

Linkkimätä + sisältöryjähdys = viitemätä

Merja Kallio

Suurin osa netin käyttäjistä on linkkiä seuratessaan saanutkin toivomansa sisällön sijaan virheilmoituksen, useimmiten 404 - sivua ei löydy. Tieteen todistettavuus ja toistettavuus kärsii, jos lähteenä käytettyjä lähteitä ei löydy.

Osallistuin avoimen julkaisemisen koulutustilaisuuteen *Avoimen julkaisemisen parhaat käytänteet -koulutus tiedelehtien tekijöille* 30.–31.1.2017, jossa Kansalliskirjaston erityisasiantuntija Juha Hakalan vetämä työpaja e-viittaamisen käytännöistä eri tieteenaloilla¹ jäi erityisesti mieleeni. Myöhemmin etsin aiheesta lisätietoa, jota ei kuitenkaan juurikaan löytynyt suomeksi.

Käytän tässä artikkelissa Juha Hakalan käyttämiä termejä linkkimätä (*link rot*) ja sisältöryjähdys (*content drift*) ja samalla kiitän häntä inspiaraatiosta lisätiedon hakemiseen.

404 - sivua ei löydy

Linkkimätää tai linkkiruttoa on varmasti ollut yhtä pitkään kuin hyperlinkkejä on käytetty, ja sitä on tutkittukin pitkään. Linkkimädän syinä ovat esimerkiksi sivustojen uudistukset, organisaatio- ja nimimuutokset tai verkko-osoitteiden muutokset. Joissain tapauksissa verkkosivuja ei vain enää ylläpidetä. Sosiaalisen median käyttäjä saattaa poistaa tuottamansa sisällön tai muuttaa sen yksityisyysasetuksia.

Verkkosivun keskimääräinen elinikä on vain 44–100 päivää². Tieteellisten artikkeleiden suhteen ongelma tunnistettiin suhteellisen aikaisin ja ratkaisuksi kehitettiin pysyvät tunnisteet, kuten DOI, URN ja handle.³

Ongelma onkin laajemmassa verkosta löytyvässä aineistossa. Paineet paperittomaan julkaisemiseen kasvavat ja esimerkiksi ministeriöiden ja asiantuntijalaitosten raportit siirtyvät vain elekt-

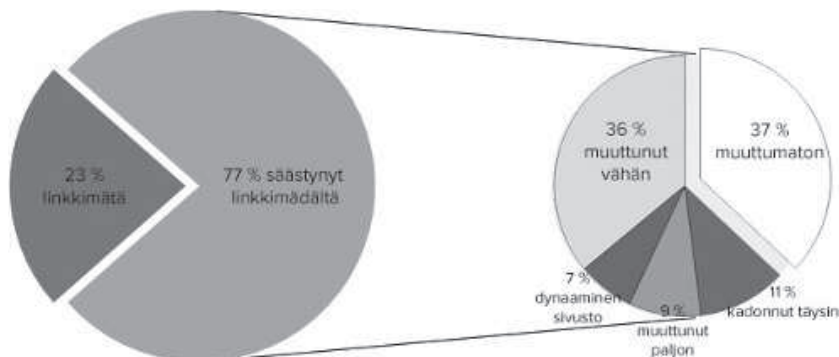
roniseen muotoon. Tutkimushankkeiden blogit herättävät keskustelua kommentoissa.

Samalla kun elämämme siirtyy verkkoon, siirtyy myös tutkimusaineisto ja tutkimuksen lähdeaineisto. Viittausten määrä muuhun verkkoaineistoon kuin tieteellisiin artikkeleihin, *URI references*, kasvaa koko ajan.^{2,4} Koska tieteellisillä artikkeleilla on usein pysyvä tunniste ja niihin suositellaan viitattavaksi hyperlinkin muodossa, käytän tässä muista verkkoaineistoviittauksista sanaa URI-viitteet*.

Linkkimädän toteaa myös tavallinen käyttäjä helposti. Sen tutkiminen voidaan automatisoida, kuten esimerkiksi *Hiberlink* -projektissa⁴, jossa tutkittiin artikkeleita vuosilta 1997–2012, tai Massicotten ja Botterin tutkimuksessa väitöskirjojen URI-viittauksista⁵. Massicotte ja Botter tutkivat kanadalaisen Concordian yliopiston väitöskirjoja ajalta 2011–2015. Tutkimuksen mukaan 23 % URI-viittauksista kärsii linkkimädästä. Automaattisessa seulonnassa toimivina URI-viittauksina kuitenkin näyttäytyvät esimerkiksi myynnissä olevat verkko-osoitteet tai räätälöidyt 404-virheilmoitussivut.

Eihän tämä olekaan se sama

Sisältöryjähdys eli *content drift* on sivuston käyttäjälle vaikeammin havaittava ongelma kuin linkkimätä. URI-viittaus näyttää toimivan ja se saattaa viedä sivustolle, jonka sisältöä käyttäjä odotti löytävänsä. Kuitenkin vain viittauksen kirjoittaja voi tietää täysin, millaista sivustoa hän on kir-



Kuvio 1. Linkkimätä ja sisältöryjähdyksen yhteisvaikutus Concordian yliopiston väitöskirjoissa

joittamishetkellä katsonut.

