



Semantinen Finlex

Laki ja oikeus avoimena linkitettyinä datana

Eero Hyvönen¹, Arttu Oksanen¹, Jouni Tuominen¹, Eetu Mäkelä¹, Minna Tamper¹, Aki Hietanen²,
Aalto-yliopisto, Semanttisen laskennan tutkimusryhmä (SeCo) ja HELDIG-keskus, Helsingin yliopisto (1) sekä
oikeusministeriö (2)

<http://seco.cs.aalto.fi/projects/lawlod>

1 MIKSI LAINSÄÄDÄNTÖ PITÄÄ JULKAISTA AVOIMENA DATANA?

Lainsäädäntömme keskeisimpiä sisältöjä, kuten lakeja, näitä tarkentavia asetuksia ja niiden soveltamiseen liittyvää oikeuskäytäntöä (oikeustapauksia) on voinut lukea verkossa oikeusministeriön tuottamassa Finlex-palvelussa¹ jo vuodesta 1997². Finlexissä julkaistaan lisäksi erilaisia viranomaismääräyksiä, valtiosopimuksia, hallituksen esityksiä ja säädösvalmisteluun liittyviä oppaita. Laki- ja oikeustieto on ollut tältä osin avointa ihmislukeijoille verkkosivuna, mutta aineistot eivät ole olleet saatavilla avoimena *datana* siten, että tietotekniset sovellukset ja toiset verkkopalvelut voisivat hyödyntää aineistoja dataa lataamalla tai avoimien rajapintojen kautta.

Monilla tahoilla on kuitenkin selvä tarve lainsäädännöllisten sisältöjen saamiseksi käyttöön datana:

- **Tietoportaalit.** Monissa verkkopalveluissa on tarvetta viitata lakien ja asetusten eri kohtiin ja näyttää niitä lukijalle. Tämä ei onnistu, jos lainsäädännön kohdat eivät ole viitattavissa ja luettavissa verkon kautta datana.
- **Juridiset verkkopalvelut.** Esimerkiksi Suomen Laki ja Edilex tarjoavat lakitietoa oikeusalan ammattilaisille, kuten tuomareille ja yritysjuristeille, mutta myös yksityishenkilöille. Nykyisten järjestelmien ylläpito perustuu suuressa määrin käsityöhön, koska dataa ei ole saatavilla koneen ”ymmärtämässä” muodossa, vaan ainoastaan eri muotoisina PDF-, Word- ym. dokumentteina.
- **Lainvalmistelutyö.** Laadittaessa uusia säädöksiä, jotka täydentävät tai korvaavat aiempia säädöksiä, joudutaan perehtymään aiemman lainsäädännön määräyksiin vaikutusten arvioimiseksi ja ristiriitaisuuksien välttämiseksi. Tietoa lainsäädännön eri versioista ja keskinäisistä riippuvuuksista ei kuitenkaan ole saatavilla kuin tekstuaalisessa muodossa. Rakenteisen datan avulla lainvalmistelutyötä tukemaan voitaisiin kehittää erilaisia apuvälineitä ja tarkistimia.
- **Lakiaineistojen toimittaminen ja julkaiseminen.** Lainsäädäntöön liittyvää tietoa tuotetaan nykyisin epäyhtenäisellä tavalla. Käytössä on erilaisia tekstiformaatteja, asiasanastoja ja käytäntöjä sisällön kuvailussa. Jos dokumentit laaditaan jo tuotantovaiheessa rakenteisena datana yhdessä sovittuja standardeja noudattaen, helpottuu niiden jatkokäsittely ja yhdistäminen toisiin dokumentteihin esimerkiksi eduskunnassa tai Finlexin kaltaisessa julkaisujärjestelmässä.
- **Media.** Yritysmailmaan ja politiikkaan liittyvissä uutisissa viitataan usein yksittäisten lakien eri kohtiin, jolloin olisi tarpeen voida johdattaa lukija alkuperäisten lakitekstien äärelle. Tämä ei onnistu, jos lainsäädännön kohdat eivät ole viitattavissa ja käytettävissä datana.
- **Älykkäämmät palvelut.** Juridisiin ongelmatilanteisiin, kuten vaikkapa avioeroon tai perinnönjakoon liittyvä tieto on lainsäädännössä usein pirstaloitunut eri lakeihin, asetuksiin ja oikeuskäytännön esimerkkitapauksiin. Ei auta, vaikka säädöksiä ja oikeustapauksia olisi saatavilla, jos kokonaisuuden

¹ <http://www.finlex.fi>

² Aki Hietanen: Free Access to Legislation in Finland: Principles, Practices and Prospects. In: Ginevra Peruginelli – Mario Ragona (ed.), Law via the Internet. Free Access – Quality of Information – Effectiveness of Rights. European Press Academic Publishing, Florence 2009.

hahmottaminen ei lukijalle, kuten tavalliselle kansalaiselle, onnistu. Lainsäädännön dokumenttien esittäminen koneen ”ymmärtämänä” eli semanttisena datana mahdollistaa aiempaa älykkäämpien sovellusten kehittämisen kansalaisille ja muille tahoille. Lakitekstejä voidaan esimerkiksi linkittää niihin liittyviin toisiin teksteihin ja oikeustapauksiin tai lakia selittäviin sanastoihin automaattisesti.

