

Afhverju finn ég ekki allt efni eftir höfundinn?

Rannsókn á vensluðum gögnum og nafnmyndastjórnun

Höfundur: Telma Sigfúsdóttir

Inngangur

Notendur Leitir.is¹ telja sig örugga um að finna allt efni eftir höfund þegar þeir leita að nafni á Leitir.is. Staðreyndin er hins vegar sú að þeir fá ekki tæmandi niðurstöður þegar þeir leita að nafnmynd einstaklings í Leitir.is, heldur einungis niðurstöður úr sumum gagnasöfnum og ekki öllum. Í tengslum við MA-nám mitt við Háskóla Íslands og starf mitt á Landsbókasafni Íslands – Háskólabókasafni gerði ég meistaranám í upplýsingafræði á vensluðum gögnum og nafnmyndastjórnun. Markmið rannsóknarinnar var að finna mögulegar leiðir til þess að opna gagnasöfnin sem eru leitarbær í Leitir.is og tengja þau saman. Með því að tengja saman gagnasöfnin er hægt að safna saman öllum nafnmyndum einstaklings og gera notendum kleift að finna allt efni eftir sama aðila, óháð því hvaða nafnmynd einstaklingsins er notuð í leitinni. Tilgangurinn er að stuðla að betra aðgengi notenda að upplýsingum.

Rannsóknin

Rannsóknin, sem var eigindleg, beindist að gagnasöfnum Landsbókasafns Íslands – Háskólabókasafns og aðgengi að þeim í leitargáttunum Lbs.leitir.is og Leitir.is. Hugmyndin með Leitir.is er að safna saman efni úr mismunandi gagnasöfnum og gera það leitarbært á einum stað, til hægðarauka fyrir notendur. Gögnin sem Leitir.is sækir í eru úr ólíkum gagnasöfnum og á mismunandi gagnasniði, s.s. MARC (MACHINE-Readable Cataloging) og DC (Dublin Core). Nafnmyndaskrárnar í gagnasöfnunum eru einnig ólíkar og í þeim er notast er við mismunandi auðkenni, s.s. URI (Uniform Resource Identifier), ORCID (Open Researcher and Contributor ID), ISNI (The International Standard Name Identifier) og VIAF (Virtual International Authority File). Nafnmyndaskrárnar tengjast ekki að neinu leyti.

Þegar leitað er eftir höfundi eða nafnmynd einstaklings í Leitir.is er það komið undir skráningunni í hverju gagnasafni fyrir sig hvaða heimtur notandinn fær. Leit með ákveðinni nafnmynd einstaklings skilar niðurstöðum úr ákveðnum gagnasöfnum en undanskilur niðurstöður úr öðrum gagnasöfnum. Notandinn hefur hvorki hugmynd um hvernig kerfið er sett upp né hugmyndaflug til þess að leita að mörgum mismunandi nafnmyndum einstaklings.

Sex sérfræðingar úr bókasafnaheiminum og hugbúnaðargeiranum, bæði hérlandis og erlendis, voru valdir til að taka þátt í rannsókninni. Gagnasöfnunin gekk út á að varpa ljósi á reynslu þeirra af tæknilegum leiðum sem hægt er að fara til þess að tengja saman gagnasöfn sem

¹ Leitir.is er byggt á hugbúnaðinum Primo VE frá Ex Libris. Leitir.is er leitargátt fyrir alla en hver safnakjarni og hvert safn í bókasafnasamlaginu á Íslandi hefur sína eigin notendasýn (e. view). Leit í Lbs.leitir.is á háskólanetinu veitir notendum aðgang að efni Landsbókasafns, efni í landsaðgangi og séráskriftum Háskóla Íslands. Til einföldunar verður rithátturinn Leitir.is notaður hér eftir um bæði Lbs.leitir.is og Leitir.is, nema annað sé sérstaklega tekið fram.

innihalda gögn á mismunandi gagnasniði. Rannsóknin leiddi í ljós að brýn þörf er á að bæta aðgengi notenda að upplýsingum í Leitir.is.

