

ONKI-projektin ontologiaeditorien vertailuraportti

Kirjoittajat: Tuomas Palonen, Osma Suominen, Sini Pessala, Mikko Lappalainen, Susanna Nykyri,
Henri Ylikotila, Matias Frosterus.

24. kesäkuuta 2013

Kansalliskirjasto / Tutkimuskirjasto ja Kirjastoverkkopalvelut



Sisällysluettelo

1. Tiivistelmä.....	3
2. Taustaa.....	3
2.1. ONKI-projektin taustaa.....	3
2.2. Ontologiaeditorit ja niiden käyttö.....	4
3. Editorien käyttövaatimukset ja testattavat toiminnot.....	5
4. Paikallisesti asennettavat editorit.....	6
4.1. TopBraid Composer: Free Edition ja Standard Edition.....	7
4.2. Protégé 4.3.....	9
5. Web-pohjaiset editorit.....	10
5.1. TopBraid EVN.....	11
5.2. PoolParty Thesaurus Server.....	12
5.3. VocBench.....	14
5.4. Muut web-pohjaiset editorit.....	15
6. Yhteenveto ja suositukset.....	15
7. Projektiin liittyvää verkkomateriaalia.....	16

1. Tiivistelmä

Kansalliskirjastossa maaliskuussa 2013 käynnistynyt ONKI-projekti vastaa yleisen suomalaisen ontologian (YSO) sekä kansallisen ONKI-ontologiapalvelun kehityksestä ja ylläpidosta. ONKI-projektia rahoittavat valtiovarainministeriö sekä opetus- ja kulttuuriministeriö. Projektin alkuvaiheen ensimmäisiä tavoitteita on ollut vertailla tällä hetkellä tarjolla olevia ontologioiden muokkaukseen käytettäviä ohjelmistoja, joista tästedes käytämme nimeä ontologiaeditori. Vertailun perusteella annetaan yhteenveto ja suositus vertailussa olleista vaihtoehdoista. Ontologia- editori on keskeisin työväline projektin jatkovaiheessa ja vastaavia työkaluja käyttävät myös Kansalliskirjaston ulkopuoliset, ONKI-projektiin liittyvät yhteistyötahot.

Raportissa esitellään projektin taustaa ja ontologiaeditorien käyttöä ja merkitystä ontologiatyössä sekä niihin liittyvät perusvaatimukset. Vertailussa olleet editorit, Top Braid Composer eri versioineen, Protégé 4.3, TopBraid EVN, PoolParty Thesaurus Server, VocBench, IQvoc sekä TemaTres esitellään kukin erikseen. Lopuksi esitetään yhteenveto ja suositus.

2. Taustaa

2.1. ONKI-projektin taustaa

Ontologiat liittyvät olennaisesti semanttisen webin teknologioihin, joita on viime vuosina tutkittu ja sovellettu paljon eri aloilla (mm. kirjastoalalla ja julkishallinnossa) niin kotimaassa kuin maailmalla. Yleisemmin aiheesta ks. Jouni Tuominen et al 2012: *Ontologiapalvelut semanttisessa webissä*¹.

Kansalliskirjaston ONKI-projektin piiriin kuuluvan YSO-ontologian taustalla on Yleinen suomalainen asiasanasto (YSA), jota niin ikään ylläpidetään Kansalliskirjastossa. Sekä YSO-ontologiaa että ONKI-palvelua on kehitetty Aalto-yliopiston FinnONTO-tutkimushankkeessa vuosina 2003–2012. Nyt vastuu on siirtynyt Kansalliskirjastolle ja fokus on muuttunut Aalto-yliopiston teknisestä tutkimushankkeesta vahvemmin käyttäjille suunnatuksi palveluksi. Projektin kehitysvaiheen on määrä kestää vuoteen 2017 asti, minkä jälkeen projekti siirtyy ylläpitovaiheeseen.

Sekä FinnONTO-tutkimushankkeessa että nykyisessä ONKI-projektissa on Kansalliskirjaston lisäksi mukana useita yhteistyöorganisaatioita, erityisesti julkiselta sektorilta. On varsin todennäköistä, että useat organisaatiot tulevat ylläpitämään omia ontologioitaan itsenäisesti Kansalliskirjaston koordinoimissa ontologiakokonaisuuksissa.

Ensimmäisiä tehtäviä ONKI-projektissa on ollut ontologiaeditorien testaus ja vertailu. Ontologiaeditori on keskeisin työkalu ontologian sisällöllisessä ylläpidossa. Vertailun tarkoituksena on kartoittaa tarjolla olevat vaihtoehdot vertailemalla eri editorien ominaisuuksia ja soveltuvuutta projektin käyttöön. Vertailua ovat tehneet tietoasiantuntijat Mikko

¹ <http://www.seco.tkk.fi/publications/2012/tuominen-et-al-ontologiapalvelut-2012.pdf>

Lappalainen, Susanna Nykyri ja Tuomas Palonen sekä tietojärjestelmäasiantuntijat Sini Pessala, Henri Ylikotila ja Osma Suominen. Työhön on osallistunut myös projektipäällikkö Matias Frosterus.

Editoreja on arvioitu toimivuuden, käyttökelpoisuuden sekä kustannustehokkuuden kannalta. Päämääränä on antaa tietopohjaa ja valintasuosituksia yhtäältä Kansalliskirjaston omaan käyttöön, toisaalta myös talon ulkopuolelle muille ONKI-palvelua käyttäville ja ontologioita ylläpitäville organisaatioille ja tahoille.

2.2. Ontologiaeditorit ja niiden käyttö

Ontologiaeditori on tietokonesovellus, jolla ontologioita rakennetaan, muokataan ja ylläpidetään. Nämä tehtävät edellyttävät editoria, jota ilman ontologian ylläpito ei ole mahdollista. Sopivan työkalun valinta on siis syytä tehdä harkiten, etenkin pitkäkestoisissa ja jatkuviissa hankkeissa.

Ontologiaeditoria käyttävät sellaiset tahot, jotka ylläpitävät omaa ontologiaansa tai ontologioitansa. Suomessa tällaisia tahoja on useita, joista monet ovat kehittäneet ontologiansa jo olemassa olleen asiasanaston tai muun kontrolloidun sanaston pohjalta. Kansalliskirjasto ylläpitää YSO-ontologiaa, mutta sen vastuulla ei tällä hetkellä ole muiden ontologioiden ylläpitoa. Kansalliskirjaston tarjoamassa kansallisessa ONKI-ympäristössä tarjotaan yhteistyötahoille palvelu, jossa yhteistyötahojen editoima ja Kansalliskirjastolle lähettämä ontologia muokataan ns. julkaisuputkessa esityskuntoon ja julkaistaan sen jälkeen kaikille avoimeen ONKI-verkkopalveluun. Lisäksi ONKI-palvelusta on kenen tahansa mahdollista ladata niin YSO-ontologia kuin muidenkin organisaatioiden kehittämiä ontologioita ja ottaa ne käyttöön omissa järjestelmissään.