- **Lainsäädännön ja -käytön tutkimus.** Lainsäädäntötyö ja soveltaminen ovat oikeustieteellisen tutkimustyön kohteena, ja tässä voidaan hyödyntää mm. data-analyysin menetelmiä. Voidaan esimerkiksi tutkia, miten lakia on sovellettu eri aikoina tai eri tuomioistuinten toimesta. Nykyisellään tällainen tutkimus on erittäin työlästä käsityötä. Työtä voitaisiin automatisoida huomattavasti, jos säädökset, niiden väliset yhteydet ja tieto säädösten soveltamisesta oikeustapauksissa on saatavilla systemaattisesti esitettynä datana.

Lakiin ja oikeustapauksiin on voinut tutustua avoimesti verkkosivujen ja paperitulosteiden välityksellä. Laki ja oikeus on periaatteessa avointa ja jokaisen kansalaisen oletetaan sitä tuntevan ja kuuliaisesti noudattavan. Suorastaan yllättävää kuitenkin on, että mikään Suomen lainsäädäntöön liittyvä data ei tiettävästi ole kuitenkaan ollut avoimesti saatavilla. Semanttinen Finlex -hankkeen³ tavoitteena on tarttua tähän haasteeseen.

2 TYÖ ALKAA

Tutkimus- ja kehitystyö Finlexiin kuuluvien sisältöjen julkaisemiseksi avoimena linkitettyinä datana, ”Semanttisena Finlexinä”, käynnistyi v. 2012. Työ oli yksi case-tutkimus Aalto-yliopiston vetämässä Linked Data Finland -projektissa⁴ (2012–2014) ja siinä olivat mukana Aalto-yliopiston ja Helsingin yliopiston ohella oikeusministeriö, Edita Publishing Oy ja Talentum Oyj Tekesin toimiessa päärahoittajana. Hanke jatkui 2015–2016 valtiovarain- ja oikeusministeriön erillisellä rahoituksella liittyen valtionhallinnossa menossa olleeseen laajempaan Avoimen tiedon ohjelmaan⁵ ja tuottavuuden kehittämistavoitteisiin. Seuraavassa kuvataan tarkemmin tämän työn taustoja, tavoitteita, haasteita ja tuloksia.

Semanttinen Finlex -hankkeessa tutkitaan ja kehitetään Suomen lainsäädännön ja oikeuskäytännön avointa julkaisemista datana semanttisen webin Linked Data -teknologioiden⁶ avulla. Yhtenä esikuvana työlle oli Alankomaissa tehty Metalex Document Server⁷ lakidatan julkaisemiseksi linkitetyn datan palveluna. EU:n piirissä oli käynnistynyt standardointityötä tavoista, joilla voitaisiin viitata kansallisiin lakeihin ja oikeustapauksiin. Tämä on tarpeen, kun esimerkiksi jäsenvaltioiden lakitietoa kootaan EU-tasolla erilaisiin palveluihin.

Tutkimustyön tuloksena syntyi ensimmäinen avoimen datan julkaisu keskeisimmistä Finlexin sisällöistä⁸, joka koostui vuoden 2012 ajantasaisista laeista (2413 kpl) ja valikoimasta niihin liittyviä korkeimman oikeuden ratkaisuja (11 904 kpl vuosilta 1926–2012) ja korkeimman hallinto-oikeuden ratkaisuja (1490 kpl vuosilta 1944–2012). Semanttinen Finlex on ollut käytettävissä Linked Data Finland -datapalvelussa⁹ sekä rajapintojen kautta että julkaisun eri datajoukot ja metadataskeemat lataamalla. Hankkeen yhteydessä kerättiin ja jalostettiin myös lainsäädännön kuvailussa eri organisaatioissa käytettyjä asiasanastoja yhtenäiseen muotoon, yhteensä lähes 30 000 termiä 26 eri sanastosta, ajatuksena niiden harmonisointi myöhemmin asteittain yhä yhteentoimivammaksi ontologiaksi.

Lainsäädäntöön liittyvän tiedon esittämistä on aiemmin tutkittu Suomessa SGML- ja XML-perustaisen teknologian

³ <http://seco.cs.aalto.fi/projects/lawlod/>

⁴ <http://seco.aalto.fi/projects/ldf/>

⁵ <http://vm.fi/avointieto>

⁶ <http://www.w3.org/standards/semanticweb/>

⁷ Rinke Hoekstra: The MetaLex Document Server legal documents as versioned linked data. In: Proceedings of ISWC 2011, Bonn, Germany. pp. 128–143. Springer-Verlag, 2011.

<http://iswc2011.semanticweb.org/fileadmin/iswc/Papers/In-Use/70320129.pdf>

⁸ Matias Frosterus, Jouni Tuominen, Eero Hyvönen: Facilitating Re-use of Legal Data in Applications – Finnish Law as a Linked Open Data Service. In: Proceedings of the 27th International Conference on Legal Knowledge and Information Systems (JURIX 2014), Kraków, Poland. pp. 115–124. IOS Press, 2014.