Hvað er nafnmynd, nafnmyndaskrá og nafnmyndastjórnun?

Nafnmynd sýnir hvernig leitaratriði eins og til dæmis nafn höfundar birtist í heimild eða gagnagrunni. Framsetning nafnmyndar getur verið mismunandi eftir heimildum og gagnagrunnum, til dæmis er nafnmyndin Jón Atli Benediktsson notuð í nafnmyndaskrá Gegnis, en nafnmynd sama einstaklings í ISNI nafnmyndagrunninum er Benediktsson, Jon Atli.

Hefðbundin nafnmyndastjórnun í bókasafnskerfum felst í því að velja eina nafnmynd sem leitaratriði (Sigrún Hauksdóttir, 2005). Ef þörf er á, eru notaðar tilvísanir í víkjandi nafnmyndir til þess að stuðla að samræmdri skráningu og gera allar nafnmyndirnar leitarbærar. Nafnmyndastjórnun getur líka falist í því að safna öllum afbrigðum saman en gera ekki endilega einni nafnmynd hærra undir höfði en öðrum. Nafnmyndir eru skráðar í sérstakar nafnmyndaskrár þar sem hver nafnmyndafærsla fær sitt einkvæma auðkenni. Risan (2020, 27. októberB) skilgreinir nafnmyndaskrá sem skrá yfir ákjósanlega ritun mannanafna og mismunandi rithátt þeirra. Tilgangur nafnmyndaskráa er að tryggja samræmda skráningu, koma í veg fyrir innsláttarvillur og tryggja að notandi geti fundið allt efni eftir höfund, sama hvaða nafnmynd úr nafnmyndafærslunni hann notar í leitinni. Risan segir einnig að nafnmyndir séu mikilvægur hlekkur í ferlinu að tengja saman ólík gagnasöfn (2020, 27. októberA). Smith-Yoshimura (2020, 16. apríl) bendir á að krefjandi og erfitt geti reynst að samræma lýsigögn úr mörgum mismunandi gagnasöfnum þegar ekkert samræmi er í nafnmyndastjórnuninni og engin tenging á milli gagnasafnanna.

Lýsigögn og vensluð gögn

Skilgreina má lýsigögn (e. metadata) sem gögn um gögn og að þau lýsi öðrum gögnum (Risan, 2020, 27. októberB; W3C, 2013). Þau eru aðskilin hinum eiginlegu gögnum, veita upplýsingar um þau og auðvelda leitir að þeim. Lýsigögn lýsa ekki einungis öðrum gögnum heldur eru þau notuð til að stjórna þeim, gera þau aðgengileg, varðveita þau og tengja saman. Lýsigögn eru alltaf sett fram á formfastan og fyrirfram skilgreindan hátt svo þau séu merkingarbær, til dæmis þarf að vera hægt að lesa úr þeim hvort „1466“ sé blaðsíðutal eða ártal. Síðan er hægt að safna því saman sem hefur verið skilgreint sem sömu tegundar í leitarskrár eða „indexa“ sem eru leitarbærir. Hver index inniheldur þá aðeins eina ákveðna tegund gagna. Höfundarindex er til dæmis leitarskrá sem leitar eingöngu í þeim gögnum sem eru skilgreind sem nöfn höfunda.

Vensluð gögn (e. linked data) innihalda lýsigögn sem eru læsileg bæði vélum og fólki. Gagnasöfn geta verið á mismunandi lýsigagnasniði, eins og MARC og DC. Til þess að gera gagnasöfn aðgengileg er mikilvægt að unnið sé eftir fyrirfram skilgreindri lýsigagnastefnu sem tekur til þess að lýsigögn séu samræmanleg. Vensluð gögn eru gögn sem hafa verið opnuð með ákveðnum aðferðum sem gerir vélum kleift að tengja þau saman. Búin eru til mynstur sem eru véllesanleg með tækni merkingarvefsins (e. semantic web). Þegar vensluð

gögn eru birt á vefnum og gerð aðgengileg öllum eru þau skilgreind sem opin vensluð gögn (e. linked open data) (W3C, 2013, 2015). Með því að breyta gögnum í opin vensluð gögn er opnað á samnýtingarmöguleika þannig að hægt sé að auðga lýsigögnin. Þannig verður mögulegt að tengja saman skyld viðföng, þótt þau séu sett fram með mismunandi hætti í ólíkum gagnasöfnum (Europeana, e.d.).