Ontologiaeditoreja on tällä hetkellä tarjolla paljon ja valikoimasta löytyy niin kaupallisia tuotteita kuin ilmaisia avoimen lähdekoodin editorejakin. Ontologiaeditorit voidaan lisäksi jakaa kahteen tyyppiin: paikallisesti asennettaviin ja web-pohjaisiin. Paikallisesti asennettavat ladataan käyttäjän koneelle kuten myös ylläpidettävä ontologia, jota käyttäjä yksityisesti muokkaa. Muutoksia ja tallennuksia voi tehdä vain yksi käyttäjä kerrallaan. Toisaalta koneelle asennettavan editorin käyttöön ei sisälly web-ympäristön vakauteen ja nopeuteen liittyviä riskejä. Web-pohjaista editoria käytetään nettiselaimella ja muokattava ontologia sijaitsee erillisellä palvelimella. Tällöin etuna on, että ontologiaa voi muokata useampi henkilö yhtä aikaa.

Semanttinen web on jatkuvasti kehittyvä teknologian ala. Nykytilanne huomioiden onkin lähestulkoon varmaa, että ontologiaeditorien kehitys jatkuu tämän raportin kirjoittamisajankohdan jälkeenkin; toisaalta tiettyjen tuotteiden kehitys saatetaan myös lopettaa. Onkin todettava, että raportti kuvaa kesän 2013 tilannetta ja on tehty tämän hetken tarpeiden näkökulmasta. Toisaalta vaihtoehtovalinnoissa on kuitenkin keskitytty sellaisiin tuotteisiin, joiden saatavuus on tällä hetkellä vakaa ja luotettava.

3. Editorien käyttövaatimukset ja testattavat toiminnot

FinnONTO-hankkeessa hankitun käyttökokemuksen ja nykyisten tarpeiden perusteella työryhmä laati vertailua varten kriittiset käyttövaatimukset ylläpitoon soveltuville editoreille. Kriittisiä vaatimuksia täyttämättömät editorit hylättiin työhön soveltumattomina. Vaatimusten lisäksi ontologiaeditorien toimivuutta testattiin suorittamalla erilaisia ontologiatyössä välttämättömiä toimintoja. Testeissä kirjattiin ylös kunkin ohjelman kohdalla yhtäältä erityisominaisuuksia siten, kun niiden katsottiin tuovan lisäarvoa, ja toisaalta työtä hankaloittavia piirteitä tai puutteita. Sopivia editorikandidaatteja kartoitettaessa hylättiin suoraan editorit, jotka eivät pystyneet lukemaan ja tallentamaan SKOS- tai OWL-muotoisia ontologioita.

Kriittisiä vaatimuksia määriteltiin seuraavasti: 1) ontologian käsitehierarkian on oltava koko ajan näkyvillä, 2) käsitteille on voitava määrittää erikielisiä nimikkeitä, 3) käsitteitä täytyy pystyä hakemaan niiden termien eli labelin perusteella² ja 4) editorin on oltava vakaa ja kaatumisriskin pieni.

Editoreilla suoritettavat toiminnot olivat seuraavat: uuden käsitteen luominen hierarkian sisään ja ns. lehtikäsitteeksi³, yläkäsitteen lisääminen käsitteelle, erikielisten nimikkeiden lisääminen, assosiativisten suhteiden lisääminen sekä hakutoiminnon osalta katkaisuhaku ja haun rajaamismahdollisuudet. Näiden lisäksi tarkistettiin selailunäkymän toimivuus, ohjelman nopeus ja stabiilius sekä mahdollisuus ontologian muokkaushistorian selailuun.

Testauksessa käytettiin YSO-ontologiaa, joka sisältää n. 26000 käsitettä ja on kolmikielinen (suomi-ruotsi-englanti). Testauksessa käytettiin YSO-ontologian SKOS- tai OWL-versiota riippuen editorin tukemista tiedostoformaateista. Näiden ominaisuuksien lisäksi huomioitiin myös editorien hinta, lähdekoodin avoimuus sekä valmistajan tuen saatavuus.

Seuraavassa käydään yksittäin läpi ne paikallisesti asennettavat ja web-pohjaiset editorit, jotka täyttivät työryhmän asettamat vähimmäisvaatimukset. Kustakin editorista käydään läpi perustoiminnot ja todetaan esiin tulleet edut, haasteet ja puutteet. Lopuksi esitellään lyhyesti muut testatut editorit ja annetaan testituloksiin pohjautuva yhteenveto ja suositus.

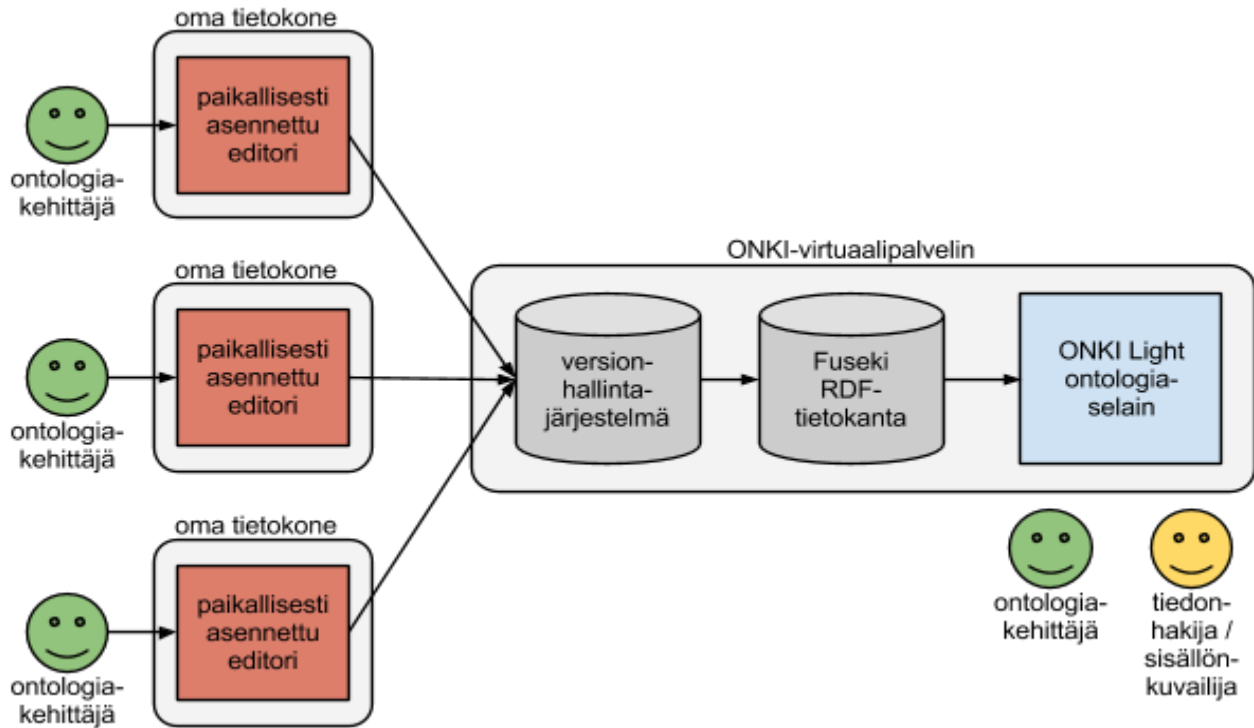
Maksullisille editoreille laskettiin myös vertailuhinnat skenaariolle, jossa ontologiaeditorin käyttäjälisenssejä on 5 kpl ja hintaan on laskettu aloitusmaksu sekä ylläpito yhteensä 4 vuoden ajalle. Vertailuhinnat luokiteltiin karkeasti kolmeen tasoon: edullinen (alle 20000 euroa), kohtalainen (50000 euroon asti) ja korkea (yli 50000 euroa). Lisäksi on huomioitu käyttäjälisenssien määrän kasvattamisen vaikutus hintaan.

