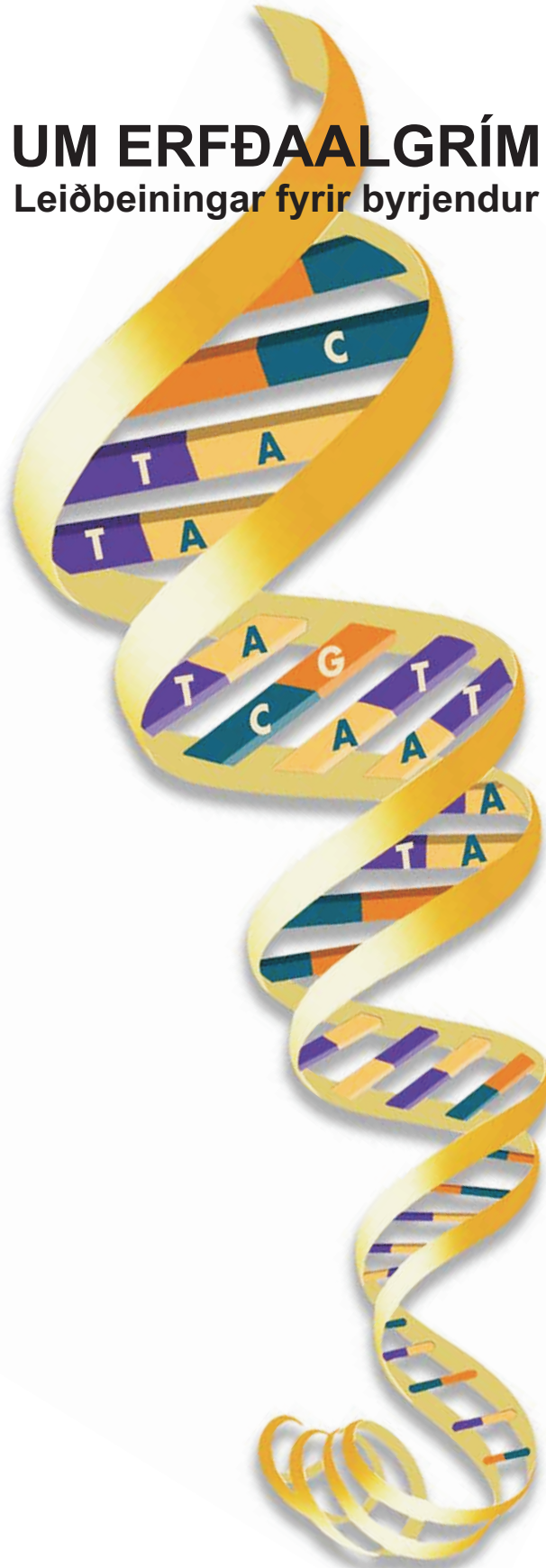


Freysteinn Alfreðsson

UM ERFÐAALGRÍM

Leiðbeiningar fyrir byrjendur



Um erfðaalgrím

Leiðbeiningar fyrir byrjendur

Markmiðið með þessum leiðbeiningum er að kynna erfðaalgrím fyrir byrjendum, með því að nota eins einföld dæmi og kostur er. Útskýringarnar gætu hentað bæði þeim sem hafa áhuga á að kynna sér hvernig hægt er að láta tölvu taka ákvörðun og þeim sem hafa undirstöðu í forritun og vilja taka fyrstu skrefin í átt að gervigreind.

Erfðaalgrím henta vel byrjendum í gervigreind þar sem þau eru bæði einföld og fljótleg í útfærslu. Sérhannað forrit hefur verið búið til fyrir þetta kennsluefni fyrir þá sem vilja sjá hvernig erfðaalgrímið virkar. Forritið heitir GAHello og er skrifað á Java forritunarmálinu. GAHello er aðgengilegt á <http://svn.isir.is/svn/ga>.

