan. Ongelmatilanteissa voitte olla yhteydessä Manne Miettiseen ([manne.miettinen@csc.fi](mailto:manne.miettinen@csc.fi), gsm. 050 3819 510).