

R programski jezik

Seminarski rad u okviru kursa
Dizajn programskih jezika
Matematički fakultet

Mihajlo Stojanović

Decembar 2019

Sažetak

U daljem tekstu prikazuju se osnove razvoja, glavne karakteristike i osnovna namena programskog jezika R. Opisani su programski jezici S i Scheme koji su najviše uticali na njegov razvoj, kao i programski jezik Julia koji je nastao pod njegovim uticajem. Za svaki od navedenih programskih jezika date su osnovne informacije, kao i razvojno stablo sa njihovim godinama nastajanja.

Sadržaj

1	Uvod	2
2	Osnovno o programskom jeziku R	2
3	Razvojno stablo	3
3.1	Programski jezik S	3
3.2	Lisp	4
3.3	Common Lisp(CL)	4
3.4	Scheme	5
3.5	Julia	5
4	Primer koda u programskom jeziku R	5
5	Zaključak	6
	Literatura	6

1 Uvod

Statistika i grafička obrada podataka imaju ključnu ulogu u različitim poslovnim oblastima. Od procesa donošenja važnih poslovnih odluka, definisanja marketinških ciljeva, naučno-istraživačkih radova, do funkcionisanja državnih institucija i privrede. Upravo programski jezik R koji je nastao 1993. godine je jezik za statističke proračune, analizu podataka i vizualizaciju. Iako je donedavno bilo uvreženo mišljenje da R koristi uglavnom akademska zajednica, dok se za analize podataka u firmama koriste neki drugi alati, danas je situacija znatno drugačija. O tome svedoči i spisak kompanija koje koriste R za analizu i prezentaciju svojih podataka[2]. U nastavku je na slici 1 prikazan logo jezika R i date su detaljnije informacije o njemu.



Slika 1: Logo programskog jezika R

2 Osnovno o programskom jeziku R

R je programski jezik i programsko okruženje za statističke proračune i grafike koji se prvi put pojavio 1993. godine[10]. R su napravili Ros Ihaka(engl. *Ross Ihaka*)¹ i Robert Dzentlmen(engl. *Robert Gentleman*)² na Aukland Univerzitetu(engl. *University of Auckland*), Novi Zeland. Jezik je postao standard među statističarima koji razvijaju statični softver i široko je korišćen za analizu podataka. R obezbedjuje širok izbor statičkih(linearnih i nelinearnih modela, klasičnih statističkih tekstova, analiza vremenskih serija, klasifikacija, klastera i ostalo) i grafičkih tehnika. Slično kao i S programski jezik, R je projektovan kao pravi programski jezik i omogućava korisnicima dodatnu funkcionalnost definisanjem novih funkcija. R može biti proširen, kroz pakete obezbedjene od korisnika, za specifične funkcije ili oblasti proučavanja. Takođe, ima bolju podršku za objektno-orijentisano programiranje nego ostali statistički programski jezici. Jos jedna od prednosti R su njegove grafičke mogućnosti koje uključuju matematičke simbole.

¹Ros Ihaka je statističar iz Novog Zelanda.

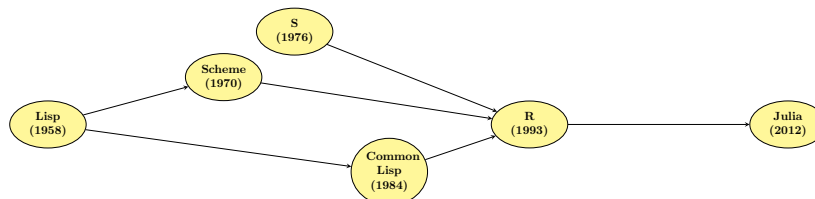
²Robert Dzentlmen je statističar i bioinformatičar iz Kanade.

Može biti korišćen kao opšti alat za izračunavanje matrica, poput konkurentskih programa Octave i Matlab. R ima sopstveni LaTeX-oliki format dokumenta, koji se koristi za predstavljanje sveobuhvatne dokumentacije, preko interneta ili kao štampana verzija.

Programski jezik R je implementacija programskog jezika S[12] kombinovanog sa semantikom koja je inspirisana programskim jezikom Scheme[11]. Nazvan je delimično prema imenima autora, a delom kao igra reči na ime S. Prva verzija izašla je 1995. godine, dok je stabilna verzija izašla 2000. godine[5]. R je deo GNU projekta³. Njegov izvorni kod je slobodan i pod uslovima koje daje GNU-ova opšta javna licenca, a prekompilirane binarne verzije su obezbedjene za različite operativne sisteme. R koristi interfejs komandne linije, kroz više grafičkih korisničkih okruženja. Razvojem R programskog jezika se danas bavi R Development Core Team.