Erfðaalgrím er aðferðafræði til að þróa lausnir á vandamálum með erfðum. Þessu má líkja við DNA í lífverum þar sem afkvæmin erfa einkenni frá báðum foreldrum sínum. Í erfðaalgríminu má líkja lausnunum við DNA þar sem afkvæmið gefur til kynna nýja lausn á vandamálinu sem inniheldur hluta af báðum lausnum foreldranna. Úr þessum lausnum eru þær bestu valdar til æxlunar í þeirri von að afkvæmin gefi hæfari lausnir sem að lokum eru nógu þróaðar til þess að leysa vandamálið að fullu. Þetta er einnig þekkt í lífríkinu þar sem hæfustu dýrin eru líklegri til þess að afla sér matar en þau óhæfu og eru því líklegri til þess að geta af sér afkvæmi. Þetta ferli þekkjum við einnig í kenningum Darwins þar sem þeir hæfustu lifa af.

Erfðaalgrímið er eitt vinsælasta algrímið úr EA (*Evolutionary Algorithm*) aðferðafræðinni þar sem það er bæði einfalt, skemmtilegt og fljótlegt í útfærslu. Það hefur verið notað víða í margvíslegum tilgangi með góðum árangri, þar á meðal í leit að útvarpsútsendingum, ákvarðanatöku í tölvuleikjum og á fjármálamarkaðinum, svo að eitthvað sé nefnt.

Erfðaalgrímið nýtist þar sem til eru fleiri en ein lausn á sama vandamálinu en á betur við þegar lausnir á vandamálinu eru mjög margar.

Einfalt dæmi um vandamál er ef erfðaalgríminu væri ætlað að finna matvöruverslun. Í því tilfelli hefði erfðaalgrímið ótal verslanir til að velja úr, ásamt mörgum leiðum til þess að komast til þeirra. Að lokum gæti erfðaalgrímið fundið lausn sem bæði innihéldi eina verslun með matvörur og leið til þess að komast þangað.

Þar sem erfðaalgrímið velur að lokum aðeins eina lausn af mörgum mögulegum lausnum segja sumir að erfðaalgrím séu eins og slembitölur með aukinni nákvæmni.

Þar sem erfðaalgrímið er nátengt erfðafræðinni byrjum við á að kynna nokkur hugtök sem þau eiga sameiginleg. Þessi nöfn verða síðan notuð í framhaldinu á þessu kennsluefni.

Hugtak	Á ensku	Í líffræði	Í erfða algrími
Erfðamengi	Genome	Allt erfðæfni í frumu	Heildarlausn
Gen	Gene	Hver eiginleiki fyrir sig	Eitt brot úr heildarlausn
Erfðamengi stofns	Genepool	Allt erfðæfni í stofni	Allar lausnir
Kynslóð	Generation	Afkvæmi fyrri kynslóðar	Nýjar lausnir myndaðar úr fyrri lausnum
Hæfni	Fitness	Hæfni lífveru í lífríki	Hæfni lausnar til þess að leysa vandamálið
Afkvæmi	Offspring	Afkvæmi lífveru	Nýtt erfðamengi búið til út frá öðrum erfðamengjum

Hugtak	Í líffræði	Í erfða algrími
P	Upphafs kynslóð	Upphafs stofn lausna
F ₁	Önnur kynslóð	Önnur kynslóð lausna, mynduð úr upphafs stofni
F ₂	Þriðja kynslóð	Þriðja kynslóð lausna, mynduð úr annarri kynslóð
F _n	n-ta kynslóð	n-ta kynslóð lausna, mynduð úr n-1 kynslóðinni

*P = Parental, F = Filius (sem þýðir sonur á latínu)

Procedure erfðaalgrímsins;

begin

set kynslóð n := 0;

búa til upphafsstofn P(n);

while hæfasta erfðamengið hefur ekki verið fundið **do**

kanna hæfni erfðamengja í stofni P(n);

velja úrval erfðamengja út frá hæfni;

búa til næstu kynslóð með úrvalinu;

skipta út gömlu kynslóðinni með þeirri nýju;

set kynslóð n := n + 1;

end

end.

Eins og sjá má á þessu ferli er það endurtekið þar til erfðamengi sem getur leyst vandamálið er fundið. Þetta ferli getur þó á þeim tíma búið til þúsundir erfðamengja sem ekki eru nógu góð til þess að fullnægja skilyrðunum.