<http://seco.cs.aalto.fi/publications/2014/frosterus-et-al-finnish-law-lod-2014.pdf>

⁹ Eero Hyvönen, Jouni Tuominen, J., Miika Alonen, Eetu Mäkelä: Linked Data Finland: A 7-star model and platform for publishing and re-using linked datasets. In: Proceedings of ESWC 2014 Demo and Poster Papers. Crete, Greece. pp. 226–230. Springer-Verlag, 2014. <http://www.seco.tkk.fi/publications/2014/hyvonen-et-al-ldf-2014.pdf>

pohjalta.¹⁰ Myös nykyinen Finlex on XML-perustainen. Finlexin data ei kuitenkaan ole ollut avoimesti saatavilla, ja julkaisussa käytetty formaatti on suunniteltu lähinnä järjestelmän toimittajan Editan Publishing Oy:n kehittämää julkaisujärjestelmää varten. W3C:n standardoima XML soveltuu hyvin dokumenttien rakenteen esittämiseen, mutta metatietojen esittämiseen W3C on sittemmin kehittänyt semanttisen webin alueelle kuuluvan RDF-tietomallin (Resource Description Framework) ja erityisen Linked Data -metodiikan datajulkaisujen toteuttamiseksi verkossa¹¹. Jo vuonna 2006 RASKE2-hankkeessa nähtiin semanttinen web ja RDF lupaavaksi uudeksi teknologiaksi lainsäädäntötyön tiedonhallinnassa¹², ja eduskunnan kirjasto oli mukana kansallisessa FinnONTO-hankkeiden sarjassa (2003–2012), jossa kehitettiin pitkäjänteisesti suomalaisen semanttisen webin infrastruktuuria ja sovelluksia¹³.

Semanttinen Finlex palkittiin vuonna 2015 Open Finland Challenge -kilpailussa Public Service & Active Citizens -sarjan parhaana hankkeena¹⁴.

3 SEMANTTINEN FINLEX JULKAISTAAN

Työn jatko käynnistyi 2015 Aalto-yliopistossa oikeusministeriön johdolla ja yhteistyössä Finlexin teknisen ratkaisun toteuttaneen Edita Publishing Oy:n kanssa. Vuoden 2016 loppuun mennessä Semanttisen Finlexin (SF) prototyyppiä kehitettiin eteenpäin seuraavilla tavoilla:

- **IRI-tunnistejärjestelmät.** SF:n eri sisältökohteiden (lait, asetukset, pykälät, oikeustapaukset, organisaatiot jne.) IRI-tunnistejärjestelmät (W3C:n standardoima Internationalized Resource Identifier) muutettiin uusien standardien ELI¹⁵ (European Law Identifier) ja ECLI¹⁶ (European Case Law Identifier) mukaiseksi¹⁷. Eurooppalainen lainsäädäntötunnus ELI tarjoaa joustavan, dokumentoidun ja yhtenäisen tavan viitata säädöksiin kansallisissa oikeusjärjestelmissä. ELI-tunnus koostuu kolmesta osasta eli pilarista:
 - 1) Säädöstunniste (HTTP URI)
 - 2) Metatiedot
 - 3) ELI-metatietojen julkaiseminen koneluettavassa muodossa

Eurooppalainen ELI-tunnusstandardi julkaistiin vuonna 2012 ja vajaassa viidessä vuodessa se on otettu käyttöön noin kymmenessä EU-maassa¹⁸. Vastaava oikeustapauksiin liittyvä ECLI-oikeuskäytäntötunnus julkaistiin standardina EU:n virallisessa lehdessä huhtikuussa 2011. ECLI koostuu määrämuotoisesta tunnistemerkinästä ja metatietokokonaisuudesta. Euroopan komissio on ollut aktiivinen ECLI-tunnuksen käyttöönotossa ja noin kymmenen EU-maan lisäksi tunnus on käytössä mm. EU-tuomioistuimessa, Euroopan ihmisoikeustuomioistuimessa ja Euroopan patenttivirastossa. Euroopan komissio on myös julkaissut Euroopan oikeusportaalissa ECLI-hakupalvelun, jossa haun piirissä on useiden miljoonien

¹⁰ Tiedonhallinta suomalaisessa lainsäädäntöprosessi. RASKE2-projektin väliraportti. Eduskunnan kanslian julkaisu 5/2004. http://www.eduskunta.fi/FI/tietoaeduskunnasta/julkaisut/Documents/ekj_5+2004.pdf

¹¹ <http://www.w3.org/standards/semanticweb/>

¹² Reija Nurmekselä, Maiju Virtanen, Antti Lehtinen, Matti Järvenpää, Airi Salminen: Suomalaisen lainsäädäntötyön tiedonhallinta. Suuntana semanttinen web. Eduskunnan kanslian julkaisu 2/2006. <http://docplayer.fi/750837-Suomalaisen-lainsaadantotyön-tiedonhallinta-suuntana-semanttinen-web.html>

