دانشگاه اصفهان

دانشکده مهندسی کامپیوتر

گزارش پروژه

درس زبان‌های برنامه نویسی

بررسی زبان R

سیدمحمدحسین هاشمی
بردیا جوادی

استاد : دکتر آرش شفیعی

پاییز ۱۴۰۳


بسم الله الرحمن الرحيم

فهرست مطالب

فهرست جداول

فهرست کدها

فصل ۱

مقدمه

۱.۱ تاریخچه زبان R

زبان برنامه‌نویسی R در ابتدا با هدف ساده‌سازی تحلیل‌های آماری و مدل‌سازی داده‌ها طراحی شد. این زبان که توسط راس ایهاکا و رابرت جنتلمن در دهه ۱۹۹۰ ابداع شد، به عنوان یک ابزار متن‌باز برای تحلیل داده‌های پیچیده توسعه یافت. در آن زمان، نیاز به زبانی که توانایی تحلیل آماری پیشرفته، مدل‌سازی ریاضی و تولید گراف‌های بصری را داشته باشد، به شدت احساس می‌شد. زبان R با الهام از زبان SAS طراحی شد و توانست مشکلاتی نظیر عدم انعطاف‌پذیری ابزارهای آماری موجود و محدودیت‌های گرافیکی آنها را برطرف کند. همچنین، متن‌باز بودن R موجب شد که جامعه‌ای پویا از کاربران و توسعه‌دهندگان حول آن شکل بگیرد، که این امر به گسترش سریع امکانات و کتابخانه‌های آن کمک کرد.

۲.۱ ویژگی‌های زبان R

R به دلیل قابلیت‌های منحصربه‌فرد خود در حوزه‌های مختلف کاربرد گسترده‌ای دارد. از جمله مهم‌ترین حوزه‌ها می‌توان به موارد زیر اشاره کرد:

- تحلیل داده‌ها و آمار پیشرفته: به‌طور خاص برای تحلیل‌های آماری پیچیده و مدل‌سازی داده‌ها طراحی شده است.
- یادگیری ماشین و هوش مصنوعی: بسیاری از الگوریتم‌های یادگیری ماشین و ابزارهای هوش مصنوعی در پیاده‌سازی شده‌اند.
- مصورسازی داده‌ها: قابلیت‌های گرافیکی پیشرفته آن را به ابزاری مناسب برای ایجاد نمودارهای حرفه‌ای و گزارش‌های بصری تبدیل کرده است.
- بیوانفورماتیک: در تحلیل داده‌های زیستی، از جمله داده‌های ژنومی و پروتئومی، کاربرد فراوان دارد.
- امور مالی و اقتصادی: بسیاری از مدل‌های مالی و پیش‌بینی‌های اقتصادی با استفاده از R پیاده‌سازی می‌شوند.

- تحقیقات علمی: به‌عنوان یک ابزار تحقیقاتی در علوم اجتماعی، روانشناسی و بسیاری از رشته‌های علمی مورد استفاده قرار می‌گیرد. علاوه بر این، R به دلیل پشتیبانی گسترده از کتابخانه‌های تخصصی، امکان تحلیل‌های پیشرفته در حوزه‌های خاص را فراهم می‌کند.
- بازاریابی و تحلیل مشتری: کسب‌وکارها از برای تحلیل رفتار مشتری، تقسیم‌بندی بازار و پیش‌بینی فروش استفاده می‌کنند.

۳.۱ هدف از ایجاد زبان و مشکلاتی که رفع کرد

R برای حل مشکلات خاصی طراحی شد که در آن زمان در زبان‌های موجود دیده می‌شد:

- رایگان و منبع‌باز بودن: در دهه ۹۰، بسیاری از ابزارهای آماری مثل SAS و MATLAB هزینه‌های بالایی داشتند. R به عنوان یک جایگزین رایگان و منبع‌باز برای این ابزارها توسعه یافت.
- سازگاری با زبان SAS: کاربران زبان SAS می‌توانستند کدهای خود را با تغییرات جزئی در R اجرا کنند. این ویژگی، پذیرش R را تسریع کرد.
- انعطاف‌پذیری بالا: R به طور خاص برای تحلیل آماری و مصورسازی داده‌ها بهینه‌سازی شده بود و نیازهای تحلیل‌گران را بهتر از زبان‌های عمومی مثل C یا Java برآورده می‌کرد.
- قابلیت توسعه‌پذیری: با ایجاد کتابخانه‌ها و بسته‌های متنوع، کاربران می‌توانستند قابلیت‌های R را برای نیازهای خاص خود توسعه دهند.