Einfalt dæmi um þetta væri bíll sem á að leggja í stæði. Til eru ótal leiðir til þess að leggja bílnum í stæðið en lokamarkmiðið er að bíllinn sé staddur í stæðinu og að hann hafi ekki valdið tjóni á leiðinni þangað. Í þessu tilfalli myndi erfðaalgrímið finna ótal óhagkvæm erfðamengi sem ekki myndu uppfylla þessi skilyrði þar sem þau innihalda leiðbeiningar sem kæmu hvorki bílnum á leiðarenda né þangað í heilu lagi.

Eftir að erfðaalgrímið hefur þróað nýjar lausnir með því að nota bestu tillögurnar úr gömlu lausnunum mun það að lokum finna lausn sem hæfir vandamálinu og leggur bílnum í stæðið með þeirri lausn.

Í dæminu sem við munum nota til hliðsjónar í þessum leiðbeiningum mun erfðaalgrímið eingöngu eiga sér eina lausn, okkur til einföldunar.

Í forritun hefur með árunum myndast hefð fyrir því að fyrsta verkefni nemandans sé að búa til forrit sem skrifar textann „Hello, World!“ á skjáinn. Því þótti tilvalið að halda í þá hefð í þessum leiðbeiningum með því að láta dæmið þróa textann „Hello, World!“ með erfðaalgrími.

UPPBYGGING ERFÐAMENGJA

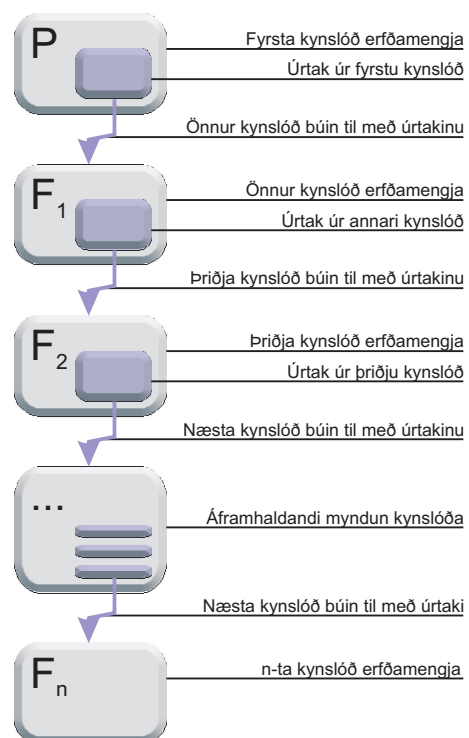
Eitt af lykilatriðum erfðaalgrímsins er erfðamengið sem lýsir lausn vandamálsins. Algengt er að nota einn textastreng eða mengi af tvíundartölum (e. binary) sem erfðamengi. Nauðsynlegt er að sníða erfðamengið að vandamálinu sem það á að leysa.

Þar sem erfðamengið lýsir eingöngu lausn á tilteknu vandamáli þá eru mögulegar uppbyggingar á erfðamengjunum jafnfjölbreytilegar og vandamálin sem þau eru sniðin til að leysa, og þess vegna ekki til neinar staðlaðar uppbyggingar á erfðamengjum sem henta öllum vandamálum. Því er það undir höfundinum komið hvernig hann hannar erfðamengið.

Til eru margar leiðir til þess að láta erfðamengin lýsa vandamálinu. Hér verður farið yfir eina þeirra. En hún felst í því að túlka stafi í textastreng. Hafa verður í huga að hægt er að mynda erfðamengi úr nánast hverju sem er.

Þegar erfðamengið er búið til úr textastreng, þá merkir hver hluti af textastrengnum eitthvað ákveðið sem lýsir bæði aðferð og gerð lausnarinnar.

Í dæminu okkar munum við nota textastreng fyrir erfðamengi sem eru jafnlangur og textinn „Hello, World!“ Óhæf erfðamengi munu því innihalda texta á borð við „eV]npC.xcrnhF“, „Hellpf worjd!“ og „Helmo, world!“.



UPPHAFSSTOFN

Upphafsstofninn er venjulega búinn til með því að mynda ný erfðamengi með slembitölum. Með því að nota slembitölur tryggjum við fjölbreytileika í erfðamengjunum og aukum líkur á að sum erfðamengin innihaldi gen sem ganga upp í lausnina á vandamálinu.

Upphafsstofninn í okkar dæmi inniheldur 3000 erfðamengi sem eru bún til með því að raða saman slembistöfum, sem eru jafnmargir og eru í textastrengnum „Hello, world!“.

