

Sosiaalista mediaa talteen?



Teemu Hänninen
IT-asiiantuntija
Xamk
Digitalia



Tuomo Räisänen
IT-asiiantuntija
Xamk
Digitalia

Kuvittele tilanne, jossa haluaisit säilöä oman sosiaalisen median sisältösi jälkipolville. Ensimmäisenä tulee mieleen, että siellähän kaikki on internetin ihmemaassa, ikuisesti, niin sanotaan. Asia ei ole näin, mistä muistutuksena esim. Yahoo Groups!, jonka pelastusoperaatiota käsiteltiin tarkemmin Failin numerossa 1/2020..

Jos haluat säilöä oman käyttöhistoriasi talteen, se onnistuu helpohkos- ti lataamalla tiedot omalta tililtäsi. Tällä hetkellä esimerkiksi Twitteris- sä voi tehdä näin max kaksi kertaa vuodessa. Saat pakatun tiedoston sähköpostiisi. Purettuasi tiedoston huomaat, että se on JSON-formaatissa, ei siis mitenkään helppolu- kuisessa muodossa. Oman twiitti- historian pitkäaikaistallentaminen on asia erikseen. Käyttötarkoitus on eri kuin alkuperäisellä sisällöllä.

Uudelleentwiittaukset ovat jonkun toisen tuottamaa sisältöä, joten niitä ei tule tallentaa luvatta. Jotta oma aineisto olisi jälkipolvienkin luettavissa, se tulisi vielä saattaa pitkäaikaistallennusmuotoon. Tässä vaiheessa useimmilla saattavat usko ja taidot loppua.

Sosiaalisen median käyttö ei ole ainakaan vähenemään päin, ja siksi sen talteenottoa on pohdittava ja edistettävä. Digialiassa vuosina 2019-2021 toteuttavassa Digitaalinen avoin muisti (DAM) -hankkeessa yksi työpaketeista omistautuu tälle aiheelle. Sosiaalinen media on jo arkipäivää eri viranomaistahojen tiedottamisessa ja asiakaspalvelussa. Esimerkiksi Kuntaliitto ja suuret kaupungit ovatkin ilmaisseet kiinnostuksensa Twitter-sisällön talteenottoon.

Eri kokoisilla kunnilla on vaihteleva määrä resursseja arkistoinnissa. Varsinkin pienimmissä kunnissa tuskin on intoa tai aikaa miettiä sosia-

lisen median pitkäaikaistallennusta. Kuntien taloudellisen tilanteen huomioon ottaen lisäkustannuksia- kaan ei kaivata. Niinpä lähtökohtana on ollut selvittää, voidaanko twiittejä kerätä ja tallentaa keskitetyksi, mieluiten avoimen lähdekoodin työkaluilla. Twitter tarjoaa API:n, jonka kautta yksittäisiä twiittejä voi koota ilmaiseksi, kunhan hankkii ns. Developer-tilin. Tilillä on toki rajoituksia, mutta tällä pääsimme alkuun.

Kuntakohtaisesti twiittejä on tallennettu organisaatioiden virallisilta tileiltä. Luottamushenkilöiden ja johtavien virkamiesten tilit vaativat henkilökohtaisen suostumuksen, mihin emme ole ryhtyneet, eikä ai- hettakaan tässä vaiheessa ole. Twiittejä riittää testiaineistoksi virallisilta tileiltäkin. Isoimmilla kaupungeilla voi olla useita virallisia tilejä.

Ratkaisun tulee olla sellainen, että tallennettavia tilejä voidaan lisätä, poistaa ja hallinnoida keskitetyksi. Tätä varten on kehitteillä web-



pohjainen käyttöliittymä. Noin sadalla kunnalla on vähintään yksi tili. Käyttöliittymässä on listattu kaikki kunnat, joilla on Twitter-tilejä, ja nämä tilit löytyvät viereisestä pudotusvalikosta, johon käyttäjä voi myös lisätä tilejä. Tietokannat nimettiin kaupunkien nimien mukaan. Yhden kaupungin tietokannassa voi olla useita tilejä.

Tätä kirjoitettaessa syksyllä 2020 olemme kokeiluun toteutettuna keränneet runsaat 200 000 twiittiä.

Tekninen ratkaisu perustuu olenaisesti ajastukseen sen sijaan, että seurattaisiin reaaliaikaisesti, milloin kukakin twiittaa. Tilejä käydään säännöllisesti läpi ja otetaan talteen uudet twiitit. Samalla päivittyvät esimerkiksi tykkäysten määrät ja muut ns. metriikat twiitistä. Kokeilujen perusteella näyttää siltä, että parin tuhannen twiitin kerääminen tunnissa onnistuu normaalilla PC:lla vaivattomasti. Koska twiittien kokonaismäärä päiväkohtaisesti on korkeintaan muutamia satoja, niin Developer-tilin rajoitukset eivät keräämisen suhteen ole vielä ongelma.

Tilien alustusten suhteen rajat sen sijaan tulevat vastaan. Taaksepäin voimme kerätä vain reilut 3000 twiittiä. Suomessa on jo kuntatilejä, jossa tämä määrä reippaasti ylittyy. Tähän on suunnitteilla ratkaisu, jossa tilin omistaja toimittaa meille itse lataamansa datan ja suoritamme sitten migraation. Tämä on vielä TODO-listalla hankkeessamme. Mutta lopputulema voisi olla, että tilinomistaja antaa meille luvan tallennukseen, toimittaa kerran oman twiittaushistoriansa, ja sen jälkeen homma toimisi automaattisesti.

Koska aineisto tallennetaan tietokantoihin, voidaan siihen kohdistaa hakuja, mikä saattaisi kiinnostaa esimerkiksi tutkijoita ja kuntaviestijöitä. Tämän jätämme kuitenkin tässä vaiheessa mahdollisissa jatkohankkeissa toteutettavaksi.

Mitä ratkaisumme sitten maksaisi kunnille? Käyttämämme ohjelmistot ovat ilmaisia, laitteistoja on kuitenkin ylläpidettävä. Sen suhteen puhutaan tilikohtaisesti ehkä kymmenistä euroista vuodessa. Hinnan ratkaisee lopullisesti se, kuinka paljon ohjelmiston ylläpito vaatii ihmistyötä. Selvitämme asiaa.

Lisätietoja Twitter-tallennusratkaisusta saa jutun kirjoittajilta Digitaliasta.