

# FLÉTTUÐ FORRITUN OG KVIK SKJALAGERÐ

Frá upphafi tölvualdar hefur flestum forritum fylgt einhver skýringartexti. Hann getur verið á margvíslegu formi. Í sumum tilfellum er markhópurinn notendur, þar sem styttri útskýringar birtast sjálfkrafa á skjánum við notkun forrithins, en lengri útskýringar má finna á sérstökum hjálparsiðum og í almennri notendahandbók. Í öðrum tilfellum er markhópurinn forritarar sem eiga að skilja nákvæmlega virkni forrithins, t.d. til að viðhalda því eða tengjast á einhvern hátt. Einnig eru dæmi um handbækur sem sýna almenna notkun forriths ásamt því að lýsa klösum og algrímum í smáatriðum, enda eru ekki alltaf skýr skil milli notenda og forritara.

Höfundar skrá oft útskýringar á formi athugasemda í kóða, bæði fyrir sjálfa sig og aðra forritara. Á seinni árum hefur færst í vöxt að athugasemdir séu skráðar á ákveðnu ívafsformi sem nýtist til að framleiða á sjálfvirkan hátt hjálparsiður og umfangsmikil uppflettirit. Dæmi um slík ívafsform eru Doxygen og Javadoc, en mörg fleiri eru til.

## FLÉTTUÐ FORRITUN AÐ HÆTTI KNUTH — FLÉTTUÐ GAGNAGREINING Í SWEAVE

Fyrir þremur áratugum kynnti tölvunarfræðingurinn Donald Knuth til sögunnar nýja aðferð og tól til að skrifa og útskýra forrit. Aðferðin sem hann nefndi literate programming (Knuth 1984, 1992) gengur út á að forrit megi skrifa á svipaðan hátt og kennslubók eða tímaritsgrein í tölvunarfræði. Líkt og í bók skiptast á texti, kóði, myndir og töflur.

Með aðferð Knuth eru útskýringar settar fram í þeirri röð sem eðlilegast væri að segja frá þeim, t.d. í fyrirlestri. Þessi röð getur verið mjög ólík þeirri röð sem tölvan les inn kóðann og framkvæmir forritið. Höfundur fléttar saman kóða og texta til skiptis í sömu skrá og getur með utanaðkomandi skipun kallað fram annars vegar þýðanlegan tölvukóða og hins vegar vel uppsetta handbók. Allur kóði forrithins er hafður í fléttuðu skránni, þó aðeins valdir hlutar séu birtir í handbókinni. Forritin TeX og LaTeX eru skrifuð og skjölud á þennan hátt, en tiltölulega fá önnur dæmi eru um að forritarar noti nákvæmlega þessa aðferð til að skrifa stór forrit. Einfalt tól fyrir fléttuða forritun er noweb (Ramsey 1994) sem er ákveðinn grunnur sem önnur töl byggja á og bera sig saman við.

Í vísindum og greiningarvinnu er stuðst við tölfræði til að draga ályktanir. Hið vinsæla tölfræðiforrit R styður fléttuða forritun með aðferð sem nefnist Sweave (Leisch 2002, Leisch og fél. 2011). Tæknileg útfærsla Sweave byggir á aðferð Knuth til að skrifa kóða og útskýringar á víxl í sama skjali, en hugmyndafræðin er talsvert breiðari og lausari í sér. Umfjöllunarefni Sweave skjals er ekki endilega lýsing á ákveðnu forriti, heldur er oft um að ræða tölfræðigreiningu á gögnum. Fléttuð gagnagreining (literate data analysis) af þessu tagi er ein gerð af kvikri skjalagerð (dynamic document generation), þar sem tölur, myndir og töflur eru sóttar á sjálfvirkan hátt í

undirliggjandi gagnaskrár/gagnagrunn og kallað er á R til að framkvæma útreikningana. Mörg önnur töl en Sweave eru fánleg fyrir fléttuða gagnagreiningu (Gandrud 2013, Xie 2013, Stodden og fél. 2014) sem styðja fjölbreytt inntaks- og úttaksform.