3 Razvojno stablo

Na razvoj jezika R su najviše uticali programski jezici **S** i **Scheme**. Jezik R predstavlja implementaciju jezika S. Postoji veliki broj promena u odnosu na S, ali ipak veliki deo S koda radi i u R. Programski jezik R koristi semantiku jezika Scheme koji predstavlja jedan od tri glavna dijalekta jezika **Lisp**, zajedno sa jezikom **Common Lisp** koji takodje ima uticaja na R i jezikom Clojure. Takodje, R je jedan od jezika koji su imali uticaj na razvoj programskog jezika **Julia**. Razvojno stablo jezika R se može videti na slici 2.



Slika 2: Razvojno stablo jezika R

3.1 Programski jezik S

S je statistički programski jezik koji je prvenstveno razvio Dzon Čembers(engl. *John Chambers*)⁴, Rik Beker(engl. *Rick Becker*) i Alan Vilks(engl. *Allan Wilks*) u Belovim laboratorijama 1976. godine. Po rečima Dzona Čembersa cilj jezika je bio "pretvoriti ideje u softver brzo i pouzdano"[3].

³GNU je operativni sistem sastavljen isključivo od slobodnog softvera.

⁴Dzon Čembers je član Američke Asocijacije statističara i instituta za Matematičku statistiku.

S je programski jezik od starta namenjen za programiranje sa podacima, odnosno organizaciju, vizualizaciju i analizu podataka. Sve do nastanka jezika S većina statističkih izračunavanja je radjena direktno preko jezika Fortran⁵. Medjutim, S je dizajniran tako da ponudi više interaktivniji pristup. Ovakav pristup uključuje grafičke uređaje (štampače i povremeno terminale karaktera) i obezbedjuje lako dostupnu dokumentaciju za funkcije.

Prva verzija napravljena 1976. godine radila je na GCOS (engl. *General Comprehensive Operating System*) operativnim sistemima. Ime je izabrano kao zajedničko slovo korišćeno u statističkim izračunavanjima, u skladu sa drugim programskim jezicima koji su dizajnirani u Belovoj laboratoriji u to vreme (poput C programskog jezika).

Velike promene su napravljene 1988. godine, pre svega u samoj sintaksi S jezika.[1] Uveden je prelazak sa makroa na funkcije, i omogućeno je prosledjivanje funkcija drugim funkcijama (poput apply). Proširen je sam koncept objekata i sintaksa je konzistentnija. Omogućena je formula notacija, rad sa data frame objektima i modifikovana je upotreba metoda i klasa. Poslednja verzija S jezika je standard S4 koji obezbedjuje napredna objektno-orijentisana svojstva. Ovakav oblik S jezika je veoma sličan modernim verzijama jezika R, zbog čega se i kaže da je R implementacija jezika S.

3.2 Lisp

Lisp je programski jezik zasnovan na matematičkoj teoriji rekurzivnih funkcija, a Lisp program je funkcija koja se primenjuje na podatke[4]. Džon Mekarti (eng. *John McCarthy*) je razvio Lisp 1958. godine dok je bio na MIT (eng. *Massachusetts Institute of Technology*). Ime LISP je nastalo od „**LI**St **P**rocessor”, a povezane liste su jedan od glavnih tipova podataka. Osnova Lisp-a je funkcionalno programiranje, ali se Lisp zbog raznih drugih svojstava smatra multiparadigmatskim programskim jezikom. Takodje, Lisp je jedan od prvih jezika koji je imao sakupljač otpadaka (eng. *Garbage Collector*). Danas najpoznatiji dijalekti Lispa su Clojure⁶, kao i Common Lisp i Scheme koji su predstavljeni u nastavku.

3.3 Common Lisp (CL)

CL je Lisp dijalekt, razvijen 1994. godine kao standardizovan i poboljšan naslednik MacLispa (dijalekt jezika Lisp koji je nastao 1966. godine.). Common Lisp je multiparadigmatski jezik opšte namene. Podržava kombinaciju imperativne, funkcijske i objektno-orijentisane programske paradigme. CL je dinamički programski jezik koji olakšava rapidan razvoj, sa iterativnom kompilacijom u učinkovite run-time programe[8].

⁵Fortran (engl. **FOR**mula **TRAN**slation) je kompajliran i imperativan programski jezik opšte namene, specijalno dizajniran za numeričku analizu i statistička izračunavanja. Napravljen je 1957. godine od strane Džona Bakusa (eng. *John Backus*) i kompanije IBM (eng. *International Business Machines Corporation*).

⁶Clojure je programski jezik opšte namene koji je stvorio Ric Hikej (eng. *Rich Hickey*) 2007. godine. Naglasak je na funkcionalnom programiranju i radi na Java platformama. Kao i drugi dijalekti Lispa, Clojure tretira kod kao podatke i ima makro sistem.