Sisältöryjähdyksen tavallisinta usein päivittyvillä sivustoilla, kuten uutissivustot. Mikäli sivu on muuttunut viimeisen viikon aikana, voidaan olettaa sen muuttuvan yhtä paljon seuraavan viikon aikana⁴.

Jopa 65 % sivustoista muuttuu melko usein, 123 tunnin välein. Muutoksia on enemmän ja useammin suosituilla sivustoilla⁴ ja pitkissä dokumenteissa⁶.

Jones SM ym.⁶ vertasivat Hiberlink -projektin aineiston URI-viitteitä verkkoarkistoista löytyviin versioihin julkaisuajankohdalta ja toteivat, että 184.065 URI-viitteen sisältö kaikkiaan 241.091 viitteestä oli muuttunut julkaisu- ja tutkimusajankohtien välillä. Toisin sanoen URI-viitteistä ainoastaan 23,65 % oli pysynyt sisällöllään samana.

Massicotten ja Botterin⁵ tutkimuksen tulos oli samansuuntainen. Väitöskirjoissa linkkimädältä vältyneistä URI-viittauksista 11 % sisältö oli kadonnut täysin, 9 % ohjasi sivustolle, jonka sisältö oli muuttunut niin paljon, ettei viittauksista voitu enää tutkia, ja 7 % URI-viittauksista ohjasi dynaamisille, jatkuvasti muuttuvilla sivustoilla. Reilu kolmasosa, 36 % sivustoista oli muuttunut vähän, mutta ulkopuolisen on vaikea tietää, onko juuri muutos tapahtunut juuri viittauksen kohteena olleessa sisällössä.

Viitemätä

Linkkimädän ja sisältöryjähdyksen yhteisvaikutus

on *reference rot*, kutsutaan sitä tässä yhtenäisen linjan säilyttämiseksi viitemädäksi*. Termiä käytettiin ensimmäisenä Hiberlink -projektissa⁶.

Viitemätä aiheuttaa sen, ettei lähteenä käytetyn URI-viittauksen nykyinen sisältö vastaa viittausajankohdan mennyttä sisältöä.

Artikkeli voi olla *immuuni*, mikäli siinä ei ole lainkaan URI-viittauksia; *terve*, mikäli kaikki artikkelin URI-viittaukset ovat toimivia ja jokainen erillinen URI-viittaus on arkistoitu; tai *saastunut*, mikäli jokin URI-viittaus ei toimi tai jokaista URI-viittauksista ei ole arkistoitu. Saastuneiden artikkeleiden osuus URI-viittauksista sisältävistä artikkeleista on jopa 70–80 %.⁴

Ilman verkkoarkistointia URI-viittauksella on riski kadota jossain vaiheessa. Väitöskirjojen URI-viittauksista tämä riski on noin kolmanneksella⁵. Avoimen tieteen periaatteet tutkimuksen tuotosten avoimuudesta ja jatkokäytöstä primärilähteistään vaarantuvat⁷

Viitemätä on ongelma, joka koskee kaikkea tieteellistä julkaisemista ja jota tulisi aktiivisesti ennaltaehkäistä.

Ratkaisu

Viitemätää voidaan torjua lisäämällä tietoisuutta ongelman laajuudesta ja URI-viittauksien riskejä sekä ohjaamalla ja tarjoamalla työkaluja tutkijoiden ja kirjastojen käyttöön.

Pysyvää tunnistetta, kuten DOI, handle tai URN, tulisi aina käyttää URI-viittauksissa, mikäli sellainen on saatavilla. Pysyvää tunnistetta



Kuvio 2. WebCite-palvelun virheilmoitus

käyttävä julkaisija on sitoutunut säilyttämään aineiston saatavilla verkkotunnuksista tai sivuston rakenteen muutoksista huolimatta. Oman kokemukseni mukaan tutkijoiden tiedot pysyvästi tunnisteista ovat rajallisia tietokannoissa käytössä olevia DOI-tunnisteita lukuun ottamatta, eikä niiden hyötyjä täysin tunnusteta lähteisiin viitattaessa.

Mikäli pysyvää tunnustetta ei ole, suositellaan verkkoarkistojen käyttöä. Verkkoarkistoihin tallennetaan yleensä kuvakaappaus (*snapshot*) sivusta, joten sivulta eteenpäin vievät linkit eivät toimi. Verkkoarkistoista voi etsiä viittausajankohdalle vastaavaa versiota. Tällaisesta verkkoarkistosta yksi esimerkki on *Internet Archive Wayback Machine* <https://archive.org/web/>.

Passiivisessa haussa valmiiksi tallennetuista arkistoista on kuitenkin muutama ongelma: verkkoarkistosta ei välttämättä löydy juuri oikeata versiota viitattusta verkkosivusta tai verkkoarkistosta ei enää ole käytössä.