HÆFNISÚTREIKNINGAR

Til þess að hægt sé að finna hæfustu erfðamengin verður að vera hægt að bera þau saman. Þetta er gert með því að reikna út hæfni hvers erfðamengis fyrir sig og að bera þá útreikninga saman við heildina. Einnig verður að vera hægt að sjá hvort fullkomin hæfni hafi verið fundin og erfðamengið innihaldi þar með svar við vandamálinu.

Í okkar tilfalli er hæfnin reiknuð út frá því hve langt hver stafur í strengnum er frá réttum staf í stafrófinu. Síðan eru tölurnar lagðar saman. Sú tala segir síðan til um hve hæft hvert erfðamengi er í samanburði við öll hin. Ef hins vegar hæfnistalan er núll þá er lausnin fundin og því óþarfi að erfðaalgrímið haldi áfram.

$$\begin{array}{cccccccccccc} H & e & l & l & o & , & w & o & r & l & d & ! \\ \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow \\ 0 & + & 0 & + & 0 & + & 0 & + & 0 & + & 0 & + & 0 \end{array} = 0 \text{ hæfni}$$

$$\begin{array}{cccccccccccc} D & m & l & j & o & , & w & o & n & l & d & ! \\ \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow \\ 4 & + & 7 & + & 0 & + & 2 & + & 0 & + & 0 & + & 0 & + & 4 & + & 0 & + & 0 & + & 0 \end{array} = 17 \text{ hæfni}$$

$$\begin{array}{cccccccccccc} D & m & l & j & v & , & \$ & w & o & n & l & f & ! \\ \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow \\ 4 & + & 7 & + & 0 & + & 0 & + & 7 & + & 0 & + & 4 & + & 0 & + & 0 & + & 4 & + & 0 & + & 2 & + & 0 \end{array} = 28 \text{ hæfni}$$



ÚRTAK

Úr hæfustu erfðamengjunum er síðan valið úrtak og í flestum tilfellum einnig slembiúrtak. Það er undir höfundinum komið hve stórt úrtakið er en það getur verið misjafnt hver æskileg stærð úrtaksins ætti að vera, en það getur farið eftir því hvernig uppbygging erfðamengjanna er.

Slembiúrtakið er nauðsynlegt þar sem sum erfðamengi innihalda örfá góð gen þrátt fyrir að erfðamengið sjálft sé með lélega hæfni. Með þessu móti eru meiri líkur á að þessi góðu gen erfist í næstu kynslóð og flýttir það fyrir því að mynda fullkomið erfðamengi í komandi kynslóðum.

Úrtak erfðamengjanna í dæminu okkar eru 500 af þeim 3000 sem mynda núverandi kynslóð erfðamengjanna, og eru þau valin út frá hæfnismati þeirra. Lægri hæfnistölur segja til um betri hæfni og eru því erfðamengin með lægstu tölurnar valin í úrtakið.

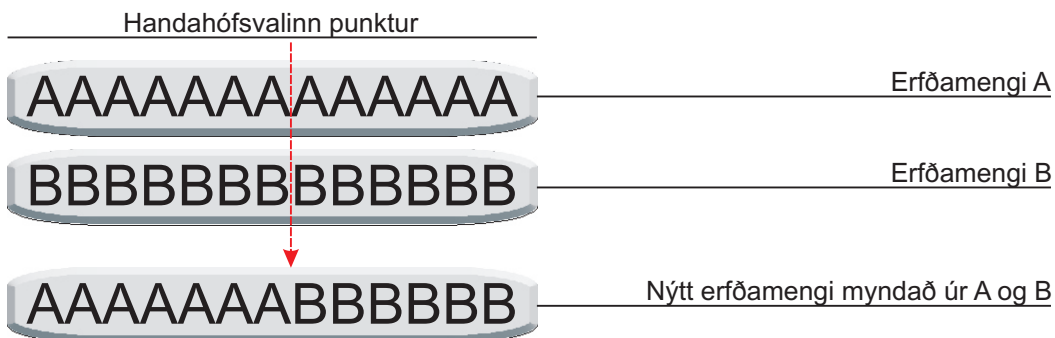
Ásamt þessum 500 eru valin 200 erfðamengi af handahófi úr erfðamengjunum.