۴.۱ مشکلاتی که در ابتدای ظهور خود رفع می‌کرد

- عدم دسترسی به ابزارهای پیشرفته تحلیل داده: با ارائه ابزارهایی قدرتمند، نیاز به نرم‌افزارهای گران‌قیمت را کاهش داد.
- مصورسازی ضعیف داده‌ها: در آن زمان، بسیاری از زبان‌های برنامه‌نویسی، ابزارهای گرافیکی مناسبی نداشتند. R با کتابخانه‌هایی مثل ggplot2، انقلابی در مصورسازی داده‌ها ایجاد کرد.
- پیچیدگی کدنویسی در زبان‌های عمومی: تحلیل‌گران داده معمولاً نیاز داشتند تا با زبان‌هایی مثل Fortran یا C کار کنند که برای تحلیل داده بهینه نبودند. R این مشکل را رفع کرد و کار با داده‌ها را ساده‌تر نمود.

۵.۱ بهبود نسبت به زبان‌های دیگر

R بسیاری از قابلیت‌ها و ویژگی‌های موجود در زبان‌های دیگر را ارتقا داد و درون خود تجمیع کرد:

- مقایسه با SAS: رایگان و متن‌باز بود، در حالی که SAS یک نرم‌افزار تجاری بود.
- مقایسه با MATLAB: در تحلیل‌های آماری پیچیده کارآمدتر بود و به جامعه کاربری بزرگی در حوزه آمار متکی بود.

- مقایسه با Python (در زمان آغاز): Python در آن زمان بیشتر برای توسعه اسکریپت کاربرد داشت و قدرت تحلیل داده‌ای آن مثل امروز توسعه نیافته بود.

۶.۱ ویژگی‌های متمایز زبان

- تمرکز بر آمار: برخلاف زبان‌های عمومی مثل Python یا C++، R از ابتدا با هدف تحلیل آماری طراحی شده است.
- کتابخانه‌های تخصصی: R دارای بیش از ۱۸،۰۰۰ بسته تخصصی در حوزه‌های مختلف مانند ژنتیک، مالی، یادگیری ماشین و غیره است.
- مصورسازی پیشرفته: R ابزارهای قدرتمندی برای ایجاد نمودارهای پیچیده و تعاملی دارد که در مقایسه با زبان‌های مشابه، برجسته‌تر است.
- جامعه کاربری بزرگ: جامعه R بسیار فعال است و منابع یادگیری رایگان بسیاری فراهم کرده است.
- قابلیت‌های تعاملی: با ابزارهایی مثل RStudio و Shiny، R امکان ایجاد داشبوردها و برنامه‌های وب تعاملی را فراهم می‌کند.