² Ontologiassa jokaisella käsitteellä on varsinainen nimike (eng. preferred label, tekninen nimi prefLabel), josta monikielisessä ontologiassa on kullekin kielelle oma vastineensa. Lisäksi käsitteillä voi olla myös yksi tai useampi vaihtoehtoinen nimike (eng. alternative label, altLabel) ja myös näillä voi olla vastineensa eri kielille. Edelleen käsitteillä voi olla vanhoja (tai korvattuja) nimikkeitä (eng. old label) erikielisinä vastineineen.

³ Lehtikäsite on hierarkian alimmainen käsite, jolla ei ole enää alakäsitteitä.

4. Paikallisesti asennettavat editorit

Paikallisesti asennettavat editorit asennetaan kunkin ontologiakehittäjän omalle tietokoneelle. Mahdolliset lisenssikustannukset ovat käyttäjäkohtaisia.



Kaavio 1. Paikallisesti asennettu ontologiaeditori ONKI-arkkitehtuurissa.

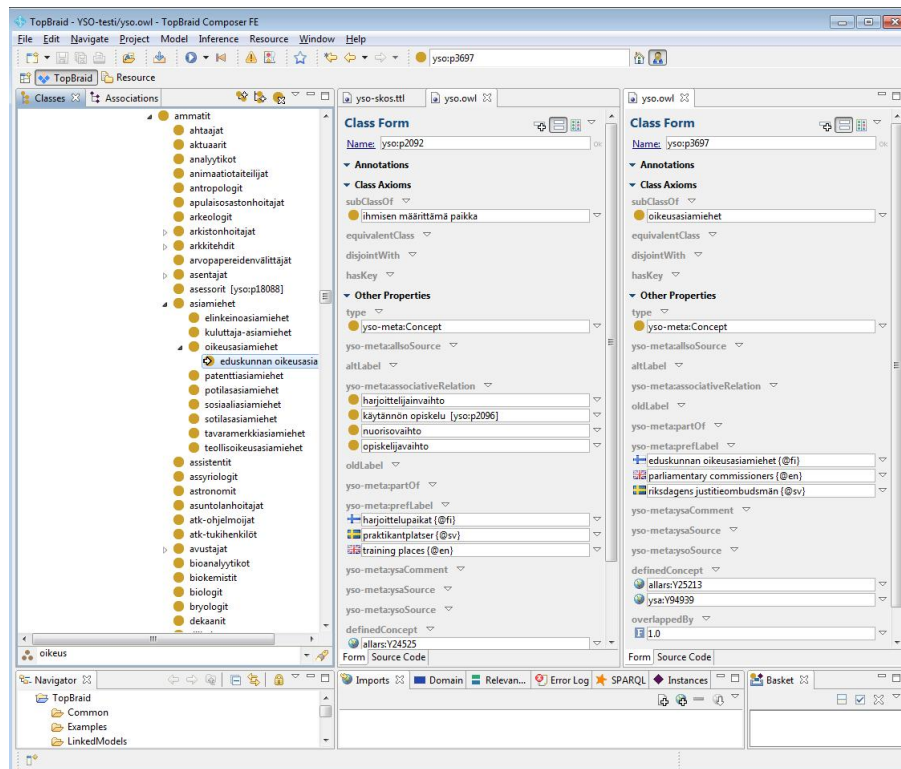
Paikallisesti asennettujen editorien hyvä puoli on se, että niissä on usein rikkaat toiminnot ontologian tarkasteluun ja muokkaukseen. Toisaalta paikallisia editoreita käytettäessä kahden käyttäjän ei ole mahdollista muokata samanaikaisesti samaa ontologiaa. Käytännössä yhteistyön koordinoimisen helpottamiseksi olisi lähes välttämätöntä käyttää yhteiskäytössä olevaa versionhallintajärjestelmää, joka voisi näkyä käyttäjille jaettuna verkkolevynä. Ontologian muokkaus tapahtuisi avaamalla ensin uusin versio ontologiasta verkkolevyltä, tekemällä tarvittavat muutokset editorissa ja lopuksi tallentamalla muutokset takaisin verkkolevylle, jolloin versionhallintajärjestelmä tallentaa uusimman version. Versionhallintajärjestelmä muistaa myös tarvittaessa aiemmat versiot, jotka voidaan palauttaa tarvittaessa käyttöön.

Ontologian julkaiseminen ONKI-järjestelmässä tapahtuisi siten, että versionhallintajärjestelmästä viedään uusin versio ontologiasta RDF-tiedostona ONKIn Fuseki-tietokantaan. Tämän voisi automatisoida niin pitkälle, että uuden ontologiaversioiden julkaisu ONKIn sisällä tapahtuisi yhdellä nappia painamalla.

4.1. TopBraid Composer: Free Edition ja Standard Edition

TopBraid Composer on yhdysvaltalaisen TopQuadrant-yrityksen ohjelmistoperhe, johon kuuluu eri laajuisia ja hintaisia ontologiaeditoreja. Näistä Maestro Edition ja Standard Edition ovat maksullisia ja Free Edition toiminnoiltaan kevyempi ja ilmainen. Testasimme Standard ja Free -versioita. Maestro Edition on versioista kallein ja suunniteltu ohjelmistokehitykseen: sen toiminnallisuus ontologiatyön kannalta ei eroa merkittävästi Standard-versiosta. Ohjelmistoperheeseen kuuluva web-pohjainen TopBraid EVN käsitellään myöhemmin erikseen. Standard-version vertailuhinta viidelle käyttäjälle asettuu edullisimpaan luokkaan. Lisenssimaksu on käyttäjäkohtainen, joten käyttäjämäärän kasvaessa lisenssikustannukset nousevat lineaarisesti.

Standard- ja Free -versiot ovat toiminnaltaan varsin yhteneväiset: tässä esitetyt huomiot koskevat molempia versioita ja eroavaisuudet mainitaan sikäli, kun niitä esiintyy. TopBraid Composer todettiin näkymältään selkeäksi ja perusolemukseltaan helppokäyttöiseksi ontologiaeditoriksi. Muutamaa myöhemmin esiteltävää teknistä ongelmaa lukuun ottamatta ontologiatyön perustoiminnot toimivat hyvin. TopBraid Composer osoittautui vakaaksi ohjelmaksi, joka ei kaadu tai jumiuudu. Jos virheilmoituksia tulee, ne ovat ymmärrettäviä ja vika on pääteltävissä. Varsinaista käsitteiden muokkaushistoriaa ei ole mahdollista seurata. Yksittäisen session lokitiedostoa on kuitenkin mahdollista lukea teknisten vikojen selvittämiseksi. Maksulliset versiot sisältävät yrityksen tarjoamaa suoraa teknistä tukea englanniksi: TopQuadrantin pääkonttori on Yhdysvaltojen länsirannikolla, mutta tukea lienee mahdollista saada myös Iso-Britannian konttorista.