¹³ Eero Hyvönen: FinnONTO-hanke loi ontologisen perustan kansalliselle webin tietoinfrastruktuurille. Tieteessä tapahtuu, no. 3, 2014. s. 12–18. <http://ojs.tsv.fi/index.php/tt/article/view/41559>

¹⁴ <http://www.aalto.fi/fi/current/news/2015-12-04-002/>

¹⁵ http://en.wikipedia.org/wiki/European_Legislation_Identifier

¹⁶ http://en.wikipedia.org/wiki/European_Case_Law_Identifier

¹⁷ ELI-standardia on kuvattu yksityiskohtaisesti seuraavissa dokumenteissa: ELI Task Force: ELI implementation methodology. Euroopan unionin julkaisutoimisto, 2015. <http://data.europa.eu/doi/10.2830/813167>; ELI Task Force: ELI: A technical implementation guide. Euroopan unionin julkaisutoimisto, 2015. <http://data.europa.eu/doi/10.2830/74251>.

¹⁸ <http://eur-lex.europa.eu/eli-register/about.html>

tuomioistuinratkaisujen metatiedot¹⁹. Laaja kiinnostus ELI- ja ECLI-standardien käyttöön on nähtävissä osana sähköisen oikeuden kehitystä Euroopassa²⁰. Euroopan julkaisutoimisto, joka vastaa EU:n toimielinten sähköisistä julkaisupalveluista, julkisti vuonna 2016 EU:n lainsäädännön ja oikeuskäytännön avoimen datan palvelun (CELLAR-tietoaaineisto- ja metatietopalvelu)²¹. EU-aineistojen uudistunut saatavuus mahdollistaa monipuolisen tietojen siirron Semanttisen Finlexin ja EU:n avoimen datan palvelujen kesken.

IRI-tunnisteiden avulla voidaan 1) indeksoida ja sisällönkuvailla tietosisältöjä yksikäsitteisellä tavalla, 2) muodostaa rajapintakyselyjä koneluettavan datan hakemiseksi ja lukemiseksi SF:stä eri muodoissa (kuten RDF, JSON ja CSV) ja 3) viitata Finlexin vastaaviin ihmisluettaviin sivuihin eri sovelluksissa.

- **Metadatatallit.** SF:n metadatatallit uusittiin niin, että lakitekstejä ei tarvitse koostaa niiden osista, kuten alkuperäisessä SF:ssä. Aiempi tietomalli myötäili alkuperäistä, dokumenttien rakennetta kuvaavaa XML-skeemaa, eikä niinkään kuvannut niiden sisältöä dokumentteina. Uudet tietomallit²² laadittiin ELI- ja ECLI-standardien suosittamien metadatatallien mukaisiksi. Metadatatallit perustuvat bibliografisten tietueiden kuvailuun kehitettyyn FRBR-käsitelmään sekä webissä laajassa käytössä olevaan Dublin Core -sanastoon.
- **Julkaisun automaattinen päivitys.** Hankkeessa toteutettiin SF:n julkaisun automaattinen päivitys Finlex-julkaisun muutoksien mukaan, jolloin käytössä on aina ajantasainen Finlexiä vastaava datajulkaisu.
- **Sanastot ja ontologiat.** Sisällönkuvaailussa käytettyä asiasanastoja ja ontologioita kehitettiin eteenpäin, mukaan lukien oikeusministeriössä valmistunutta Asseri-sanastoa. SF-aineistoja linkitettiin myös Tieteen termipankin oikeustieteellisen tietokannan²³ määritelmiin (n. 1650 määritelmää/sivua).
- **Uudet datajulkaisut.** Semanttiseen Finlexiin otettiin mukaan uusia tietosisältöjä Finlexistä, kuten alkuperäiset säädökset.
- **Datan validointi.** Järjestelmän automaattista datan laadun validointia kehitettiin eteenpäin.

Semanttinen Finlex julkaistiin virallisesti 10.3.2016 Säätytalolla pidetyssä seminaarissa²⁴, jonka järjestivät oikeusministeriö, Aalto-yliopisto ja Edita Publishing Oy yhteistyössä valtiovarainministeriön kanssa. Datapalvelu on käytettävissä osoitteessa

<http://data.finlex.fi>

josta löytyy myös palvelun tarkempi dokumentaatio. Palvelu sisälsi julkaisuhetkellä laajan joukon alkuperäisiä säädöksiä (n. 45 000 kpl suomeksi ja ruotsiksi eli yhteensä 90 000 kpl), näistä toimitettuja ajantasaisia lakeja (n. 2 800 kpl suomeksi ja ruotsiksi eli yhteensä n. 5 600 kpl), korkeimman oikeuden (KKO) ratkaisuja (n. 5 400 kpl suomeksi ja n. 5 400 kpl ruotsiksi) ja korkeimman hallinto-oikeuden (KHO) päätöksiä (n. 7 000 suomeksi ja n. 2 000 ruotsiksi). Sisältöihin voidaan viitata joustavasti ELI- ja ECLI-tunnisteilla. Esimerkiksi ELI-osoite

<http://data.finlex.fi/eli/sd/2008/521/ajantasa/20160101>

palauttaa vuoden 2008 säädöksen numero 521 (Vakuutusyhtiölaki) ajantasaisen version ajanhetkellä 1.1.2016. Lain sisällä voidaan viitata mihin tahansa lainkohtaan ja sen tietyllä hetkellä voimassa olleeseen versioon. Siten ELI-osoite

<http://data.finlex.fi/eli/sd/2008/521/luku/1/pykala/1/ajantasa/20160101>

palauttaa Vakuutusyhtiölain 1 luvun 1 pykälän ajanhetkellä 1.1.2016.

