

Informaatiotutkimuksen päivät 2014
6. - 7. marraskuuta, Oulun yliopisto, Oulu

ABSTRAKTI

Eeva Savolainen

Radio- ja televisioaineiston kuvailun haasteita ja mahdollisuuksia Kansallisen audiovisuaalisen instituutin radio- ja televisioarkistossa

Eeva Savolainen, Kansallinen audiovisuaalinen instituutti, eeva.savolainen@kavi.fi

Digitaalisen tekniikan kehitys ja vuoden 2008 alussa voimaan tullut Kulttuuriaineistolaki (L 28.12.2007/1433) mullistivat suomalaisen radio- ja televisioaineiston arkistoinnin ja tutkimusmahdollisuudet. Kansallinen audiovisuaalinen instituutti (KAVI) on tallentanut vuoden 2009 alusta lähtien keskeisten kotimaisten televisio- ja radiokanavien ohjelmavirran kokonaisuudessaan. Muilta kotimaista ohjelmistoa lähettäviltä kanavilta tallennetaan vuosittain viikon näyte. Tietyt ohjelmatyypit kuten dokumentit ja draama tallennetaan arkistoon myös alkuperäislaatuina, sillä digitaalinen kaappaustallenne on muodoltaan pakattua, eikä siten sovellu uusintaesityksiin tai ohjelmatuotantoon.

Radio- ja televisiotietokanta Ritva

Radio- ja televisioarkiston Ritva-tietokanta avattiin yleisölle marraskuussa 2010. (Savolainen 2011, 200-201.) Katselu- ja kuuntelupisteitä on vuonna 2014 pääsääntöisesti yliopistokirjastojen yhteydessä. (KAVIn verkkosivut) Ritvaan kertyy vuosittain noin miljoona lähetystä. 1.10.2014 lähetystä oli 6,2 miljoonaa, joista suoratoistollisia lähetystä oli 2,7 miljoonaa. Loput tietueet sisältävät metatietoa ohjelmista ja lähetyksajoista. Kotimaisten sarjojen kartunta on noin 300 kpl vuodessa. Alkuperäistason talletusten määrä vaihtelee: esimerkiksi vuonna 2013 talletettiin 4200 ohjelmaa (KAVA tulokset 2013).

Ritvan metadata koostuu eri lähteistä tiedonsiirtoina saadusta metadatasta ja luetteloiden käsin tuottamasta metadatasta. Keskeistä tiedonsiirtona saatavaa metadataa tarjoavat kotimaisten televisio- ja radiokanavien yhteinen tietojärjestelmä Venetsia, radiokuuntelua ja televisionkatselua mittaava Finnpanel sekä elektroniset ohjelmaoppaat (Savolainen 2011, 200). Yleisradiolta saadaan tekijä- ja yhtiötietoja. Lisäksi ohjelmatietojen siirroista on tehty sopimuksia yksittäisten kanavien kanssa. Tiedonsiirtoina saadaan tietoa ohjelmien sisällöistä (esim. nimi, jaksokohtainen kuvaus, genre) sekä lähetystiedoista (kanava, päivämäärä, alkamis- ja loppumisaika). Aineiston määrä ja automaattinen metadata tarjoavat kuvailutyölle sekä haasteita että mahdollisuuksia.

Automatisoinnin ansioita ja ongelmia

Automatisoidun tallennuksen ja metatietoprosessin suurimpia ansioita on nopeus: tallennetut ohjelmat saadaan asiakkaiden käyttöön noin kolmen päivän kuluttua ohjelman lähetyksestä,

kun perustietoja ei tarvitse syöttää järjestelmään käsin. Kääntöpuolena ovat järjestelmään jäävät virheet, joita tekniset apuvälineet tai pistokokein tehtävä laadunvarmistus eivät havaitse: aineiston määrä on niin suuri, ettei jokaista tietuetta ja tallennetta ole mahdollista käydä yksityiskohtaisesti lävitse. Asiakkailla on kuitenkin mahdollisuus antaa huomaamistaan virheistä palautetta, jonka perusteella järjestelmään tehdään korjauksia ja huomautuksia.

Automaattisen metadatan ongelmia ovat ristiriitaisuuksista johtuvat virheet ja epätasaisuus. Ohjelmanmuutoksista ja eri lähteistä saatujen tietojen ristiriitaisuuksista johtuvat virheet korjataan käsin, sillä tekniset apuvälineet eivät toistaiseksi pysty tulkitsemaan esimerkiksi kaikkia lähetyisaikojen muutoksista johtuvia virheitä. Eri kanavilta saadaan yksityiskohdiltaan hyvin erityyppistä metadataa, ja joissakin harvoissa tapauksissa metadataa ei saada lainkaan. Metatiedon suurpiirteisyys tai sen puutekaan ei estä asiakkaiden pääsyä tallenteisiin, mutta vaikuttaa eri kanavien ohjelmistojen haettavuuteen.