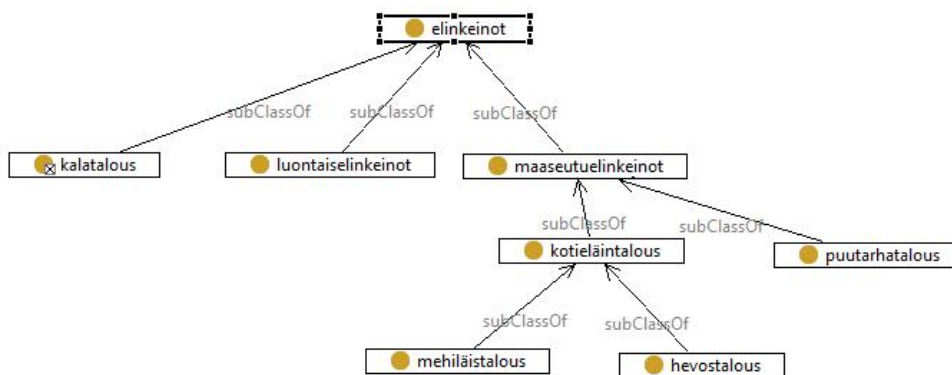


Kuva 1. TopBraid Composer Free Edition. Avoinna kaksi käsitettä.

Yksi TopBraid Composerin eduista on se, että siinä on mahdollista avata kaksi erillistä käsitteitä, joita voi editoida yhtä aikaa. Lisäksi käsitteidenäkymässä on mahdollista klikata toista käsitettä (jolla on jokin suhde avattuun käsitteeseen), jolloin näkyviin tulevat myös kyseisen käsitteen tiedot. Tällainen usean käsitteen tietojen yhtäaikainen tarkastelu nopeuttaa ja helpottaa ontologiatyötä. Lisäetuina ovat Standard (ja Maestro) -versiosta löytyvät *Graph-* ja *Diagram-*toiminnot, joilla käsitesuhteita voi visualisoida ja editoida graafisesti: toiminnoilla voi helposti esimerkiksi visualisoida tietyn palan hierarkiaa ja muokata tässä näkymässä myös käsitteiden välisiä suhteita (ks. kuva 2). Free-versiossa näitä visualisointitoimintoja ei ole.

TopBraid Composeriin liittyy myös jonkin verran teknisiä ongelmia, jotka ovat käytännön työn kannalta melko vakavia mutta kuitenkin ratkaistavissa. Suoritetussa testissä suurena haasteena oli saada SKOS-muotoisen ontologian suomenkieliset nimikkeet näkyviin hierarkiassa, mutta ongelma saatiin ratkaistua lataamalla SKOS Core -skeema projektiin ja asettamalla tämän jälkeen kieliasetukset kohdalleen. Ontologiatyössä tärkeä käsitteiden asetusten muokkaus osoittautui myös haastavaksi. Free-versiosta puuttuu muokkauksen mahdollistava Form Layout -näkyvä. Merkittävä ongelma on myös, että käsitettä muokattaessa käsitesuhteiden lisäämiseen liittyvä käsitteiden haku ei täysin toimi: testissämme haku kohdistui vain osaan YSO-käsitteistä, ei koko ontologiaan. Tiettyjä käsitteitä ei pystynyt liittämään toisiinsa, mutta ongelma onnistuttiin ratkaisemaan. Lisäksi moniperintäinen⁴ käsite ei tuoreeltaan sellaiseksi muokattuna näy kahdessa paikassa hierarkiassa, kuten sen pitäisi.

Yhteenvetona todetaan, että TopBraid Composer voidaan suositella ontologiatyöhön: se on perustoiminnoiltaan helppokäyttöinen ja muutamia hyödyllisiä erikoisominaisuuksia sisältävä työkalu, jossa teknisiä ongelmia on suhteellisen vähän.



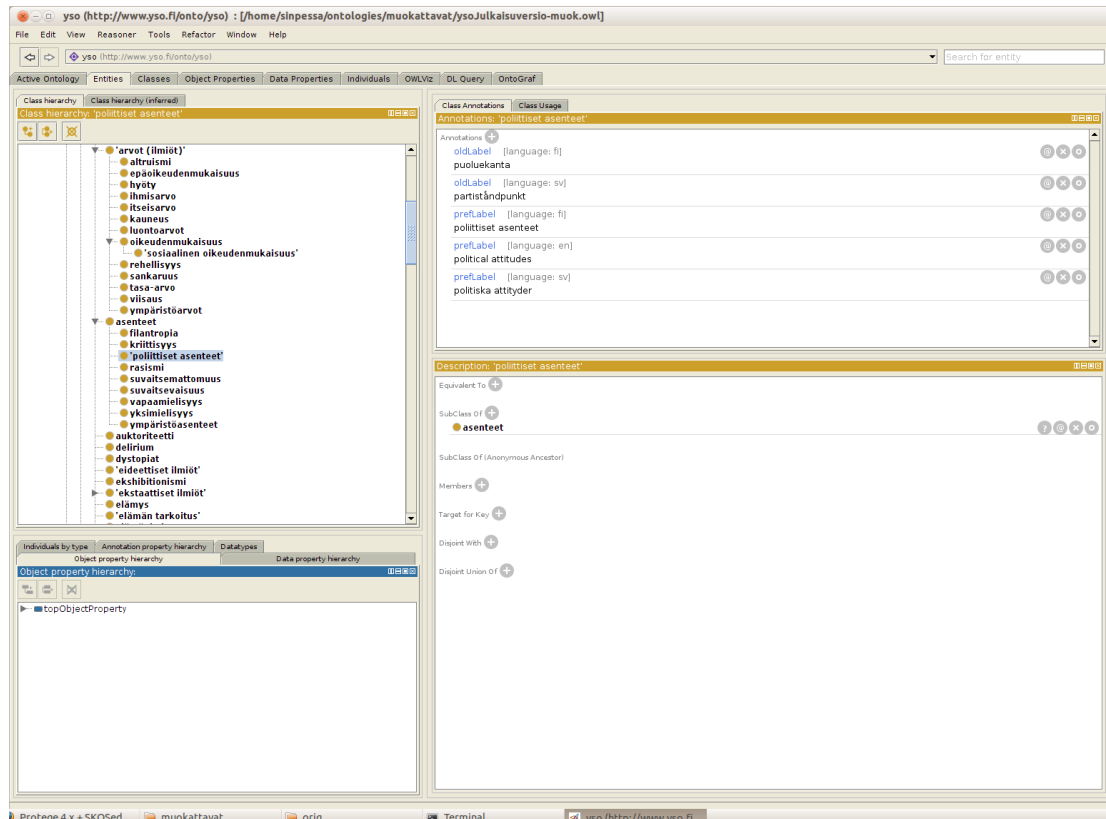
Kuva 2. TopBraid Standard Edition: esimerkki hierarkian visualisoinnista.

⁴ Moniperintä tarkoittaa yhden käsitteen sijaintia kahdessa tai useammassa hierarkian kohdassa. Esimerkiksi sudet-käsite voisi olla sekä koira-eläimet että petoeläimet -käsitteiden alakäsite.

4.2. Protégé 4.3.

Protégé on tunnetuimpia ontologiaeditoreja. Se on kehitetty ja sitä ylläpidetään Stanford Universityssä, Yhdysvalloissa. Se on ilmainen avoimen lähdekoodin ohjelma. Testattavana oli Protégén tuorein versio 4.3. Protégé on näkymältään selkeä editori, jonka hyviä puolia on se, että useaa käsitettä voi tarkastella yhtä aikaa erillisinä välilehtinä. Myös haku toimii hyvin ja moitteettomasti. Tosin mikäli haluaa esim. vaihtoehdoisen nimikkeen haun kohteeksi, on haku tehtävä käyttäen SPARQL-kyselykieltä.