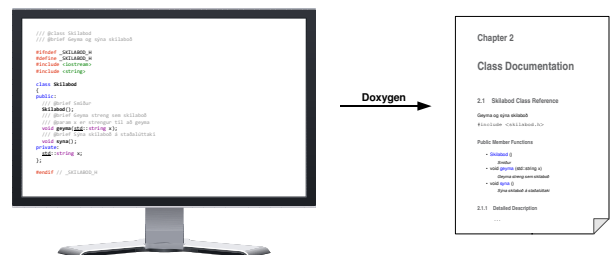
## SAMANBURÐUR Á AÐFERÐUM

Munurinn á aðferðum við fléttuða forritun og kvika skjalagerð (tafla 1) snýr einkum að því hvort allur kóði forrithins sé hafður með, þ.e. hvort hægt sé að þýða allt forritið úr fléttuðu skránni. Athugasemdir á ívafsformi eins og Doxygen teljast strangt til tekið ekki vera fléttuð forritun nema úttakið líkist hefðbundnu kennslubókarformi.

Tafla 1. Aðferðir og algeng töl fyrir fléttuða forritun og kvika skjalagerð.

Aðferð	Inntak	Töl	Úttak
Fléttuð forrit	Kóði með fléttuðum útskýringum	noweb	Forrit + handbók
Forrit með ívafi	Kóði með athugasemdum á ívafsformi	Doxygen	Forrit + hjálparsiður / uppflettirit
Kvik handbók	Skjal með fléttuðum sýnidæmum	Sweave	Handbók
Kvik skýrsla	Skjal með fléttuðri greiningu	Sweave	Skýrsla / vísindagrein

Algengara er að úttak slíks ívafs sé flókinn vefur HTML hjálparsiðna og/ eða þykk PDF samsuða, í báðum tilfellum ætlað til uppflettingar frekar en að lesa í kaflaröð. Textinn ber þá oft með sér að vera að miklu leyti sjálfvirk úttak í bland við texta sem höfundur forrithins hefur skrifað (mynd 1).



Mynd 1. Forritun með ívafi er ekki fléttuð forritun í strangasta skilningi (sjá texta) en er náskyld aðferð, mikið notuð í stærri forritunarverkefnum. Athugasemdir sem hefjast á /// @ eru á Doxygen formi.



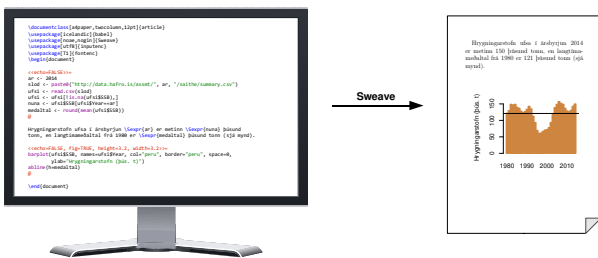
Arni Magnússon og Sigurður Þór Jónsson, sérfræðingar á Hafrannsóknastofnun

Í kvikum handbókum og skýrslum er hins vegar skjalið í forgrunni. Þar gegnir fléttaður kóði því hlutverki að kalla fram sértækt úttak, t.d. mynd, töflu eða sýnidæmi, en er ekki heilsteypt forrit ætlað til almennrar notkunar. Fléttaði kóðinn tryggir að inntakið gefi nákvæmlega það úttak sem sýnt er í skjalinu. Þetta getur verið afar gagnlegur eiginleiki fyrir handbækur, skýrslur og vísindagreinar.

## ÁVINNINGUR

Í fléttaðri forritun skrifar höfundurinn kóða og skýringar á sama tíma og í sömu skrá. Þetta eykur líkur á skýringar verði yfirleitt skráðar niður, auk þess sem vel þekkt er meðal forritara að oft finnst óvænt betri lausn þegar hugsunin er færð í orð. Með því að hafa kóða og textaskýringar í sömu skrá minnka líkur á misræmi milli forrits og handbókar. Kvik handbók veitir ekki aðeins gegnsæ sýnidæmi fyrir lesandann, heldur auðveldar hún viðhald á forritinu. Sýnidæmin eru þannig sjálfvirkar prófanir á núverandi virkni forritsins, þar sem undirliggjandi pakkar taka oft breytingum. Handbókin veitir höfundi sjálfkrafa aðhald við að lýsa breytingum sem hafa orðið á forritinu.