¹⁹ https://e-justice.europa.eu/content_ecli_search_engine-430-fi.do

²⁰ Aki Hietanen: E-Justice in Finland – Trends and Challenges. In: Thomas Gottwald (ed.), E-Justice in Österreich – Erfahrungsberichte und europäischer Kontext. pp. 757–770. Editions Weblaw, Bern, 2013.

²¹ <http://joinup.ec.europa.eu/software/cellar/description>

²² Arttu Oksanen: Lainsäädännön ja oikeuskäytännön mallintaminen ja julkaiseminen linkitettyinä avoimena datana. MSc Thesis, Aalto University, School of Electrical Engineering, Degree Programme in Electronics and Electrical Engineering, October, 2016. <http://seco.cs.aalto.fi/publications/2016/oksanen-msc-thesis-2016.pdf>

²³ <http://tieteentermipankki.fi/wiki/Oikeustiede>

²⁴ <http://seco.cs.aalto.fi/events/2016/2016-03-10-semanttinen-finlex/>

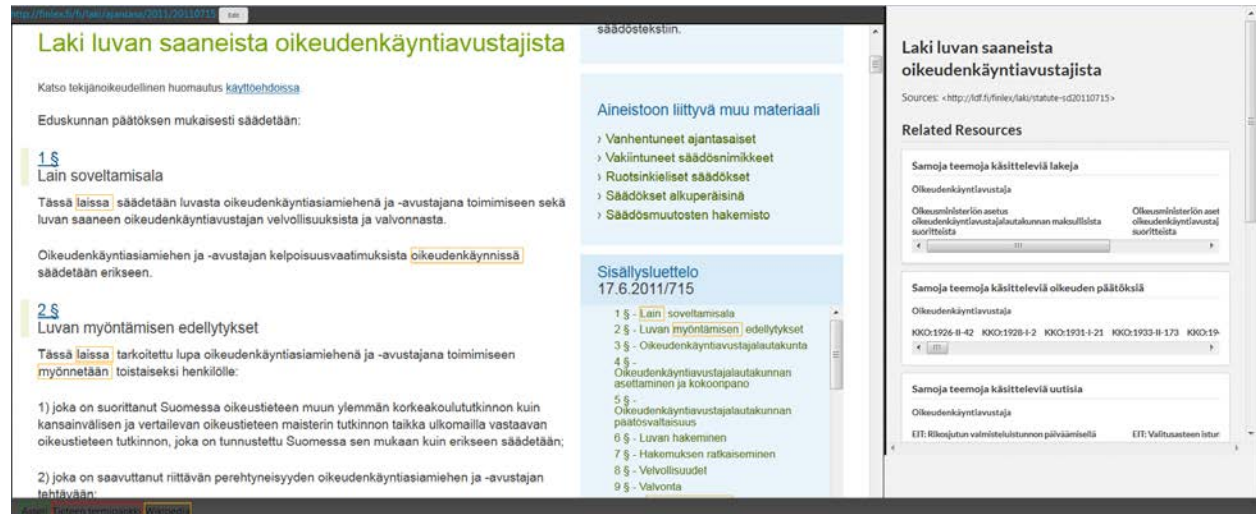
Vastaavasti ECLI-tunnisteen sisältävä osoite <http://data.finlex.fi/oikeus/ECLI:FI:KKO:2016:65> tuo korkeimman oikeuden ennakkopäätöksen KKO:2016:65.

Lisäksi palvelussa julkaistaan oikeushallinnon piirissä käytössä olevaa terminologiaa ontologioina. Tällaisia sanastoja ovat mm. Asseri-sanasto, joka on 26 oikeushallinnon sanastosta koostettu n. 6000 termin kokoelma, Finlex-palvelussa käytössä oleva säädöskokoelman asiasanasto sekä KKO:n ja KHO:n asiasanastot. Lisäksi käytetään ulkoisia sanastoja, kuten n. 47 000 käsitteen KOKO-ontologiaa²⁵, Eurovoc-sanastoa²⁶ ja Wikipedian (DBpedia) käsitteitä.

4 SOVELLUSKEHITYSTÄ

Semanttisen Finlexin hyödyllisyyttä tutkitaan demonstraattoreiden avulla.