Protégéhen liittyy kuitenkin myös paljon työtä hankaloittavia tekijöitä. Ominaisuuksien lisääminen käsitteelle toimii oudolla ja epäjärjestelmällisellä logiikalla: esimerkiksi assosiatiiviselle suhteelle ei ole omaa kenttää, vaan se lisätään jonkin toisen ominaisuuden kautta. Lisäksi kun luo uuden käsitesuhteen, ei voi käyttää hakua lainkaan, vaan on etsittävä haluttu käsite hierarkiasta, mikä on todella raskasta ja aikaa vievää. Aktivoitu käsite on helppo hukata hierarkianäkymässä. Lisäksi käsitteen URIn näkeminen on hankalaa. Virheiden selvityksen tekee ongelmalliseksi se, että dataa ei ole mahdollista tarkastella RDF-kolmikkoina. Ohjelma on ajoittain myös melko hidas ja välillä ponnahdusikkunat jäävät tyhjiksi, mikäli ohjelma jumittuu. Testissä todettiin, että YSON suomenkieliset nimikkeet eivät näkyneet, mutta niiden saaminen esiin onnistui suhteellisen pienellä selvityksellä. Lisäksi kriittinen huomio on se, että SKOS-muotoisen ontologian muokkaaminen ei onnistu Protégellä lainkaan.

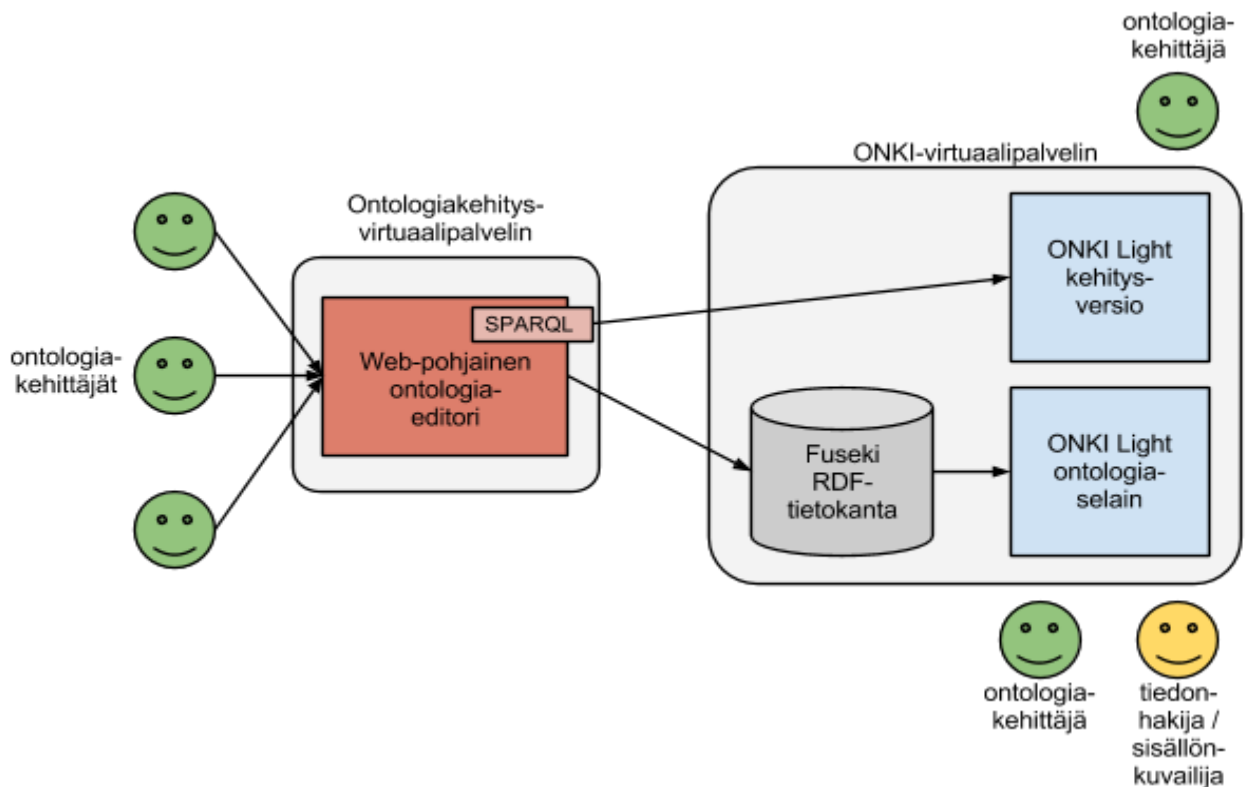


Kuva 3. Protégé 4.3.

Protégén ongelmat ovat sitä luokkaa, että sitä ei voi suositella ensisijaiseksi valinnaksi. Vaikka se onkin ilmainen ja perustuu avoimeen lähdekoodiin, on huomioitava, että teknisten ongelmien ratkaiseminen vaatii todennäköisesti teknistä asiantuntemusta ja resursseja.

5. Web-pohjaiset editorit

Web-pohjaista editoria käytettäessä editorista tarvitaan vain yksi asennus, joka voitaisiin tehdä yhteiskäytössä olevalle virtuaalipalvelimelle. Ontologiakehittäjät käyttävät editoria selaimen kautta. Lisenssikustannukset tässä mallissa muodostuvat editorin peruslissenssistä sekä mahdollisista käyttäjämäärään sidotuista lissenssimaksuista. Nämä kustannukset kuitenkin kattavat koko projektin tarpeet ja samaa editoria voivat tämän jälkeen käyttää useat organisaatiot.



Kaavio 2. Web-pohjainen ontologiaeditori ONKI-arkkitehtuurissa

Web-pohjaisen editorin hyviä puolia on se, ettei sitä erikseen tarvitse asentaa kunkin ontologiakehittäjän omalle tietokoneelle sekä se, että ontologioita voi useampi ihminen muokata samanaikaisesti. Ontologian viimeisin versio on koko ajan kaikkien nähtävillä.

Web-pohjaisiin editoreihin kuuluu usein myös SPARQL-rajapinta. Tämän rajapinnan kautta olisi esimerkiksi mahdollista toteuttaa erillinen ONKI-kehittäjäversio, joka näyttää reaaliajassa uusimman ontologiaversioon, joten ontologiakehittäjät voivat nähdä miten heidän tekemänsä muutokset heijastuvat ONKI-käyttöliittymään jo ennen varsinaista julkaisua.

Varsinaiset julkaisut kannattaa kuitenkin tehdä samaan tapaan kuin paikallisen editorin tapauksessa eli niin, että senhetkinen uusin versio viedään nappia painamalla Fuseki-tietokantaan, josta ONKI näyttää sen. Versionhallintajärjestelmän käyttö muutosten hallintaan on suositeltavaa, muttei samassa määrin välttämätöntä kuin paikallisten editorien tapauksessa, varsinkin jos editoriin itseensä sisältyy jonkinlainen versiohistorian tarkastelu- ja palautusmahdollisuus.

