



Sähköiset aineistot historian opin- näytetöissä

Tiina Valonen tarkastelee internetlähteiden käyttöä Turun yliopiston Suomen historian vuosina 2000–07 valmistuneissa opinnäytetöissä.

■ Nykyään internet on keskeinen väline opiskelussa. Laajakaistaliittymien ja langattomien verkkoyhteyksien yleistyminen ja kohtuulliset käyttö hinnat ovat mahdollistaneet nopean ja helpon tiedonhaun internetistä. Verkosta löytyy monipuolisesti tietoa erilaisista aiheista. Tietoverkkojen käyttö murtaa perinteistä paikkasidonnaisuutta: tietoa ei enää tarvitse lähteä hakemaan kirjastosta tai arkistosta, vaan informaatiota voi etsiä kotitietokoneella.

Tieteellisen tutkimuksen kannalta internetin toimintaan ja käyttöön liittyy kuitenkin omat ongelmansa. Tietoa on

tarjolla paljon, mutta takeita sen luotettavuudesta on vaikea saada. Luotettavuuden ohella tiedon säilyminen ja muuttumattomuus verkkoympäristössä herättää kysymyksiä.

Käsitellen¹ seuraavassa internetlähteiden käyttöä Turun yliopiston Suomen historian opinnäytetöissä. Mukana tarkastelussa on 101 vuodesta 2000 vuoden 2007 lokakuun alkuun mennessä valmistunutta pro gradu -tutkielmaa ja sivuainelaudaturtyötä. Yhtäältä tarkastelen, kuinka paljon internetlähteitä on käytetty ja keiden ylläpitämiltä sivustoilta tietoa on etsitty. Toisaalta käsitellen sitä, miten tutkielmissa käytetyt aineistot ovat säilyneet internetissä. Toimivatko kirjoittajien ilmoittamat osoitteet edelleen, ovatko tutkielmissa käytetyt tiedot löydettävissä verkosta ja ovatko aineistot säilyneet joidenkin ylläpitäjien sivuilla toisia paremmin? Käytettyjä sivustoja on voitu muokata ajan kuluessa, ja sivustot voivat olla nykyään erilaisia kuin opinnäytetyöntekijöiden tutustuessa niihin. Tiedon säilymistä tarkastelen asiasisällön kannalta vertailemalla nykyistä sisältöä opinnäytetyössä hyödynnettyihin tietoihin.

Internetlähteiden käyttö

Internetissä olevien aineistojen käyttö yleistyi Suomen historian opinnäytetöissä 2000-luvulla. Keskimäärin joka toisessa tutkielmassa oli hyödynnetty verkosta löytyvää materiaalia: 101 tutkielmasta 45:ssä oli käytetty internetlähteitä.

Internetlähteiden käytössä trendi oli selvästi kasvava. Kun esimerkiksi vuonna 2000 valmistuneista tutkielmista vajaassa viidesosassa oli hyödynnetty verkkolähteitä, niin kaikissa vuonna 2007 palautetuissa opinnäytetöissä oli käytetty internetlähteitä. Ainoan poikkeuksen nousevaan trendiin muodosti vuosi 2005, jolloin internetlähteiden käyttöprosentti laski.

Myös viittausten määrä eri osoitteisiin lisääntyi tutkielmissa vuosina 2000–07. Vuonna 2000 valmistuneissa opinnäytetöissä, joissa oli käytetty internetlähteitä, tietoa oli haettu keskimäärin kahdesta eri osoitteesta. Vuoden 2007 tutkielmissa tietoa oli etsitty keskimäärin neljästä osoitteesta.

1. Tiina Valonen *Sähköiset aineistot Suomen historian vuosina 2000–2007 valmistuneissa opinnäytetöissä*. Julkaisematon selvitys. Turun yliopiston Suomen historian oppiaine, 2007.

Naiset käyttivät verkon tarjoamia mahdollisuuksia tiedonhakuun selvästi useammin kuin miehet; naisista noin joka toinen käytti internetlähteitä opinnäytetyössään. Miehistä puolestaan alle 40 prosenttia käytti internetlähteitä. Sukupuolten välinen ero näkyi myös käytettyjen osoitteiden määrissä. Naisten kirjoittamissa tutkielmissa, joissa oli hyödynnetty sähköistä aineistoa, viitauksia eri osoitteisiin oli keskimäärin neljä. Miehillä vastaava luku oli kaksi.

Tiedon säilyminen

Internetosoitteiden toimivuutta tarkastelin luokittelemalla osoitteet kolmeen ryhmään. Toimiviin osoitteisiin jaottelin tarkat osoitteet, jotka toimivat suoraan, ja osoitteet, joista tutkielmissa oli annettu ainoastaan pääsivun osoite, mutta joilta käytetty tieto löytyi. Toisen ryhmän muodostivat osoitteet, joiden toiminnassa ilmeni ongelmia. Ryhmään kuului monenlaisia tapauksia. Muutamat osoitteet ohjautuivat uusille sivustoille. Joissakin tapauksissa tarkka osoite ei toiminut, vaikka pääsivut olivat edelleen toiminnassa. Osa sivustoista ei ollut enää toiminnassa tai ne olivat siirtyneet uuden ylläpitäjän käyttöön. Kolmannen ryhmän puolestaan muodostivat osoitteet, joita ei voinut luokitella osoitteen toimivuuden perusteella. Näissä tapauksissa lähteen internetosoitetta ei ollut merkitty opinnäytetöihin. Lisäksi yksi osoite oli selvityksen tekoaikana tilapäisesti poissa käytöstä.