3.4 Scheme

Scheme je multiparadigmatski programski jezik opšte namene, koji podržava funkcionalno i imperativno programiranje[9]. Kreirali su ga Derald Susman(eng. *Gerald J. Sussman*) i Gaj Stil(eng. *Guy L. Steele*) 1975. godine u MIT-ovim laboratorijama za računarska istraživanja i veštačku inteligenciju. Nastao je pod uticajem Algola⁷ i funkcionalnog Lisp programskog jezika. Prvobitno je bio namenjen za istraživanja i podučavanja studenata.

Scheme je prvi jezik iz Lisp familije koji je uveo statički doseg identifikatora(eng. *static (lexical) scope*) i koji je zahtevao samu implementaciju kako bi se izvela optimizacija repne rekurzije, dajući time veću podršku funkcionalnom programiranju kao i drugim tehnikama koje koriste rekurziju. Implementacija repne rekurzije se obavlja pomoću naredbi skoka, pa se time izbegava nepotrebna upotreba memorijskog prostora. Memorijski prostor za podatke se alokira dinamički i automatski se dealocira pomocu sakupljača otpada(eng. *garbage collector*). Scheme je jako tipiziran programski jezik, čiji se parametri prenose po vrednosti dok se funkcije tretiraju jednako sa ostalim tipovima podataka[6]. Smatra se da je Scheme imao veliki uticaj na nastanak i razvoj Common Lisp jezika, drugog dijalekta jezika Lisp.

3.5 Julia

Programski jezik Julia su razvili Džef Bezanson (eng. *Jeff Bezanson*), Stefan Karpinski (eng. *Stefan Karpinski*), Viral Šah (eng. *Viral B. Shah*) i Alan Edelman (eng. *Alan Stuart Edelman*) 2012. godine. Julia je multiparadigmatski programski jezik koji kombinuje karakteristike imperativnog, funkcionalnog i objektno-orijentisanog programiranja[7]. Dizajniran je da odgovori na zahteve razvijenih numeričkih i naučnih izračunavanja poput programskog jezika R, ali je takodje pogodan za opšte programiranje. Da bi se ovo postiglo jezik Julia je napravljen na osnovu matematičkih programskih jezika, pozajmljujući dosta karakteristika od dinamičkih jezika poput Lispa(Common Lisp i Scheme).

4 Primer koda u programskom jeziku R

```
1000 year = as.integer(readline(prompt="Enter a year: "))
1001 if((year %% 4) == 0) {
1002     if((year %% 100) == 0) {
1003         if((year %% 400) == 0){ print(paste(year, "is a leap year")) }
1004         else { print(paste(year, "is not a leap year")) }
1005     } else { print(paste(year, "is a leap year")) }
1006 } else{ print(paste(year, "is not a leap year")) }
```

Listing 1: R kod koji proverava da li je godina prestupna

⁷Algol(eng. *ALGO*rithmic *L*anguage) je imperativan, proceduralni i struktuiran programski jezik nastao 1958. godine.

5 Zaključak

U ovom tekstu ukratko su prikazane osnovne karakteristike i razvoj programskog jezika R. Prikazani su jezici za koje se smatra da su najviše uticali na njegov nastanak, kao i jezik na koji je R najviše uticao. Prikazano je njegovo osnovno razvojno stablo. Za detaljnije informacije preporučuje se korišćenje zvanične stranice programskog jezika R[10].

Literatura

- [1] J.M.; Wilks A.R. Becker, R.A.; Chambers. *The New S Language: A Programming Environment for Data Analysis and Graphics*. Wadsworth Brooks/Cole, 1988.
- [2] Deepanshu Bhalla. Companies using r. *Listen data*, January 2016.
- [3] John M Chambers. *Programming with Data: A Guide to the S Language*. Springer, 1998.
- [4] David Chisnall. Influential programming languages, part 4: Lisp. *informIT*, January 2011.
- [5] Gary Ernest Davis. Over 16 years of r project history. *Revolution analytics*, March 2016.
- [6] R. Kent Dybvig. *The Scheme Programming Language, Fourth Edition*. The MIT Press., 2009.
- [7] Julia. Zvanična stranica programskog jezika julia. <https://julialang.org/>.
- [8] Common Lisp. Zvanična stranica programskog jezika Common Lisp. <https://common-lisp.net/>.
- [9] MIT. The scheme programming language. October 2003.
- [10] R. Zvanična stranica programskog jezika R. <https://www.r-project.org/about.html>.
- [11] Scheme. Zvanična stranica programskog jezika Scheme. <https://schemers.org/>.
- [12] W. N. Venables and B. D. Ripley. *S Programming*. Springer, 2000.