Með kvikli skjalagerð er hægt að flýta verulega fyrir við að útbúa skýrslur sem líkjast fyrri skýrslum, t.d. gerðar á öðrum tímamarki eða með annað viðfang. Allar breytingar verða auðveldari, því þegar gögnum er breytt eða mismunandi reikninalganir prófaðar þá uppfærast tölulegar niðurstöður, myndir og töflur sjálfkrafa (mynd 2). Aðferðafræðin styður líka við stofnanaminni: starfsfólk kemur og fer, en nýr starfskraftur getur tekið upp kvika skýrslu og séð nákvæmlega skref fyrir skref hvernig ákveðin greining er framkvæmd.



Mynd 2. Fléttað gagnagreining (t.d. með Sweave) uppfærir allt skjalið sjálfkrafa þegar undirliggjandi gögnum eða reikniáðferðum er breytt. Með því að gera fléttaðu skrána (vinstra megin) aðgengilega á vefnum getur hver sem er endurtekið greininguna og skilið til hlítar. Táknin <<>= og @ afmarka R kóða sem sækir gögn, reiknar og teiknar.

Á allra síðustu árum eru ritstjórar virtustu vísindatímarita farnir að hvetja höfunda til að senda inn greinar á kviku formi. Þannig fá aðrir vísindamenn í hendur keyranlegt skjal sem gefur sömu niðurstöður og birtar eru í tímaritsgreininni, eða sambærilegt sýnidæmi. Öldum saman hefur endurtakanleiki verið lykilaðri í rannsóknum sem teljast vísindalegar, en á síðustu áratugum hefur molnað verulega undan þessum hornsteini framfara. Með auknu umfangi rannsókna og flóknari greiningaraðferðum er nú oft útilokað fyrir lesanda að endurtaka rannsóknina og fá sömu niðurstöður. Líta má á slíkar greinar sem auglýsingu um nýja aðferð eða uppgötvun, en kvik skjöl eru heppilegur miðill til að lýsa greiningaraðferðum það nákvæmlega að þær séu endurtakanlegar.

Það er við hæfi að lýsa þessari tækniþróun og vinnulagi í Tímariti Skýrslutækniátlags Íslands. Kvik skjalagerð með fléttaðri gagnagreiningu er nútímaleg og opin skýrslutækni sem býður upp á marga kosti og fléttað forritun er grunnurinn að þessari nálgun.

## HEIMILDIR

Gandrud, C. 2013. Reproducible Research with R and RStudio. Boca Raton: CRC.

Knuth, D.E. 1984. Literate programming. The Computer Journal 27:97–111.

Knuth, D.E. 1992. Literate Programming. Stanford: Center for the Study of Language and Information.

Leisch, F. 2002. Sweave: Dynamic generation of statistical reports using literate data analysis. Í: W. Härdle & B. Rönz (ritstj.) Compstat 2002: Proceedings in Computational Statistics, Heidelberg, Physica, bls. 575–580.

Leisch, F., M. Eugster & T. Hothorn. 2011. Executable papers for the R community: The R2 platform for reproducible research. Procedia Computer Science 4:618–626.

Ramsey, N. 1994. Literate programming simplified. IEEE Software 11(5):97–105.

Stodden, V., F. Leisch & R.D. Peng. 2014. Implementing Reproducible Research. Boca Raton: CRC.

Xie, Y. 2013. Dynamic Documents with R and knitr. Boca Raton: CRC.

## **Literate programming and dynamic generation of reports**

Literate programming and literate data analysis are efficient ways to produce documented computer programs and dynamic reports. This paper examines the gradient between Knuth's literate programming (generating a program and handbook), source code with markup (program and API reference), dynamic handbook (user manual with tested examples), and literate data analysis (document with underlying code). Today, scientific journal editors are encouraging authors to submit manuscripts supplemented with literate data analysis, in order to make scientific analysis reproducible.