2000-luvulla valmistuneissa Suomen historian opinnäytetöissä tietoa oli etsitty yhteen-

sä 160 internetosoitteesta, joista 93 toimi, 63:ssa ilmeni ongelmia ja neljää ei voinut luokitella. Tapauksia, joita ei voinut luokitella, ei ole otettu huomioon tarkasteltaessa tiedon säilymistä.

Osoitteiden toimivuus oli sidoksissa lukuaikaan. Esimerkiksi viidennes vuonna 2000 etsityistä lähteistä toimi ilman ongelmia, kun taas vuonna 2006 luetuista osoitteista melkein 90 prosenttia toimi ilman ongelmia.

Se, että osoitteiden toiminnassa ilmeni ongelmia, ei välttämättä merkinnyt sitä, että tiettyä tietoa ei olisi enää mahdollista löytää sivustoilta. Tiedon saattoi löytää etsimällä, jos sivustot olivat edelleen toiminnassa ja saman ylläpitäjän käytössä. Jos osoite ohjautui uudelle sivustolle, tieto oli mahdollista löytää joko automaattisesti tai etsimällä. Esimerkiksi vuonna 2003 luetuista 39 osoitteesta 19 osoitteen toiminnassa ilmeni ongelmia. Yhdeksässä tapauksessa tieto löytyi joko sivustolta uudesta paikasta tai uudelta sivustolta. Koko tarkasteluajan 63 osoitteesta, joiden toiminnassa ilmeni ongelmia, käytetyistä tiedoista 27 oli mahdollista löytää sivustoilta. Sen sijaan luetuista tiedoista 36 oli kadonnut.

Opinnäytetöissä käytetyistä 156 internetlähteestä 120 tapauksessa tieto löytyi joko suoraan tai etsimällä. Toisin sanoen 77 prosenttia käytetyistä tiedoista löytyi edelleen sivustoilta ja 23 prosenttia oli kadonnut. Tiedon säilymisen ja lukuajan yhteys tuli selkeästi esille. Kadonneiden aineistojen suhteellinen osuus laski

tasaisesti tultaessa lähemmäs nykyhetkeä. 2000-luvun alkupuolella luetuista tiedoista oli kadonnut 50–60 prosenttia. Vuonna 2004 luetuista tiedoista viidesosaa ei ollut enää mahdollista löytää sivustoilta, ja vuoden 2006 aikana luetuista aineistoista vain neljää prosenttia ei ollut löydettävissä sivustoilta.

Ylläpitäjien sivustoilta kadonneita tietoja on mahdollista etsiä internetarkistoista. Käytin kadonneiden tietojen etsimiseen yhdysvaltalaisista arkistoa nimeltään Wayback machine, osoitteesta <http://www.archive.org>. Sivustoille on kerätty internetaineistoa vuodesta 1996 alkaen. Sivustolla voi hakukenttään kirjoittaa etsityn osoitteen, ja hakutulokset ilmoitetaan sivustojen päivitysaikojen mukaan.

Kokonaisuudessaan 36 kadonneesta osoitteesta 17 löytyi arkiston sivuilta. Sen sijaan 19 osoitteesta ei sivuilta löytynyt. Näistä 19 tapauksesta viisi osoitetta tuotti hakutuloksen, mutta opinnäytetöissä luettuja tietoja ei löytynyt. Esimerkiksi kahdessa tapauksessa käytetyn aineiston otsikko löytyi, mutta itse aineistoa ei ollut mahdollista lukea. Yksi yhdistys oli kieltänyt sivustojensa tallentamisen arkistoon.

Kokonaisuudessaan käytetyistä tiedoista oli kadonnut 12 prosenttia sekä ylläpitäjien sivustoilta että Wayback machinesta. Lukuajan vaikutus oli suuri. Vuonna 2000 luetuista tiedoista oli kadonnut 20 prosenttia. Sen sijaan vuonna 2006 ja 2007 luetuista tiedoista kaikki löytyivät joko ylläpitäjien sivuilta tai Wayback machines-ta.

Ylläpitäjät

Internetlähteitä oli opinnäytetoissa käytetty ennen kaikkea tutkimuskirjallisuutena. Yksittäisiin kysymyksiin liittyvää yleistietoa oli hyödynnetty runsaasti. Myös tieteellistä kirjallisuutta oli luettu paljon. Verkosta oli etsitty myös jonkin verran kuvamateriaalia. Sen sijaan tutkimuksen alkuperäisaineistona internetlähteitä oli käytetty hyvin vähän.