Automaattisesti saatavat ohjelmien perustiedot vapauttavat parhaimmillaan työaikaa syvällisemmälle sisällönkuvailulle ja tarjoavat hyvän lähteen sisällönkuvailutyöhön. Koska perustiedot saadaan automaattisesti, voidaan keskittyä sisältötietoihin. Toistaiseksi sisällönkuvailussa on keskitytty parantamaan erityisesti kotimaisten ohjelmien haettavuutta.

Kotimaisten ohjelmien kuvailu

Kotimaisiin pisteohjelmiin ja sarjoihin lisätään tuotantomaaksi Suomi. Lisäksi jokaiselle kotimaiselle sarjalle luodaan sarjakohtainen kuvaus, johon etsitään tietoja automaattisten metadatalähteiden lisäksi mm. lopputeksteistä, lehdistä ja internetistä. Kuvaukseen kerätään tiedonhakua tukevia tietoja kuten sarjan tuotantovuosi, tuotantomaa, kuvaus sisällöstä, asiasanat, keskeiset tekijät ja tuotantoyhtiöt. Automaattisesti saatujen jaksokohtaisten kuvausten pohjalta on mahdollista koota koko sarjaa koskeva sisällönkuvaus nopeammin kuin esimerkiksi ohjelmia katselemalla.

Järjestelmässä on käytössä ohjelmien nimien ja kuvausten perusmuotoistaminen, mikä vähentää katkaisumerkin käytön tarvetta hauissa. Kuvailutyön kuten asiasanoituksen avuksi perusmuotoistamisesta ei sen sijaan toistaiseksi ole, sillä koneen sisällönkuvauksesta keräämät sanat eivät pysty ainakaan sarjatasolla tavoittamaan sisällön monimuotoisuutta yhtä hyvin kuin asiantuntijan tekemä asiasanoitus.

Alkuperäistasoisena talletetut ohjelmat kuvaillaan KAVIn materiaalitietokanta Tenhoon, josta osa kuvailutiedosta siirretään Ritvaan. Koska Ritva ja Tenho toimivat toistaiseksi omina hakujärjestelminään, on tärkeää että molemmissa tietokannoissa on riittävästi tiedonhakua tukevia tietoja. Tiedonsiirroilla pyritään välttämään samojen tietojen syöttämistä kahteen eri kertaan.

Tulevaisuuden suuntia

Käyttäjänäkökulmasta Ritva tarjoaa jo nyt monipuolista ohjelmatietoa ja useita hakumahdollisuuksia. Järjestelmän käytettävyydessä on kuitenkin kehittämistä, kuten tekninen suunnittelija Tommi Lehtonen on todennut opinnäytetyössään (Lehtonen 2014). Kuvailun näkökulmasta Ritva-tietokannan tulevaisuuden keskeinen haaste on aineiston löydettävyyden turvaaminen. Tähän antavat mahdollisuuksia olemassa olevat ja uudet metadatalähteet. Lisäksi tarvitaan uudenlaisia hakua ja ohjelmatietojen ryhmittelyä tukevia teknisiä ominaisuuksia. Tulevaisuudessa sisällönanalyysiin toivotaan lisätukea esimerkiksi kehittyvistä puheen- ja kuvantunnistusmenetelmistä.

Lähteet

KAVA tuloraportti (2013). <http://intra/system/files/KAVA%20tuloraportti%202013%20.pdf> (29.8.2014) Käytettävissä KAVIn verkossa.

KAVIn verkkosivut. <https://kavi.fi/fi/rtva/ohjelmien-katselu-ja-kuuntelu/ohjelmavirta> (28.8.2014)

L 28.12.2007/1433. Laki kulttuuriaineistojen tallettamisesta ja säilyttämisestä. Valtion säädöstietopankki Finlex, Ajantasainen lainsäädäntö. <http://www.finlex.fi> (28.8.2014)

Lehtonen, Tommi (2014). User Experience Evaluation of the Radio and Television Media Management System 2014
http://publications.theseus.fi/bitstream/handle/10024/71335/TommiLehtonenThesis_2014_50.pdf?sequence=1 (28.8.2014)

Savolainen, Eeva (2011). Guarding the heritage: radio and television culture preservation in Finland. *Journal of Scandinavian Cinema* 1 (2) (2011): 199-203.