Ensimmäisenä sovelluksena Semanttinen Finlex -datapalvelun rajapinnan käytöstä on toteutettu Semanttinen Finlex -selain, joka linkittää verkossa olevan Finlex-palvelun HTML-sivujen ja PDF-dokumenttien sanoja ja pidempiä ilmauksia reaaliajassa ja automaattisesti niitä taustoittaviin tietoihin. Ideana on helpottaa lukijalle lakiteksteihin tutustumista tarjoamalla hänelle automaattisesti lakeihin liittyvää hyödyllistä lisätietoa, niiden lukukonteksti.



Kuva 1: Semanttinen Finlex-selain tarjoaa lukukontekstin Finlexin aineistojen lukijalle linkitetyn datan avulla.

Semanttinen Finlex -selain (kuva 1) linkittää termien määritelmiä kolmesta eri datalähteestä: 1) Finlexin Asseri-sanastosta, 2) Helsingin yliopiston Tieteen termipankin oikeustieteellisestä tietokannasta²⁷ ja 3) Suomenkielisestä Wikipediasta. Linkitetty termi näkyy käyttäjälle värikoodattuna laatikkona tekstissä. Viemällä kursori laatikon päälle, voi lukea siihen liittyvää lisätietoa infolaatikon kautta. Laatikon väri ilmaisee linkin tietolähteen, jotka on lueteltu selainsivun alareunassa. Lisäksi kunkin Finlex-sivun oikealle puolelle luodaan automaattisesti joukko semanttisia suosittelevia lakeihin liittyviin muihin Semanttisen Finlexin aineistoihin. Tässä tapauksessa linkit johtavat 1) samoja teemoja käsitteleviin toisiin lakeihin, 2) samoja teemoja käsitteleviin oikeustapauksiin ja 3) samoja teemoja käsitteleviin uutisiin. Tässä demonstraattorissa hyödynnettiin SeCo-ryhmässä kehitettyä Contextual Reader -järjestelmää²⁸. Tavoitteena on auttaa lukijaa ymmärtämään aiempaa paremmin lain sisältöjä.

²⁵ <http://finto.fi/koko/fi/>

²⁶ <http://eurovoc.europa.eu>

²⁷ Suomalaisen Lakimiesyhdistyksen Oikeustieteellinen tietokanta on Termipankin ainoa tieteellinen sanasto, jota ei ole julkaistu avoimen datan lisenssillä, eikä linkitystä sen termeihin ole valitettavasti voitu julkaista tekijänoikeudellista syistä Semanttisessa Finlexissä tai sen selaimessa.

²⁸ Eetu Mäkelä, Thea Lindquist, Eero Hyvönen: CORE – A Contextual Reader based on Linked Data. In: Proceedings of Digital Humanities 2016, long papers, Kraków, Poland. pp. 267–269. 2016. <http://seco.cs.aalto.fi/publications/2016/makela-et-al-core-dh2016.pdf>. CORE-hankkeen kotisivulla on useita demonstraattoreita järjestelmästä: <http://seco.cs.aalto.fi/projects/core/>.

Toisena demonstraattorina toteutettiin semanttinen haku- ja suosittelukone, jonka avulla lakisäilytöjä voidaan hakea paitsi perinteisillä monimerkityksillä merkijonoilla myös ontologisilla käsitteillä. Hakutuloksia voidaan myös linkittää automaattisesti toisiinsa semanttisten suosituslinkkien avulla. Semanttinen haku ja suosittelu perustuvat kieliriippumattomiin käsitteisiin ja mahdollistavat näin myös hakujen suorittamista ja suosittelua yli kielirajojen.

Hankkeessa luotiin perusta lakiaineistojen automaattiselle semanttiselle sisällönkuvailulle kieliteknologian avulla²⁹. Samoja menetelmiä hyödyntämällä kehitettiin myös käsitepilvi (kuva 2) jokaiselle ajantasaiselle säädökselle helpottamaan säädösten sisällön hahmottamista.



Kuva 2. Käsitepilvi sosiaalihuoltolaille (30.12.2014/1301).

Vuosina 2017–2018 kehitystyötä viedään eteenpäin Aalto-yliopiston ja Helsingin yliopiston HELDIG – Helsinki Centre for Digital Humanities -keskuksen³⁰ toimesta osana laajempaa Severi-projektia³¹. Siinä tutkimus- ja kehityskohteena ovat erityisesti datapalvelun varaan kehitettävät älykkäät toiminnallisuudet loppukäyttäjille sekä sisältöjen automaattinen kuvailu (annotointi), linkitys ja anonymisointi.

Semanttisen Finlexin tutkimus- ja kehitystyötä vetää Aalto-yliopiston ja Helsingin yliopiston Semanttisen laskennan tutkimusryhmässä professori Eero Hyvönen yhdessä tutkijoiden Jouni Tuominen, Eetu Mäkelä, Arttu Oksanen ja Minna Tamper kanssa. Tätä työtä rahoittaa Tekes ja Severi-projektin konsortio. Finlexin kehittämistä johtaa Aki Hietanen oikeusministeriöstä. Jari Linhala ja Risto Talo ovat vastanneet Finlexin teknisestä toteutuksesta ja datan tuotannosta projektin käyttöön.