۷.۱ ارزیابی بر اساس معیارهای مختلف

- خوانایی: برای کاربران با پس‌زمینه آمار و علوم داده خواناتر است. با این حال، نحو خاص آن ممکن است برای برنامه‌نویسان سنتی دشوار باشد.
- قابلیت اطمینان: در تحلیل‌های آماری و شبیه‌سازی‌ها بسیار قابل‌اعتماد است. با این حال، برای پروژه‌های بزرگ و داده‌های حجیم ممکن است نیاز به بهینه‌سازی داشته باشد.
- هزینه: رایگان و متن‌باز است، که آن را به انتخابی ایده‌آل برای دانشجویان و شرکت‌های نوپا تبدیل می‌کند.
- بهره‌وری و کارایی: برای تحلیل‌های آماری کوچک و متوسط، R بسیار سریع و کارآمد است. اما در مقایسه با زبان‌هایی مثل Python برای پروژه‌های چندمنظوره یا داده‌های حجیم ممکن است کارایی کمتری داشته باشد.
- هزینه یادگیری: برای کسانی که با آمار آشنایی دارند، آسان‌تر است. اما برای برنامه‌نویسان تازه‌کار، زبان‌هایی مثل Python ساده‌تر هستند.
- قابلیت جابجایی: بر روی تمامی سیستم‌عامل‌های رایج (ویندوز، مک، لینوکس) بدون مشکل اجرا می‌شود.
- ابزارهای توسعه: ابزارهایی مانند RStudio و Jupyter Notebook پشتیبانی قدرتمندی برای توسعه و تحلیل داده ارائه شده‌اند.

۸.۱ مقایسه با زبان‌های مشابه

۱.۸.۱ خوانایی

- R: طراحی شده با تمرکز بر آمار و تحلیل داده. کدهای R ممکن است برای تازه‌کاران کمی دشوار به نظر برسند، اما برای کسانی که در تحلیل آماری تجربه دارند، خوانا و ساده است.
- Python: خوانایی بالاتر به دلیل سینتکس ساده و عمومی. یادگیری و استفاده از آن برای افراد تازه‌کار آسان‌تر است.
- MATLAB: خوانا و ساده، مخصوصاً برای مهندسان. ساختار نوشتار آن شبیه به ریاضیات است.
- SAS: نسبتاً سخت‌تر به دلیل سینتکس خاص و قدیمی.

۲.۸.۱ سادگی

- R: برای تحلیل داده ساده و بهینه است، اما سینتکس آن برای کارهای غیرتحلیلی ممکن است پیچیده باشد.
- Python: ساده‌تر به دلیل طراحی چندمنظوره و استفاده گسترده در زمینه‌های مختلف.
- MATLAB: ساده اما گران است، زیرا نیاز به لایسنس دارد.
- SAS: پیچیده‌تر، به خصوص برای کاربران غیرمتخصص.

۳.۸.۱ کتابخانه‌ها

- R: غنی‌ترین مجموعه کتابخانه‌های آماری و تحلیل داده را دارد (CRAN، Bioconductor).
- Python: دارای کتابخانه‌های عمومی قوی مانند Pandas، NumPy، و scikit-learn است. اما برای تحلیل‌های پیشرفته، به R نمی‌رسد.
- MATLAB: کتابخانه‌های ریاضی و شبیه‌سازی قوی دارد، اما برای تحلیل آماری گسترده نیست.
- SAS: ابزارهای تخصصی برای تحلیل داده دارد، اما کتابخانه‌های خارجی محدودند.

۴.۸.۱ کاربرد

- R: تحلیل داده، مدل‌سازی آماری، یادگیری ماشین.
- Python: همه‌منظوره، از تحلیل داده تا هوش مصنوعی و توسعه وب.
- MATLAB: محاسبات عددی، شبیه‌سازی و پردازش سیگنال.
- SAS: تحلیل داده‌های صنعتی و کسب‌وکاری.

۵.۸.۱ کارایی

- R: برای مجموعه داده‌های بزرگ ممکن است کند باشد مگر با استفاده از بسته‌های بهینه‌سازی شده.
- Python: سریع‌تر با کتابخانه‌هایی مانند NumPy و استفاده از Cython.
- MATLAB: کارایی بالا برای مسائل ریاضی و شبیه‌سازی.
- SAS: کارایی بالا در تحلیل داده‌های سازمانی.

۶.۸.۱ هزینه

- R: رایگان و متن‌باز.
- Python: رایگان و متن‌باز.
- MATLAB: بسیار گران و نیازمند لایسنس.
- SAS: گران‌قیمت و مناسب شرکت‌ها.