Nafnmyndaskrár gagnasafna Landsbókasafns Íslands – Háskólabókasafns og staðan á Íslandi í dag

Fyrirtækið Ex Libris er framleiðandi bókasafnskerfisins Ölmú og leitargáttarinnar Primo VE. Alma er veflægt bókasafnskerfi sem notað er á bókasöfnunum á Íslandi og var tekið í notkun í júní 2022. Kerfið ber heitið Gegnir á íslensku, óháð því hvaða hugbúnaður liggur að baki. Notendaviðmótið Primo VE er síðan byggt ofan á Ölmú. Á Íslandi gengur leitargáttin undir nafninu Leitir.is. Bæði Alma og Primo VE eru skýjalausnir sem eru hýstar í gagnaveri Ex Libris í Amsterdam. Eins og í bókasafnskerfinu Aleph500, sem var í notkun á undan Ölmú, fer öll vinnsla eins og umsýsla með útlánnum, lánþegum, millisafnalánnum, skráningu og aðföngum, fram í bókasafnskerfinu. Að auki fer umsýsla rafrænna áskrifta fram í Ölmú sem og stillingar og uppsetning Primo VE sem áður var gert í sérstökum kerfum, SFX og Primo BackOffice.

Gagnasöfn Landsbókasafns Íslands – Háskólabókasafns sem eru leitarbær í gegnum Lbs.leitir.is eru bæði innri og ytri gagnasöfn. Innri gagnasöfn eru annars vegar bókfræðigrunnurinn í Ölmú og hins vegar bókfræðigögn sem fengin eru frá Ex Libris og eiga við um rafrænt áskriftarefni. Ytri gagnasöfnin eru Opin vísindi² og Skemman³. Gögn úr ytri gagnasöfnunum IRIS⁴ og Hirslan⁵ voru áður fyrr sótt og gerð leitarbær í Leitir.is en því hefur nú verið hætt. Hluti gagnanna úr IRIS og Hirslan er leitarbær í Leitir.is með öðrum leiðum þótt þau séu ekki sótt beint.

Hvaða þýðingu hefur kerfisuppsetningin fyrir notandann?

Gagnagrunnarnir sem eru leitarbærir í Leitir.is eru á ólíkum lýsigagnasniðum, ýmist á MARC-sniði eða DC. Notuð eru ólík auðkenni í gagnagrunnunum og grunnarnir tengjast ekki að neinu leyti. Það veldur því að ákveðnar færslur verða alltaf útilokaðar í leitarniðurstöðum. Jafnvel þótt leitarstrengurinn (þ.e. röðun nafniða og stafsetning þeirra) sé eins, falla leitarniðurstöðurnar ekki saman ef gögnin eru á mismunandi gagnasniði. Í okkar íslenska umhverfi eru um 300 bókasöfn af öllum gerðum saman í samlagi, grunnskólasöfn, framhaldsskólasöfn, almenningssöfn, sérfræðisöfn, háskólasöfn og Landsbókasafn Íslands – Háskólabókasafn. Öll söfnin deila einni nafnmyndaskrá sem er lokuð inni í bókasafnskerfinu

2 Opin vísindi er rafrænt varðveislusafn fyrir efni í opnum aðgangi í íslenskum háskólum.

3 Skemman er rafrænt varðveislusafn íslenskra háskóla yfir lokaverkefni nemenda og rannsóknarrit kennara og fræðimanna.

4 IRIS (Icelandic Research Information System) er kerfi sem sýnir rannsóknavirkni íslenskra háskóla og stofnana sem eiga aðild að kerfinu. Gögnin í IRIS hafa verið fjarlægð úr Leitir.is (desember 2022) en eru enn leitarbær í prófunarumhverfi Leitir.is og verða skoðuð þar.