Verkkoarkistosta etsimistä parempi vaihtoehto on oma aktiivisuus viittaushetkellä. URI-viitteen voi tallentaa itse viittaushetkellä ja varmistaa, että lukijan saatavilla on myöhemminkin sivustosta täysin sama versio. Tätä artikkelia kirjoittaessani testasin lähteissäni suositeltua WebCitation-sivustoa, joka näytti tallentavan URI-viitteeni ja viiveen jälkeen sain sähköpostin, jossa oli linkki arkistoon. Valitettavasti linkkiä seurattessani sain vastaan virheilmoituksen, joka ei muuttunut sivua päivittämällä, selaimen vaihdolla tai linkkiä myöhemmin seuraamalla.

Suosittelavin vaihtoehto olisi hyödyntää *Zotero*

-viitteidenhallintaohjelmaa ja tallentaa URI-viitteestä verkkoarkistoon kuvakaappaus (*snapshot*) heti viittauksen tehdessään. Käyttöä voi helpottaa *Zoteron* selaimen laajennuksella⁴, joka on saatavilla *Chromelle*⁸ ja *Firefoxille*. Testatessani *Zoteroa* se tuntui toimivan paremmin *Firefoxilla*, mutta käyttökokemus saattaa johtua myös siitä, että *Firefoxin* käyttö on minulle luontevampaa.

Zotero ei käyttöliittymältään kuitenkaan ole helpoiten lähestyttävissä oleva viitteidenhallintaohjelma. Muita ohjelmia käytettäessä URI-viitteen voi tallentaa esimerkiksi *archive.is* -sivustolle osoitteessa <http://archive.is/>.

Tutkimushanke tai väitöskirjan kirjoitusprosessi saattaa viedä vuosia. Huomio URI-viittauksiin tulisivin kiinnittää heti projektin alusta lähtien. Verkkoarkistointi ja oikean URI-viittaus-tapa kannattaa ottaa käyttöön heti aineistonkeruun kanssa. Muuten toimittajalla tai kustantajalla saattaa olla jo ennen julkaisuprosessia käsissään aineistoa, jonka URI-viittaukset kärsivät viitemädästä. Jos URI-viittaukset jätetään julkaisujankohdan jälkeen kirjaston tai arkiston harteille, saattaa osa URI-viittauksista olla jo siinä vaiheessa menetetty.

*) En löytänyt sanalle suomenkielistä vastinetta.

Lähteet

¹ Hakala, Juha (2017). *E-viittaamisen uudet käytännöt* (2017). Noudettu 7. huhtikuuta 2017, osoitteesta <http://urn.fi/URN:NBN:fi-fe201702061462>

² Habibzadeh, P. (2013). *Decay of References to Web sites in Articles Published in General Medical Journals: Mainstream vs Small Journals*. *Applied Clinical Informatics*, Vol. 4: Issue 4 2013. Noudettu 7. huhtikuuta 2017, osoitteesta <https://doi.org/10.4338/ACI-2013-07-RA-0055>

³ Hans-Werner Hilse and Jochen Kothe (2006). *Implementing Persistent Identifiers: Overview of concepts, guidelines and recommendations*. London / Amsterdam: Consortium of European Libraries and European Commission on Preservation and Access, 2006. ISBN 90-6984-508-3. Noudettu 7. huhtikuuta 2017, osoitteesta <http://nbn-resolving.de/urn:nbn:de:gbv:7-isbn-90-6984-508-3-8>

⁴ Klein, M. (2014). *Scholarly Context Not Found: One in Five Articles Suffers from Reference Rot*. (Research Article).

PLoS ONE, 9(12). Noudettu 7. huhtikuuta 2017, osoitteesta <http://dx.doi.org/10.1371/journal.pone.0115253>

^{5.} Massicotte, Mia and Botter, Kathleen (2017). *Reference Rot in the Repository: A Case Study of Electronic Theses and Dissertations (ETDs) in an Academic Library*. Information Technology and Libraries, Vol 36, No 1. Noudettu 7. huhtikuuta 2017, osoitteesta <http://dx.doi.org/10.6017/ital.v36i1.9598>

^{6.} Jones SM, Van de Sompel H, Shankar H, Klein M, Tobin R, et al. (2016). *Scholarly Context Adrift: Three out of Four URI References Lead to Changed Content*. PLOS ONE 11(12): e0167475. Noudettu 7. huhtikuuta 2017, osoitteesta <http://dx.doi.org/10.1371/journal.pone.0167475>

^{7.} Avointiede - Tutkimuksen toistettavuus (2015). Noudettu 11. huhtikuuta 2017, osoitteesta <http://avointiede.fi/tutkimus-pas>

^{8.} Zotero Blog » Blog Archive » A Better “Save to Zotero” Button in Chrome (2016). Noudettu 11. huhtikuuta 2017, osoitteesta <https://www.zotero.org/blog/a-better-save-to-zotero-button-in-chrome/>
<https://www.zotero.org/groups/lahdemata/items/>

Tietoa kirjoittajasta:

Merja Kallio

Vaasan yliopiston ja Vaasan ammattikorkeakoulun julkaisusihteeri, Tritonia.

merja.kallio@tritonias.fi