ÆXLUN ERFÐAMENGJANNA

Næst þarf að mynda nýja kynslóð af erfðamengjum með því að nota úrtakið.

Til eru nokkrar aðferðir til að mynda næstu kynslóð erfðamengja. Aðferðinar sem henta best fara að mestu eftir því hvernig erfðamengin eru uppbyggð. Algengasta aðferðin er að nota rússneska rúllettu (e. Russian roulette).

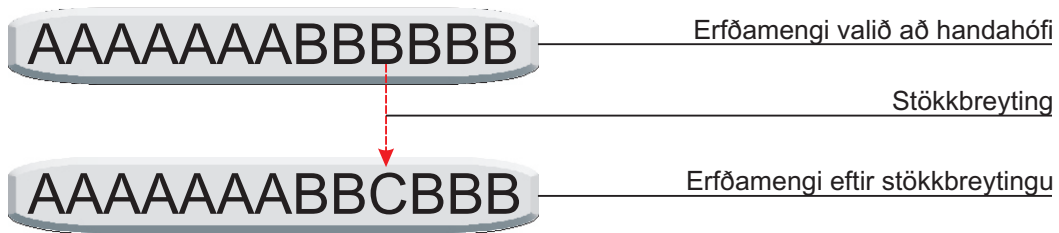
Rússneska rúllettan felst í því að taka af handahófi erfðamengjapör úr úrtakinu, taka þau í sundur á handahófskenndum stað og að lokum mynda nýtt erfðamengi sem inniheldur sinn hlutann af hvoru erfðamenginu. Í sumum tilfellum getur þó reynst æskilegt að skipta erfðamengjunum í fleiri parta.



Í okkar tilfalli er þetta alveg eins. Strengirnir 700, sem mynda erfðamengin, eru settir saman með því að skipta hverjum streng fyrir sig í tvennt. Síðan er nýr strengur búinn til með því að setja hvorn hlutann fyrir sig saman. Þetta ferli er framkvæmt 3000 sinnum og er þá gömlu erfðamengjunum hent og nýju erfðamengin notuð í staðinn.

STÖKKBREYTINGAR

Þar sem hætta er á að ekkert erfðamengjanna innihaldi nauðsynleg gen er það möguleiki að aldrei verði hægt að finna lausn á vandamálinu. Því gegna stökkbreytingar gríðalega mikilvægu hlutverki í erfðaalgrímum þótt í einstaka tilfellum eigi þær ekki við. Stökkbreytingar í erfðaalgríminu eru útfærðar sem tilviljanakenndar breytingar á úrtaki úr erfðamengjunum.



Dæmið okkar nýtir sér stökkbreytingar með því að skipta annað slagið út einum staf í erfðamengjunum 3000 sem mynduð voru úr fyrri kynslóð erfðamengja. Með þessu móti tryggjum við ef okkur vantar einn staf í erfðamengjunum 3000, t.d. stafinn H í „Hello”, að þá mun sá stafur að lokum fást með stökkbreytingum og erfðaalgrímið geti að lokum leyst vandamálið.

HELLO WORLD

Nú þegar öll hugtök erfðaalgrímsins hafa verið útskýrð, leggjum við lokahönd á dæmið okkar.

Eftir að búið er að mynda, fjölga og stökkbreyta erfðamengjunum þá er ferlið endurtekið þar til það finnur textastrenginn „Hello, World!”.

Forritið skrifar út hæfasta erfðamengið í hverri kynslóð, ásamt því hve hæft það er. Að lokum þegar það hefur myndað erfðamengið „Hello, World!”, hættir það vinnslu þar sem hámarks hæfni hefur verið náð.

Hvaða efni er gott að skoða í framhaldi af þessu?

Eins og sjá má er erfðaalgrímið afar einfalt og hægt að nýta það við ótal aðstæður, svo lengi sem bæði vandamálið og lausnaraðferðirnar eru þekktar.

Í framhaldinu getur verið sniðugt fyrir þá sem áhuga hafa á að útfæra erfðaalgrímið að reyna leysa 8 drottninga vandamálið. Það felst í því að raða 8 drottningum á taflborð þannig að engin þeirra ógni annarri. Og þá gæti verið hentugt að hafa upplýsingar um staðsetningu drottninganna í erfðamengjunum.