5 Hirslan er vísinda- og fræðsluefnissafn Landspítala. Skráningu í kerfið hefur verið hætt og lýsigögnin keyrð inn í IRIS.

Ölmu og tengist einungis bókfræðifærslum í því kerfi, en brýn þörf er á að opna gögnin til þess að geta tengst öðrum nafnmyndaskrá. Eins og staðan er í dag, á nafnmyndaskráin í Ölmu enga möguleika á að tengjast öðrum nafnmyndaskrá, ekki einu sinni nafnmyndaskrá in öðrum gagnagrunnum sem eru leitarbærir í gegnum Leitir.is.

The screenshot shows the search results for the article "Remote Sensing Image Classification Using Attribute Filters Defined over the Tree of Shapes" by Jón Atli Benediktsson, Jon Atli, Cavallaro, Gabriele; Dalla Mura, Mauro; Plaza, Antonio. The page includes a sidebar with navigation options (EFST, SENDA TIL, SKOÐA RAFRÆNT, NÁNAR, TENGLAR), a search bar, and a list of metadata for the article. The metadata includes the title, authors, publisher (IEEE Geoscience & Remote Sensing Society), year (2016), DOI (10.1109/TGRS.2016.2530690), and a list of keywords (Gray-scale, Morphology, Remote sensing, Shape, Spatial resolution, Vegetation). The page also features a "Skoða rafrænt" section with a search bar and a "Nánar" section with a list of metadata.

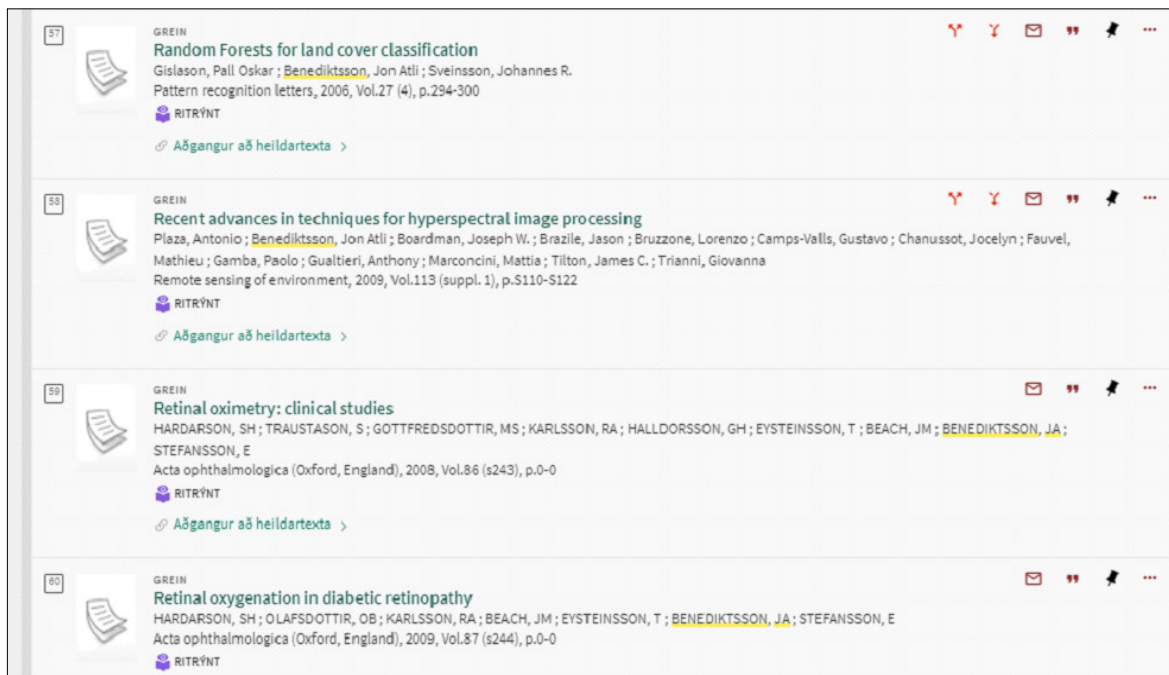
Titill	Remote Sensing Image Classification Using Attribute Filters Defined over the Tree of Shapes
Aðild	Benediktsson, Jon Atli > Cavallaro, Gabriele > Dalla Mura, Mauro > Plaza, Antonio >
Útgefandi	IEEE Geoscience & Remote Sensing Society
Útgáfuár	2016
Skyldir titlar	Tengt við : DOI: 10.1109/TGRS.2016.2530690 > Tengt við : info:eu-repo/grantAgreement/EC/FP7/606962 >
Birtist í	Gabriele Cavallaro, Mauro Dalla Mura, Jón Atli Benediktsson, Antonio Plaza. "Remote Sensing Image Classification Using Attribute Filters Defined over the Tree of Shapes." IEEE Transactions on Geoscience and Remote Sensing (Volume:54, Issue:7) >
Efni	Gray-scale > Morphology > Remote sensing > Shape > Spatial resolution > Vegetation >
Auðkenni	ISSN : 0196-2892
Tungumál	Enska
Abstract	Peer Reviewed EU FP7 Theme Space project North State, The program J. Verne 2014 under Project 31936TD, and the INS2I JCJC 2015 under project "IDES"