5 SISÄLLÖNTUOTANNON HAASTEITA

Koko hankkeen yksi keskeinen haaste liittyy tapaan, jolla lainsäädäntöä Suomessa toimitetaan ja julkaistaan. Lainsäädännöstä on Finlexissä olemassa kolme keskeistä aineistoa³²:

²⁹ Minna Tamper: Extraction of Entities and Concepts from Finnish Texts. MSc Thesis, Aalto University, School of Science, Degree Programme in Computer Science and Engineering, December, 2016. <http://seco.cs.aalto.fi/publications/2016/tamper-msc-thesis-2016.pdf>

³⁰ <http://heldig.fi>

³¹ <http://seco.aalto.fi/projects/severi/>

³² Aki Hietanen, Sähköinen, autenttinen ja ajantasainen – sähköisen säädösjulkaisemisen kehitysvaiheista. In: Heikki

1. **Säädökset alkuperäisinä (SA).** Kokoelma sisältää säädösten tekstit sellaisina kuin ne on eduskunnasta, hallitukselta tai muusta lähteestä saatu ja julkaistu säädöskokoelmassa. Finlexissä on kaikki säädökset v. 1987 alkaen ja keskeinen osa vanhemmista säädöksistä. Uudet säädökset voidaan helposti lisätä Finlexin kokoelmaan myös alkuperäisessä PDF-muodossa.
2. **Ajantasainen lainsäädäntö (AL).** Tämä kokoelma sisältää päivitettyt säädöstekstit, joissa lakiin tai asetukseen tehdyt muutokset on sisällytetty alkuperäiseen säädöstekstiin. Tässä tarvitaan käsityötä, josta kohta tarkemmin lisää.
3. **Säädösmuutosten hakemisto (SH).** Edellisten synkronoimiseksi toisiinsa ylläpidetään Finlexissä lisäksi erillistä säädösmuutosten hakemistoa, jossa kerrotaan uusien lakien ja asetusten voimaantulosta tai vanhan kumoamisesta.

Kun julkaistaan uusi säädös, kuten laki tai asetus, se tulee Finlexin SA-kokoelmaan. Uusi säädös tyypillisesti korjaa tai täydentää aiempia säädöksiä muuttaen niiden eri kohtia. Kullakin ajanhetkellä voimassa olevat säädökset muodostavat ajantasaisen lain. Haasteena on, että ajantasaista lainsäädäntökokoelmaa (AL) ei nykyisellään voida täysin automaattisesti tuottaa aiemmista ajantasaisista laista ja niiden jälkeisistä säädöksistä, vaan tässä joudutaan tekemään tekstin tulkintaa ja editoimaan muuttuneita lakeja käsityönä.

Tarkastellaan esimerkkinä kuvassa 3 olevaa eduskunnan säädöstä 1291/2015, joka on annettu 2.11.2015:

1291/2015

Laki

kiinteistöverolain muuttamisesta

Eduskunnan päätöksen mukaisesti
muutetaan kiinteistöverolain (654/1992) 11 §:n 3 momentti, 12 §:n 1 momentti, 12 a §:n 1 ja 2 momentti, 12 b §, 13 §:n 1 momentti ja 14 §:n 1 momentti, sellaisina kuin ne ovat, 11 §:n 3 momentti ja 12 §:n 1 momentti laissa 861/2014, 12 a §:n 1 ja 2 momentti laissa 1026/1999, 12 b § laissa 811/2013, 13 §:n 1 momentti laissa 862/2014 ja 14 §:n 1 momentti laissa 969/2009, seuraavasti:

11 §

Yleinen kiinteistöveroprosentti

Kunnanvaltuuston määräämää *yleistä kiinteistöveroprosenttia* sovelletaan tämän lain mukaan veronalaisiin kiinteistöihin, jollei 12, 12 a, 12 b, 13, 13 a tai 14 §:stä muuta johdu. Yleiseksi kiinteistöveroprosentiksi on määrättävä vähintään 0,86 ja enintään 1,80.

12 §

Vakituisten asuinrakennusten veroprosentti

Kunnanvaltuusto määrää pääasiassa vakituiseen asumiseen käytettävien rakennusten veroprosentiksi vähintään 0,39 ja enintään 0,90.

Kuva 3. Esimerkki säädöksestä.

Säädöksen johdannossa kerrotaan sanallisesti, minkä eri säädösten kohtia (kokonaisia pykäläiä tai momentteja, kohtia jne.) säädöksellä muutetaan, ja sitten kerrotaan kohta kohdalta, miten niitä on muutettu. Esimerkiksi listan ensimmäisen muutoksen kiinteistöverolain 11 §:n 3 momentin kohdalla ajantasaisessa laissa luki säädöksen julkaisemispäivänä (kuva 4):

Kunnanvaltuuston määräämää *yleistä kiinteistöveroprosenttia* sovelletaan tämän lain mukaan veronalaisiin kiinteistöihin, jollei 12–14 §:stä muuta johdu. Yleiseksi kiinteistöveroprosentiksi on määrättävä vähintään 0,80 ja enintään 1,55. ([7.11.2014/861](#))

Kuva 4. Esimerkki kuvan 1 säädöksen muuttamasta momentista.