5.1. TopBraid EVN

TopBraid Composer -perheeseen kuuluu myös web-pohjainen, SKOS-muotoisten ontologioiden editoimiseen tarkoitettu TopBraid Enterprise Vocabulary Net (EVN). Testin perusteella EVN on toimiva ja helppokäyttöinen ontologiatyökalu. Hinnaltaan TopBraid EVN kuuluu korkeimpaan luokkaan jo viidellä käyttäjällä. EVN:stä maksetaan kertaluontoinen peruslisenssi ja tähän lisätään käyttäjälisenssi, jonka hinta määräytyy käyttäjien määrän mukaan: hintaryhmät ovat 1-5 käyttäjää, 6-10 käyttäjää, 11-20 käyttäjää ja 21-50 käyttäjää.

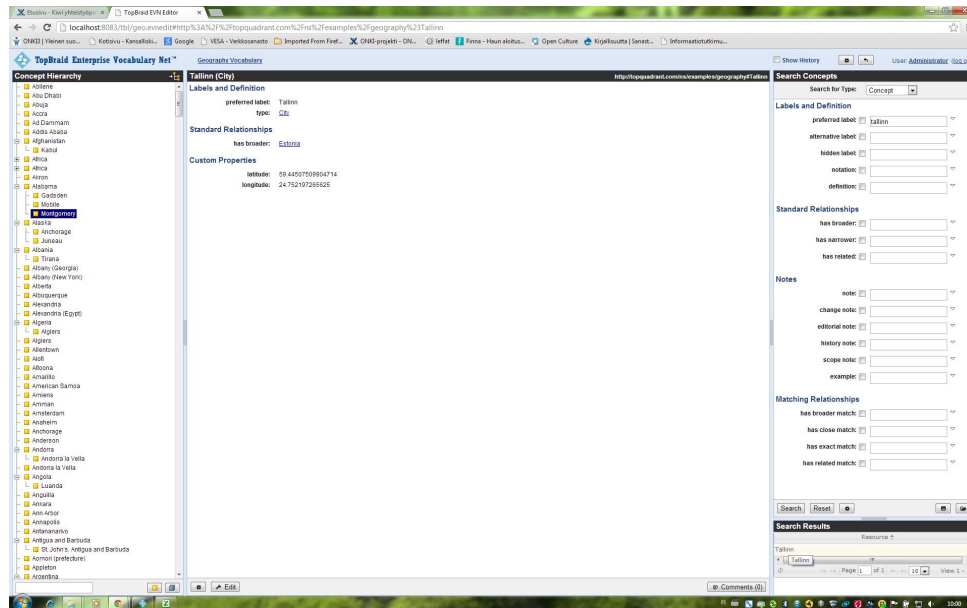
TopBraid EVN täytti testin kriittiset vaatimukset ja sisälsi muutenkin vain vähän ongelmakohtia. Siitä löytyy SPARQL-rajapinta, joka mahdollistaa erillisen kehittäjäversion ontologiasta. Selainnäkyminen on selkeä ja EVN toimii vakaasti ilman kangertelua. EVN eroaa monella tavalla TopBraidin koneelle asennettavista versioista. Maestro ja Standard Editionin graafista näkymää EVN:ssä ei ole, mutta toisaalta siinä on ominaisuuksia, joita ei löydy edellä mainituista.

EVN:n hyviin puoliin kuuluu se, että käsitteitä voi siirrellä hierarkiassa vapaasti niitä raahaamalla. EVN myös päivittää toisen käyttäjän tekemät muutokset hierarkiaan joka sivulatauksella, mikä onkin tärkeää, jos työtä tekee useampi samaan aikaan. Koska EVN on selainpohjainen, kahta (tai useampaakin) käsitettä voi helposti tarkastella ja muokata yhtä aikaa avaamalla lisää välilehtiä. Hakutoiminto on EVN:ssä monipuolinen ja pitkälle kehitetty: hakuja voi mm. tallentaa. EVN:ssä on kiinnitetty myös erityistä huomiota ontologian yhdessä muokkauksen toimivuuteen ja vaiheisiin: ontologiasta on ensin olemassa työversio, jossa tehtyjä muutoksia on mahdollista kommentoida ennen hyväksymistä. Eri käyttäjille voi antaa erilaisia oikeuksia. Editorissa on mahdollista myös selata käsitteiden muokkaushistoriaa, mikä voi tarjota ontologian ylläpitäjälle tärkeää tietoa. Ohjelma tarjoaa myös erilaisia tilastotietoja muokattavasta ontologiasta. EVN tukee myös eri käsitetyyppejä ja käsitteen tyyppin vaihtaminen on suoraviivaista. EVN:n käyttöliittymää on myös mahdollista muokata lisäämällä siihen plug-in-liitännäisiä: tähän tosin tarvitaan tietoteknistä asiantuntemusta.

EVN:ssä jokaiselle käsitteelle on muodostettu oma nettisivu, joka pohjautuu käsitteen URI-tunnisteseen. Tämä mahdollistaa käsitteen nopean löytämisen, sillä EVN-sivun avaamisen ja käsitteen etsimisen sijaan riittää, että liittyy halutun käsitteen osoitteen selaimeen.

Ongelmia tuli testissä vastaan verrattain vähän. Merkittävä ongelma on kuitenkin se, että EVN:ssä ei saa määriteltyä erikseen oletuskieltä hierarkian näyttämiseen, vaan oletuskieli luetaan käyttäjän selaimen asetuksista. Asetuksen vaihtaminen on yksinkertaista, mutta saattaa

vaikuttaa muihin käytettyjen nettisivujen kielisiin, mikäli EVN:ssä halutaan käyttää hierarkiassa eri kieltä kuin millä normaalisti haluttaisiin tarkastella nettisivuja.



Kuva 4. TopBraid EVN

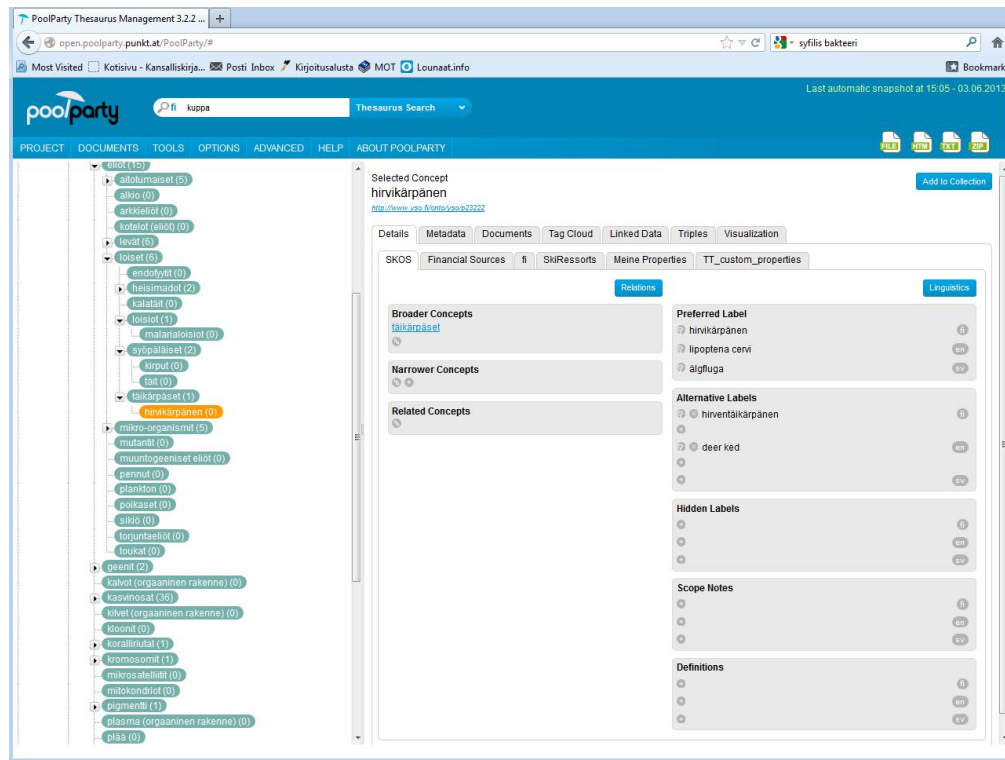
Yhteenvedon todetaan, että TopBraid EVN on suositeltava työkalu ontologiatyöhön. Web-pohjaisena ja usean käyttäjän yhtäaikaista työtä mahdollistavana se tarjoaa merkittävän edun koneelle asennettaviin ohjelmiin nähden. EVN on toimiva, monipuolinen ja helppokäyttöinen ontologiaeditori, josta löytyy perustoiminnallisuuden lisäksi hyödyllisiä erikoistoimintoja. Kielenmäärittämisongelma on kuitenkin syytä ratkaista ennen ylläpitotyön aloittamista.