MYND 1: LEITARNIÐURSTADA ÚR LBS.LEITIR.IS. LEITARSTRENGUR: JÓN ATLI BENEDIKTSSON

Ekki er hægt að gera ráð fyrir að notendur séu sérfræðingar í leitaraðferðum og notendur finna hvergi upplýsingar um kerfisuppsetninguna. Miklar líkur eru á að notandinn treysti því að hann fái tæmandi yfirlit yfir verk höfundar þegar hann leitar en reyndin er allt önnur eins og áður hefur komið fram. Mögulegt er að einhverjir notendur viti af þessari brotalöm og geti þá brugðist við með því að breyta leitinni á meðan aðrir notendur missa af leitar- niðurstöðum.

Eftirfarandi dæmi um hindrunina sem notendur verða fyrir sýna að brýn þörf er á því að brugðist verði við þessu vandamáli og lausnir fundnar til að bæta aðgengi notenda að upp- lýsingum. Á mynd 1 má sjá leit í leitarprófilnum⁶ „Mitt bókasafn“ á Lbs.leitir.is. Þar er leitað að „Jón Atli Benediktsson“ og leitin skilar greininni „Remote Sensing Image Classification Using Attribute Filters Defined over the Tree of Shapes“ úr Opnum vísindum.

6 Leitarprófill er skilgreint mengi leitarbærra gagna sem notandinn getur valið úr fellilista í leitarglugganum á Leitir.is



MYND 2: SEINNI LEITARNIÐURSTADA ÚR LBS.LEITIR.IS. LEITARSTRENGUR: BENEDIKTSSON, J. A.

Nafnmyndin í Opnum vísindum er „Benediktsson, Jon Atli“ en þegar leitað er í sama leitarprófil með leitarstrengnum „Benediktsson, J. A.“ finnst þessi tiltekna færsla ekki eins og mynd 2 sýnir. Leitarniðurstaðan er í stafrófsröð og þar sést að greinin „Remote Sensing Image Classification Using Attribute Filters Defined over the Tree of Shapes“ kemur ekki fram í niðurstöðulista.

Hvað getum við gert?

Þrátt fyrir að Leitir.is sækji gögn úr mismunandi gagnagrunnum tengjast gögnin ekki saman í leitarniðurstöðum. Á Landsbókasafni Íslands – Háskólabókasafni er einnig boðið upp á leit í hverjum gagnagrunni fyrir sig, eins og Skemman.is, Opinvisindi.is og Iris.is. En hugmyndin með Leitir.is er að safna saman mismunandi gagnasöfnum í eina leitargátt til hægðarauka fyrir notendur svo þeir þurfi ekki að leita í mörgum kerfum og leitarvélum.