Hyödynnetyistä internetlähteistä suurin osa oli luettu kotimaisten ja ulkomaalaisten yliopistojen ylläpitämiltä sivustoilta. Melkein neljäsosa kaikista aineistoista oli haettu näiltä sivustoilta. Tietoa oli etsitty melko paljon tieteellisten yhteisöjen, valtionhallinnon, kaupallisten organisaatioiden sekä erilaisten liittojen, järjestöjen, seurojen, säätiöiden ja yhdistysten ylläpitämiltä sivustoilta. Jokaisen näiden neljän ylläpitäjäröhmän osuus oli 12–13 prosenttia kaikista käytetyistä internetlähteistä. Lisäksi medioiden ylläpitämiltä sivustoilta tietoja oli etsitty jonkin verran. Sen sijaan vain vajaa viisi prosenttia aineistoista oli etsitty joko kuntien tai arkistojen, kirjastojen ja museoiden ylläpitämiltä sivuilta. Uskonollisten yhteisöjen ja yksityishenkilöiden ylläpitämille sivuille oli tehty vain muutamia viittauksia.

Parhaiten opinnäytetoissa käytetyt tiedot olivat säilyneet yksityishenkilöiden ja uskonollisten yhteisöjen ylläpitämiltä sivustoilla. Tosin näille sivustoille viittauksia oli tehty hyvin vähän. Yli 90 prosenttia niin tieteellisten yhteisöjen kuin medioiden ylläpitämiltä sivustoilta haetuista tiedoista

oli edelleen luettavissa näillä sivustoilta. Valtionhallinnon sekä arkistojen, kirjastojen ja museoiden sivustoilla tiedoista oli säilynyt 80 prosenttia. Yliopistojen, kaupallisten organisaatioiden sekä liittojen, järjestöjen, seurojen, säätiöiden ja yhdistysten sivuilta tiedoista oli puolestaan säilynyt keskimäärin 70 prosenttia. Sen sijaan kuntien ylläpitämiltä sivustoilla tiedot olivat säilyneet huonoiten, vajaa 60 prosenttia opinnäytetoissa hyödynnetyistä tiedoista oli enää löydettävissä.

Etsittäessä kadonneita tietoja ylläpitäjien sivustoilta Wayback machinesta näiden aineistojen osuudet laskivat. Ylläpitäjien sivustoilta sekä Wayback machinesta löytyi yhteensä 95 prosenttia tieteellisten yhteisöjen, 92 prosenttia medioiden, 92 prosenttia yliopistojen, 90 prosenttia valtionhallinnon, 89 prosenttia kaupallisten organisaatioiden, 85 prosenttia arkistojen, kirjastojen ja museoiden, 85 prosenttia kuntien ja 74 prosenttia liittojen, järjestöjen, seurojen, säätiöiden ja yhdistysten sivuilta luetuista tiedoista.

Internetin arkea lähdetutkimuksessa

Internetlähteiden käytön yleistyminen Suomen historian opinnäytetoissa 2000-luvulla on ymmärrettävää, koska internetaineistojen tarjonta on monipuolistunut. On todennäköistä, että internetlähteiden käyttö tulee opinnäytetoissa tulevaisuudessa lisääntymään, koska yhä enemmän tutkimuskirjallisuutta ja alkuperäisaineistoja on saatavissa sähköisessä muodossa.

Tiedon säilyminen verkkoympäristössä on asia, jota tulisi pohtia. Melkein neljäsosa opinnäytetoissa käytetyistä internetistä luetuista tiedoista oli hävinnyt kirjoittajien ilmoittamilta sivustoilta. Noin puolet ylläpitäjien sivustoilta kadonneista aineistoista oli mahdollista löytää internetarkistosta. Toisin sanoen kymmenesosa käytetyistä tiedoista oli kadonnut sekä ylläpitäjien sivustoilta että Wayback machinesta. Tämä heikentää olennaisesti tutkielmissa käytettyjen tietojen jäljitettävyyttä ja mahdollisuuksia myöhemmin arvioida tutkielmien tietopohjan luotavuutta.

Olisi tärkeää laatia selkeät ohjeet, miten internetlähteisiin tulisi viitata tutkielmissa. Esimerkiksi pelkän pääsivun osoitteen ilmoittaminen työssä ei ole kovinkaan mielekää, koska käytetyn tiedon etsimiseen sivustoilta voi mennä paljon aikaa. Toisaalta 40 prosenttia tutkielmissa annetuista tarkoista osoitteista ei toiminut suoraan, jolloin tietoa piti etsiä joko sivustoilta uudesta paikasta tai uusilta sivustoilta. Toisaalta Wayback machinen käytössä tarkan osoitteen ilmoittaminen auttaa hakemista huomattavasti. Arkiston käytössä myös sivustojen päivitysaika nopeuttaa löytämistä. Tiedon etsimisessä sen sijaan lukuajalla ei ole kovinkaan suurta merkitystä. Osoitteiden ilmoittamistarkkuuden suhteen olisi hyvä löytää jokin ratkaisu. Esimerkiksi yhtenä ratkaisuna voisi tarkan osoitteen ja sivustojen päivitysajan ilmoittamisen ohella kuvailla käytettyä aineistoa.

Tiina Valonen