۷.۸.۱ قدرت آماری

- R: بالاترین قدرت آماری، با گسترده‌ترین ابزارها و مدل‌ها.
- Python: قدرت آماری متوسط. کتابخانه‌های آن برای تحلیل داده کافی هستند، اما به گستردگی R نیستند.
- MATLAB: قدرت آماری کمتر نسبت به R و Python.
- SAS: قدرت آماری بالا، اما بیشتر در حوزه‌های سازمانی کاربرد دارد.

۸.۸.۱ قابلیت جابجایی

- R: قابل نصب روی تمام سیستم‌عامل‌ها (Windows، macOS و Linux).
- Python: بسیار قابل جابجایی با پشتیبانی جهانی.
- MATLAB: نیاز به لایسنس برای هر سیستم‌عامل.
- SAS: معمولاً به صورت سروری استفاده می‌شود و به راحتی قابل انتقال نیست.

۹.۸.۱ نتیجه‌گیری

- اگر قدرت آماری و تحلیل داده اولویت است: R بهترین گزینه است.
- اگر نیاز به یک زبان همه‌منظوره با کاربری گسترده دارید: Python انتخاب بهتری است.
- اگر نیازمند به زبانی برای شبیه‌سازی هستید: MATLAB برتری دارد.
- اگر تحلیل‌های کسب‌وکاری سازمانی مدنظر است: SAS گزینه مناسبی است.

۹.۱ نوع پیاده‌سازی

- تفسیرشده: R به طور اصلی یک زبان تفسیرشده است، به این معنی که کدها به صورت خط به خط توسط مفسر اجرا می‌شوند. این قابلیت به تحلیل‌گران اجازه می‌دهد تا کدها را به صورت تعاملی اجرا کنند و سریعاً نتایج را ببینند.
- کامپایل جزئی (Just-in-Time Compilation): از نسخه ۱۳.۲ به بعد، R از کامپایلر JIT[1] بهره می‌برد که با استفاده از بسته compiler ارائه شده است. JIT بهینه‌سازی‌هایی برای اجرای سریع‌تر کدها فراهم می‌کند.
- پیاده‌سازی ترکیبی: اگرچه R عمدتاً تفسیرشده است، برای برخی از عملیات‌های پیچیده‌تر، از کدهای C و Fortran استفاده می‌شود که کامپایل شده‌اند تا کارایی بیشتری داشته باشند.

۱۰.۱ مفسرها و کامپایلرهای موجود برای R

۱. GNU R (R پایه):

- توسعه‌دهنده: تیم توسعه R که یک گروه جهانی از متخصصان و دانشمندان داده است.
- ویژگی‌ها:
  - مفسر اصلی زبان R.
  - پیاده‌سازی به زبان C و Fortran.
  - استفاده از بسته compiler برای کامپایل جزئی.
- مزایا:
  - منبع‌باز و رایگان.
  - پایدار و استاندارد، با پشتیبانی وسیع از جامعه کاربری.

۲. FastR:

- توسعه‌دهنده: تیم Oracle Labs به عنوان بخشی از پروژه GraalVM.
- ویژگی‌ها:
  - یک پیاده‌سازی جایگزین برای R با هدف افزایش سرعت اجرا.
  - از JIT برای اجرای سریع‌تر کدها بهره می‌برد.
- مزایا:
  - بهینه‌سازی برای کاربردهای داده‌های حجیم و عملکرد سریع‌تر نسبت به GNU R.
  - سازگاری با ابزارهای GraalVM.

۳. Renjin:

- توسعه‌دهنده: BeDataDriven، یک شرکت هلندی.
- ویژگی‌ها:
  - پیاده‌سازی R در JVM[2].

[1] Just-in-Time
[2] Java Virtual Machine

- مناسب برای ادغام با سیستم‌های مبتنی بر جاوا.

- مزایا:
  - کارایی بالا در محیط‌های سازمانی جاوا.
  - امکان ادغام مستقیم با زیرساخت‌های جاوا.

۴. pqR:

- توسعه‌دهنده: Radford Neal، یک آماردان برجسته.
- ویژگی‌ها:
  - یک نسخه بهینه‌شده از GNU R با تمرکز بر اجرای سریع‌تر.
  - بهره‌گیری از پردازش موازی برای بهبود کارایی.
- مزایا:
  - سرعت بالاتر برای تحلیل داده‌ها.
  - مناسب برای کاربران حرفه‌ای تر.