Tämä momentti on siis uuden säädöksen myötä vanhentunut, mutta tämä ei näy vielä AL:ssa. Lainsäädännöllinen sisällöntuotantoprosessi ei tuota valmiita uusia ajantasaisia versioita laeista, vaan ikään kuin paikkoja nuttuun paikkojen päälle, ja ajantasaista lakia parsitaan kokoon käsityönä aina uusien säädösten tullessa voimaan. Kulloisenkin ajantasaisen lain hahmottaminen vaatii lukijalta muutoshistorian selvittämistä viimeisimmästä ajantasaisesta laista lähtien. Tässä voidaan hyödyntää säädösmuutosten hakemistoa (SH).

Finlexin kannalta käsityön tarve katkaisee julkaisuprosessin ja johtaa myös siihen, että ajantasaiseen lakiin syntyy toimitustyön kautta tekijänoikeuksia, jolloin datan avaaminen hankaloituu. On aiheellista miettiä, kannattaisiko prosessia muuttaa niin, että se tuottaa aina myös uusia versioita muuttuneista laeista ja asetuksista säädösten ohella. Käytännössä tämä tarkoittaisi sitä, että joko hallituksen esityksen liitteenä tai eduskuntakäsittelyn aikana tuotettuna (epävirallisena) tausta-aineistona olisi saatavilla säädöksen tuleva ajantasainen versio. Asetusten kohdalla tämä edellyttäisi, että ministeriössä asetuksen valmistelija kokoaisi ajantasaisen version, ajantasainen asetusteksti tuotettaisiin valtioneuvoston yhteisenä palveluna (esimerkiksi valtioneuvoston kansliassa) tai vaihtoehtoisesti tämä päivitystyö ulkoistettaisiin.

Finlex on suomalaisten säädösten virallinen julkaisufoorumi ja vuodesta 2011 lähtien säädökset on julkaistu vain sähköisinä autenttisina tiedostoina³³. Finlexissä julkaistaan kaikki Suomen säädöskokoelmassa ilmestyvät säädökset, mutta kaikista säädöksistä ei tuoteta ajan tasalle toimitettuja versioita. Ajantasaiset versiot julkaistaan kaikista laeista, suurimmasta osasta valtioneuvoston asetuksia ja säädöksistä, joita muutetaan säännöllisesti. Ajantasaisten säädösten tuottamisessa karsintaa tehdään yhtäältä kustannussyistä, toisaalta sen vuoksi, että säädöksiä on hyvin paljon ja monilla säädöksillä on tuskin mielenkiintoa Finlexin käyttäjäkunnalle, kuten tavallisille kansalaisille. Laajempia toimitettuja versioita ajantasaisesta lainsäädännöstä on saatavilla kaupallisten palveluiden kautta, esimerkkinä Talentum Oyj:n Suomen Laki³⁴ ja Edita Publishing Oy:n Edilex³⁵. Nämä tarjoavat asiakkailleen myös kaikenlaista muutakin hyödyllistä kontekstietoa lainsäädännön eri kohdista ja toisiin aineistoihin (mm. oikeustapauksiin, EU-säädöksiin ja oikeuskirjallisuuteen) linkityksiä, joita niin ikään tehdään paljolti käsityönä. Finlexissä ei tällä hetkellä julkaista eri ajanhetkinä voimassa ollutta ajantasaista lainsäädäntöä aikasarjana, vaan ainoastaan viimeisin versio. Jos kansalainen on esimerkiksi syytteessä pari vuotta sitten tehdystä rikoksesta, on hänen vaikea saada selville, millaisen lain version mukaan asia tullaan käsittelemään.

Kiinnostava kysymys on, miten monipuolista lakipalvelua viranomaisten pitäisi tarjota kansalaisille maksuttomana peruspalveluna Semanttisen Finlexin kautta, ja mikä osuus jää yritysten kaupallisen toiminnan varaan. Tämä problematiikka ei ole uusi, vaan sitä käsiteltiin jo parikymmentä vuotta sitten Finlexin maksuttoman verkkopalvelun valmisteluvaiheessa³⁶.

Semanttisen Finlexin tavoitteena ei ole korvata kehittyneitä kaupallisia palveluita, mutta kuitenkin tarjota lakitietoa kansalaisille helpommin sulavassa muodossa, ja yrityksille harmonisoitua hyvälaatuista dataa sovellusten kehittämiseksi. Tavoitteena on, että käsityön tarve ja päällekkäinen aineistojen toimitustyö vähenisi erilaisia säädösjulkaisuja tuottavissa tai hyödyntävissä yrityksissä.

³³ Laki Suomen säädöskokoelmasta (25.2.2000/188), 12§ muuttui säädöksellä 1197/2010, joka tuli voimaan 1.1.2011.

³⁴ <http://www.suomenlaki.com>

³⁵ <http://www.edilex.fi>

³⁶ Valtiovallan perusvelvoite voimassa olevan normiston julkaisemiseen. Työryhmän mietintö. Oikeusministeriö, Helsinki, 1996.