5.2. PoolParty Thesaurus Server

PoolParty Thesaurus Server on itävaltalaisen Semantic Web Companyn web-pohjainen ontologiaeditori, joka täyttää perusvaatimukset. Ohjelmaa tarjotaan kahtena maksupakettina: a) kuukausilisenssinä, jolloin käyttäjämäärä on rajoittamaton ja lisenssi sisältää pilvipalvelun ja teknisen tuen tai b) kertaluontoisesti, mikä sisältää oman serveri-instanssin ja niin ikään rajoittamattoman määrän käyttäjiä. Molempien hinta asettuu kohtalaisen kategorian yläpäähän, mutta käyttäjämäärän kasvattaminen ei vaikuta hintaan.

PoolParty läpäisi asetetut vähimmäisvaatimukset ja todettiin monilta osin toimivaksi kokonaisuudeksi. Myös PoolPartysta löytyy muokkaushistoria ja tilastotietoa. PoolPartysta löytyy niin ikään SPARQL-rajapinta. Joitakin puutteita kuitenkin havaittiin, jotkut vähäpätöisiä, toiset hieman vakavampia. PoolParty on nopea ja toimii vakaasti. Se on näkymältään suhteellisen selkeä ja helppokäyttöinen. Web-pohjaisena sen avulla voi useampi henkilö muokata samaa ontologiaa yhtä aikaa. Erikoisominaisuuksiin kuuluu, että käsitettä luotaessa

PoolParty määrittää itse luodun käsitteen URIn: tämä poistaa riskin kahden identtisen URIn syntymisestä.



Kuva 5. PoolParty.

Kriittisin puute liittyy hakuun ja siihen sisältyvään automaattiseen täydennykseen: ehdotettujen hakutulosten järjestys on epälooginen (paras osuma ei tule ensimmäisenä) eikä hakutuloksia myöskään pysty vierittämään alaspäin, jos tuloksia tulee paljon. Sama ongelma tulee vastaan myös käsitesuhteita lisättäessä, koska siinä käytetään tavallisesti myös haun automaattista täydennystä. YSON kohdalla tämä on ongelmallista, sillä tietyntalokuisia käsitteitä (esim. oikeus-) on paljon. PoolPartyn kehittäjien mukaan automaattisen hakutoiminnon palauttamaa tulost määrää on mahdollista nostaa, mutta tämäkään ei välttämättä ratkaise ongelmaa kaikissa tapauksissa.

Toinen vastaan tulleista PoolPartyn ongelmista oli hierarkian päivityvyys: testissä havaittiin, että käsitteen nimikettä muuttaessa tai yläkäsitettä vaihtaessa tieto ei päivittynyt hierarkianäkymään toisille ontologiaa samanaikaisesti selanneille. Toisaalta käsitteen tyyppiä ei voi vaihtaa, koska PoolParty ei tue eri käsitetyyppejä. Käsitteenäkymän asetuksia ei pysty vapaasti muuttamaan, mutta valittavana on laaja ja suppea valikoima SKOS-kenttiä. Lisäksi erillisille välilehdille on mahdollista määrittää omia ominaisuuksia ja suhdetyyppejä. PoolPartyyn liittyy lisäksi muutamia epäkäytännöllisyyksiä: esimerkiksi aktivoitu käsite hukkuu helposti hierarkianäkymässä ja sommittelun vuoksi käyttäjä joutuu skrollaamaan paljon.

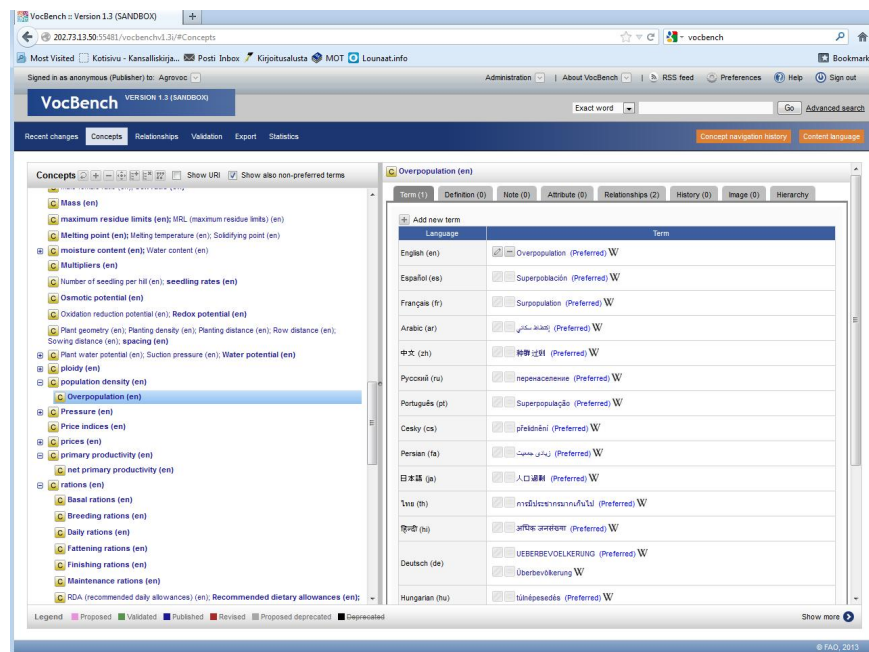
Puutteistaan huolimatta PoolParty toimii melko hyvin ja on yksi vertailun helppokäyttöisimmistä editoreista.

5.3. VocBench

VocBench on YK:n elintarvike- ja maatalousjärjestö FAOn kehittämä pääasiallisesti SKOS-muotoisten sanastojen editointiin tarkoitettu työkalu. Se on ilmainen ja seuraava versio julkaistaan avoimella lähdekoodilla.⁵ Tässä vertailussa testattiin ohjelman versiota 1.3.1. Vocbenchiin ei ollut mahdollista tuoda YSO-ontologiaa, joten kaikkia testejä ei voitu tehdä. Keskeiset toiminnot ja tiedot saatiin kuitenkin selvitettyä.

VocBench on web-pohjainen ja mahdollistaa monen käyttäjän yhtäaikaisen käytön. Editori on melko selkeä ja täyttää ontologiatyön kriittiset vaatimukset. Se toimii vakaasti ja sen eduksi voidaan katsoa, että käsitteiden muokkaushistoriaa voi selata. Muokatut asiat voi myös peruuttaa *Validation*-näkyvän kautta. VocBenchiin kuitenkin liittyy monia ongelmia ja epäkäytännöllisyyksiä, jotka tekevät ontologiatyöstä epävarmaa.