۵. Microsoft R Open (MRO):

- توسعه‌دهنده: مایکروسافت (Microsoft).
- ویژگی‌ها:
  - نسخه بهینه‌شده R با بهبود عملکرد.
  - شامل بسته‌های از پیش کامپایل‌شده و بهینه‌سازی شده برای تحلیل داده.
- مزایا:
  - یکپارچگی با ابزارهای مایکروسافت مانند Azure Machine Learning.
  - سرعت بیشتر در پردازش داده‌های حجیم.

پیاده‌سازی	مزایا	معایب
GNU R	استاندارد و قابل اعتماد، منبع‌باز، جامعه کاربری گسترده	سرعت پایین‌تر نسبت به نسخه‌های بهینه‌شده
FastR	سرعت بیشتر، سازگاری با GraalVM	نیاز به پیکربندی پیشرفته
Renjin	سازگاری با جاوا، مناسب برای محیط‌های سازمانی	محدودیت در پشتیبانی برخی از بسته‌های R
pqR	پردازش موازی، کارایی بالا	جامعه کاربری کوچک‌تر
Microsoft R Open	سازگاری با ابزارهای مایکروسافت، سرعت بیشتر	وابستگی به اکوسیستم مایکروسافت

جدول ۱.۱: مقایسه پیاده‌سازی‌های مختلف زبان R

فصل ۲

نحو و معنا شناسی

۱.۲ نحو و معناشناسی

کلمات کلیدی زبان R شامل موارد زیر است و نمی‌توان از آن‌ها به عنوان نام متغیر یا تابع استفاده کرد:
if، else، repeat، while، function، for، in، next، break، TRUE، FALSE، NULL، NA،
Inf، NaN، NA_integer_، NA_real_، NA_complex_ و NA_character_
توضیح کلمات کلیدی:

۱.۱.۲ if - اجرای کد بر اساس شرط مشخص

Listing 2.1: if

```
if (x > 0) {
  print("is positive")
}
```

۲.۱.۲ else - بلوک جایگزین در صورت عدم تحقق شرط

Listing 2.2: else

```
if (x > 0) {
  print("is positive")
} else {
  print("is negative or zero")
}
```

۳.۱.۲ repeat - اجرای حلقه بی‌نهایت تا زمان استفاده از break

Listing 2.3: repeat

```
i <- 1
repeat {
   print(i)
   if (i == 5) break
   i <- i + 1
}
```

۴.۱.۲ while – اجرای حلقه تا زمانی که شرط برقرار باشد

Listing 2.4: while

```
i <- 1
while (i <= 5) {
   print(i)
   i <- i + 1
}
```

۵.۱.۲ function – تعریف توابع جدید

Listing 2.5: function

```
my_function <- function(a, b) {
   return(a + b)
}
print(my_function(3, 4))
```

۶.۱.۲ for – اجرای حلقه روی مجموعه‌ای از مقادیر

Listing 2.6: for

```
for (i in 1:5) {
   print(i)
}
```

۷.۱.۲ in – تعیین عضویت یا استفاده در حلقه

Listing 2.7: in

```
x <- 5
print(x %in% c(3, 5, 7)) # output: TRUE
```

۸.۱.۲ next – عبور از مرحله جاری و رفتن به مرحله بعدی حلقه

Listing 2.8: next

```
for (i in 1:5) {
  if (i == 3) next
  print(i)
}
```

۹.۱.۲ break – خروج از حلقه

Listing 2.9: break

```
for (i in 1:5) {
  if (i == 3) break
  print(i)
}
```

۱۰.۱.۲ TRUE و FALSE – مقادیر منطقی (Boolean) در R

Listing 2.10: TRUE, FALSE

```
x <- TRUE
y <- FALSE
print(x & y) # output: FALSE
```