Tilgangurinn með rannsókninni er að nýta hana til að leysa vandann við lélegar leitarheimtur þegar leitað er að nöfnum einstaklinga í Leitir.is. Rannsóknin sýnir að margar leiðir eru færar til að tengja gögnin saman. Um leið og ákveðið hefur verið að fara af stað með slíka vinnu þarf að byrja á greiningarvinnu til að ákveða hvaða leið er ákjósanlegust fyrir okkur. Það er ljóst að hvaða leið sem verður fyrir valinu krefst hún mikillar tæknikunnáttu og verður kostnaðarsöm og tímafrek.

Það fyrsta sem við þurfum að gera er að opna gögnin og koma á sameiginlegum auðkennum í öllum gagnagrunnum til að eiga möguleika á að vensla þau saman. Nafnmyndafærslur þurfa að innihalda einkvæm auðkenni til þess að hægt sé að tengja saman tvær eða fleiri mismunandi nafnmyndir sama einstaklings og koma í veg fyrir að tveir einstaklingar með sama eða svipað nafn fái eina og sömu nafnmyndafærsluna.

Einn möguleiki sem við höfum til að leysa málið er að nýta tækni sem Ex Libris býður upp á og nota API-vefþjónustur til að láta gögnin „tala“ saman. Með þeirri tækni er hægt að senda API-köll á milli gagnagrunnanna og það ætti að skila betri leitarheimtum á Leitir.is. Með þeirri aðferð eru gagnasöfnin sjálf ekki tengd saman, heldur er hvert þeirra tengt við Ölmú og Leitir.is með API-vefþjónustum. Það myndi skila betri niðurstöðum en nú er raunin, en ekki er víst að leiðin myndi skila fullkominni niðurstöðu. Önnur leið er að smíða utanálíggjandi nafnmyndaskrá þar sem nafnmyndaskrá allra gagnasafnanna er hlaðið inn og gögnin „tala“ við bókasafnskerfið Ölmú og Leitir.is. Aftur er forsenda fyrir betri leitarheimtum sú að allar nafnmyndir innihaldi einkvæm auðkenni, svo hægt sé að tengja þær saman. Þriðja leiðin er að búa til utanálíggjandi venslatöflu þar sem gögnin eru opnuð og vensluð í gegnum venslaskrá. Fjórtða leiðin er að búa til svokallaðan „resolver“ sem væri frístandandi SQL-grunnur og ótengdur öllum öðrum gagnagrunnum. Hann myndi innihalda öll auðkenni úr öllum gagnagrunnunum og venslatöflu myndu tengja auðkennin saman. Einungis auðkennin eru sótt og önnur gögn í gagnagrunnunum eru ekki snert.

Lokaorð

Niðurstöður rannsóknarinnar nýtast til að fá yfirsýn yfir helstu viðfangsefni sem þarf að takast á við svo hægt sé að opna og tengja saman gögn. Rannsóknin er mikilvægt skref í að bæta aðgengi notenda að upplýsingum í Leitir.is. Það er alveg ljóst að við þurfum í náinni framtíð að hrinda af stað þróunarvinnu til að bæta leitarheimtum í Leitir.is.

Heimildir

Europeana. (e.d.). *Linked open data*. <https://pro.europeana.eu/page/linked-open-data>

Risan, A. (2020, 27. októberA). *Masterdata og autoriteter*. <https://nva.sikt.no/registration/0187b7d01d46-de66831d-c031-4b86-bcfo-502f45e4e222>

Risan, A. (2020, 27. októberB). *Metadata - orden i kaos*. <https://nva.sikt.no/registration/0187b7c91e55-85f9ebba-272a-4ae4-ab79-358c218e0fa8>

Sigrún Hauksdóttir. (2005). Nafnmyndaskrá Gegnis. *Fregnir* 30 (3), 46-47.

Smith-Yoshimura, K. (2020, 16. apríl). Presenting metadata from different sources in discovery layers. *Hanging Together – the OCLC Research blog*. <https://hangingtogether.org/presenting-metadata-from-different-sources-in-discovery-layers/>

W3C. (2013, 27. júní). *Linked Data Glossary*. <https://dvcs.w3.org/hg/gld/raw-file/default/glossary/index.html>

W3C. (2015). *Data – W3C*. <https://www.w3.org/standards/semanticweb/data>