Ongelmia ovat mm. se, että hierarkianäkymän perusteella ei voi tietää, mikä käsite on aktivoitu: sekä aiemmin klikatut käsitteet että nykyinen ovat väriltään sinisiä. Nimikkeen muutokset eivät päivity hierarkianäkymään suoraan, mutta selaimen reload-toiminnon kautta päivitys onnistuu. Käsitteen ominaisuudet näkyvät hankalasti eri välilehtinä, mikä hidastaa työskentelyä. VocBench toimii myös melko hitaasti eikä ohjelma ladatessaan ilmoita, mitä on tapahtumassa. Lisäksi kerran poistetut käsitteet jäävät hierarkiaan ja ovat edelleen muokattavissa.



Kuva 6. VocBench.

⁵ FAO:n mukaan VocBenchistä on jo olemassa avoimeen lähdekoodiin perustuva versio 2, mutta kirjoitushetkellä tämä versio ei ole vielä testattavissa. Kyseisessä versiossa on myös SPARQL-rajapinta.

Vaikka editorin perustoiminnallisuus on kunnossa, ei VocBench ole web-pohjaisten editorien suositeltavimmasta päästä – ainakaan ohjelman tämänhetkinen versio. Se, että ohjelma on ilmainen ja siitä on suunnitteilla uusi avoimeen lähdekoodiin perustuva versio, on kuitenkin varteenotettava seikka ja tilannetta tämän editorin kohdalta onkin hyvä seurata tulevaisuudessa.

5.4. Muut web-pohjaiset editorit

Testattaviin kuuluivat myös iQvoc ja TemaTres. Näistä ensin mainittu on sanastojen editointiin tarkoitettu saksalaisen innoQ-yrityksen kehittämä editori. Se ei kuitenkaan täyttänyt asetettuja vaatimuksia: hierarkia ei ole iQvocissa koko ajan näkyvillä vaan häviää, kun siirrytään käsitteenäkymään. Myöskään alimerkkijonohaku ei toimi eli käsitteen nimike pitää kirjoittaa kokonaisuudessaan hakukenttään. Näistä syistä iQvocia ei voi suositella ontologiatyöhön.

TemaTres on avoimen lähdekoodin editori, jonka on kehittänyt Diego Ferreyra. Se on näkymältään epäselvä, epäkäytännöllinen ja keskeneräisen oloinen. Hierarkiaa ei näy lainkaan, käsitteet näkyvät aakkosjärjestyksessä ja käsitteenäkymä latautuu hankalasti käsillä olevaan välilehteen sanastonäkymän tilalle eli käsitelistaa ja yksittäistä käsitettä ei voi tarkastella yhtä aikaa. Näistä syistä TemaTresia ei voi lainkaan suositella ontologiatyöhön.

6. Yhteenveto ja suositukset

Ohjelmiston valinnassa keskeisiä kriteerejä ovat toiminnallisuus ja hinta. Toiminnallisuuteen liittyen työryhmä on laatinut taulukon (ks. alla), jossa testatuille editoreille ja niiden ominaisuuksille on annettu pisteet asteikolla yhdestä viiteen (viisi on paras pistemäärä). Mitä hintaan tulee, maksullisten editorien ohella löytyy useita ilmaisia ontologiaeditoreja, mutta näistä kaikki eivät ole luotettavia työkaluja. Tärkeä kysymys on myös se, halutaanko käyttää verkkopohjaista ohjelmaa, joka mahdollistaa monen käyttäjän yhtäaikaisen ontologian muokkaamisen. Tässä on syytä priorisoida tarpeet ja miettiä, mihin ja miten organisaatio haluaa käyttää editoria ja ontologiaansa. Vertailussa luotettaviksi havaitut verkkopohjaiset ontologiaeditorit ovat kalliita tuotteita, joten mikäli resurssit ovat puutteelliset, on syytä harkita, riittäisikö koneelle asennettava ontologiaeditori, jota muokkaa vain yksi ylläpitäjä kerrallaan.

Koneelle asennettavista editoreista suositellaan TopBraid Composer -ohjelmistoperhettä. Ilmainen TopBraid Composer Free Edition sisältää ontologiatyön kannalta kaiken olennaisen. Maksullinen Standard Edition tarjoaa enemmän toimintoja. Kalliimpi Maestro Edition ei ole suositeltava, sillä sen tuoma lisäarvo ei ole ontologiatyön kannalta merkittävä. Ilmainen Protégé on TopBraidin Composer Free Editioniin nähden selkeästi epäluotettavampi ohjelma ja siksi sitä ei suositella.

Web-pohjaisista ilmaisista ohjelmista ainoa, joka täyttää ontologiatyön perusedellytykset on VocBench. Tärkeimmät toiminnot ovat kunnossa, mutta ohjelma ei kuitenkaan ole kaikilta osin

täysin luotettava: ohjelman sivuilla mm. todetaan, että asennus on vaikea eikä teknistä tukea ole tarjolla. VocBenchin kehitystyön seuraaminen on kuitenkin suositeltavaa, sillä uusi versio ohjelmasta on tulossa. Maksullisista ohjelmista suositellaan ensisijaisesti TopBraid EVN:ää, toissijaisesti PoolPartyä. TopBraid EVN todettiin testatuista ohjelmista parhaaksi: siinä yhdistyy helppokäyttöisyys, monen käyttäjän yhtäaikaisuus ja monipuolinen toiminnallisuus. Toisaalta on huomioitava, että TopBraid EVN oli hinnaltaan vertailun korkein ja myös PoolPartyn hinta oli kohtalaisen korkea.

Editori	Tyyppi	Helppo- käyttöi- syys	Perus- toiminno t	Lisä- toiminno t	Luotetta- vuus	Hinta	Lähdekoodi
TopBraid Free	Paik.	4	4	3	4	ilmainen	suljettu
TopBraid Standard	Paik.	4	4	4	4	edullinen	suljettu
Protégé 4.3	Paik.	3	3	3	2	ilmainen	avoin
PoolParty Thesaurus Server	Web	4	4	3	3	kohtalainen	suljettu
TopBraid EVN	Web	4	5	4	4	korkea	suljettu ⁶
VocBench 1.3.1	Web	3	2	3	2	ilmainen	toistaiseksi suljettu

Taulukko 1. Ontologiaeditorien ominaisuuksien pisteytys.

7. Projektiin liittyvää verkkomateriaalia

ONKI-projektin Wiki

<https://wiki.helsinki.fi/display/ONKI>

FinnONTO-tutkimushankkeen kotisivu

<http://www.seco.tkk.fi/projects/finnonto/index.fi.php>

Valtiovarainministeriö: Hallinnon ja aluekehityksen ministerityöryhmä puolsi kansallisen ontologiapalvelun rahoitusjärjestelyä vuodelle 2013

http://www.vm.fi/vm/fi/03_tiedotteet_ja_puheet/01_tiedotteet/20121115Hallin/name.jsp

⁶ TopBraid EVN:n lähdekoodi on suljettu, mutta ohjelman kylkeen on mahdollista lisätä ulkopuolisia plug-in-liitännäisiä käyttäjän tarpeiden mukaan.