۱۱.۱.۲ NULL – نشان‌دهنده مقدار خالی یا بدون مقدار

Listing 2.11: NULL

```
x <- NULL
print(is.null(x)) # output: TRUE
```

۱۲.۱.۲ NA – مقدار غیرموجود یا ناشناخته (Not Available)

Listing 2.12: NA

```
x <- c(1, NA, 3)
print(is.na(x)) # output: FALSE TRUE FALSE
```

۱۳.۱.۲ Inf و Inf- – مقدار بی‌نهایت مثبت و منفی

Listing 2.13: Inf, -Inf

```
print(1 / 0) # output: Inf
print(-1 / 0) # output: -Inf
```

۱۴.۱.۲ NaN – عدد غیرقابل تعریف (Not a Number)

Listing 2.14: NaN

```
print(0 / 0) # output: NaN
```

۱۵.۱.۲ NA_integer_، NA_real_، NA_complex_ و NA_character_ – انواع خاص NA برای داده‌های مختلف

Listing 2.15: NA_integer_, NA_real_, NA_complex_, NA_character_

```
x <- NA_integer_
print(typeof(x)) # output: integer
```

۲.۲ گرامر ساده برای بخشی از زبان

گرامر زیر یک زیرمجموعه کوچک از R را برای ساختارهای کنترلی و عملیات ریاضی توصیف می‌کند:

Listing 2.16: Simple R grammar

```
<program>     ::= <statement> | <statement> ";" <program>
<statement>   ::= <assignment> | <conditional> | <loop> |
    <expression>
<assignment>  ::= <identifier> "<-" <expression>
<conditional> ::= "if" "(" <expression> ")" "{" <program> "}" [
    "else" "{" <program> "}" ]
<loop>        ::= "for" "(" <identifier> "in" <expression> ")" "{"
    <program> "}"
<expression>  ::= <term> | <term> <operator> <expression>
<term>        ::= <number> | <identifier> | "(" <expression> ")"
<operator>    ::= "+" | "-" | "*" | "/"
<number>      ::= [0-9]+
<identifier>  ::= [a-zA-Z_][a-zA-Z0-9_]*
```

۳.۲ برنامه نمونه در زبان

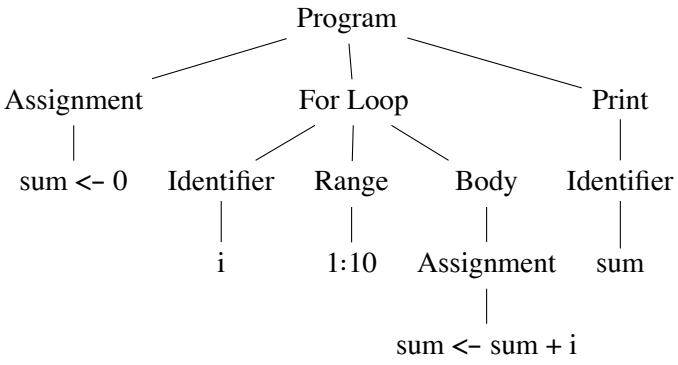
برنامه زیر یک حلقه برای محاسبه مجموع اعداد ۱ تا ۱۰ است:

Listing 2.17: Simple R program

```
sum <- 0
for (i in 1:10) {
   sum <- sum + i
}
print(sum)
```

۴.۲ درخت تجزیه برای برنامه نمونه

درخت تجزیه برنامه ۳.۲ شامل عناصر زیر است:



۵.۲ تقدم و وابستگی عملگرها

در R، تقدم عملگرها به صورت پیش‌فرض تعریف شده است. برخی از عملگرهای رایج با ترتیب تقدم:

۱. داخل پرانتز – ()): بالاترین تقدم.

۲. توابع: توابع مانند sin()، log() تقدم بیشتری از عملگرهای ریاضی دارند.

۳. عملگر توان: بیشترین تقدم میان عملگرهای ریاضی.

۴. ضرب و تقسیم – (* و /): بالاتر از جمع و تفریق.

۵. جمع و تفریق – (+ و –): پایین‌تر از ضرب و تقسیم.

وابستگی عملگرها:

- بیشتر عملگرها از چپ به راست ارزیابی می‌شوند (مثلاً +، –، *، /).

- عملگر توان از راست به چپ ارزیابی می‌شود.

مثال:

Listing 2.18: Precedence of operators

```
x <- 2 + 3 * 4^2
```

توضیح گام به گام:

۱. توان: $۴^۲ = ۱۶$.

۲. ضرب: $۳ * ۱۶ = ۴۸$.

۳. جمع: $۲ + ۴۸ = ۵۰$.

۶.۲ توصیف گرامر بر اساس تقدم عملگرها

برای پیروی از تقدم عملگرها در گرامر، می‌توانیم گرامر را اصلاح کنیم:

Listing 2.19: Operation grammar

```
<expression> ::= <term> | <term> "+" <expression> | <term> "-"
    <expression>
<term>       ::= <factor> | <factor> "*" <term> | <factor> "/" <term>
<factor>     ::= <number> | "(" <expression> ")" | <factor> "^"
    <factor>
```

این گرامر تقدم توان را بالاتر از ضرب و تقسیم و تقدم ضرب و تقسیم را بالاتر از جمع و تفریق تضمین می‌کند.

۷.۲ معناشناسی عملیاتی و توصیف ساختارها

مقایسه برخی از ساختارهای پایه در زبان‌های برنامه نویسی در زبان‌های R، C و اسمبلی:

۱.۷.۲ تخصیص (Assignment)

- کد R

Listing 2.20: R Assignment

```
sum <- 0
```

- معادل در C

Listing 2.21: C Assignment

```
int sum = 0;
```

- معادل در اسمبلی (فرض پردازنده x86)

Listing 2.22: Assembly Assignment

```
MOV sum, 0
```

۲.۷.۲ حلقه for

- کد R

Listing 2.23: R for

```
for (i in 1:10) {
   sum <- sum + i
}
```

- معادل در C

Listing 2.24: C for

```
for (int i = 1; i <= 10; i++) {
   sum += i;
}
```

- معادل در اسمبلی

Listing 2.25: Assembly loop

```
MOV i, 1
MOV sum, 0
LOOP_START:
ADD sum, i
INC i
CMP i, 10
JLE LOOP_START
```

۳.۷.۲ چاپ (Print)

- کد R

Listing 2.26: R Print

```
print(sum)
```

- معادل در C

Listing 2.27: C Print

```
printf("%d", sum);
```

فصل ۳

متغیرها و نوع های داده ای

۱.۳ انقیاد

۱.۱.۳ انقیاد نوع و مقدار

انقیاد نوع (Type Binding)

نوع‌گذاری پویا (Dynamic Typing): در R، متغیرها نیازی به تعریف نوع در زمان اعلان ندارند. نوع متغیر در زمان اختصاص مقدار (assignment) تعیین می‌شود. این نوع‌گذاری به صورت زمان اجرا (Runtime) انجام می‌شود. نوع متغیر می‌تواند در طول برنامه تغییر کند.
مثال:

Listing 3.1: Type Binding

```
x <- 42     # integer
x <- "Hello" # character
```

در این مثال، متغیر x ابتدا یک عدد صحیح (integer) است، اما بعداً به یک مقدار رشته‌ای (character) تغییر می‌یابد. این نشان‌دهنده نوع‌گذاری پویا است.

۲.۱.۳ انقیاد مقدار (Value Binding)

مقداردهی متغیرها در زمان تخصیص مقدار به متغیر انجام می‌شود. از عملگر <- یا = برای اختصاص مقدار به متغیر استفاده می‌شود. انقیاد مقدار در R در زمان اجرا (Runtime) انجام می‌شود، زیرا مقادیر می‌توانند تغییر کنند.

Listing 3.2: Value Binding

```
y <- 10       # Initialization
y <- y + 5    # New value
```

در این مثال، مقدار y ابتدا ۱۰ است و بعداً به ۱۵ تغییر می‌کند.

۳.۱.۳ تعریف متغیرها

صریح یا ضمنی بودن تعاریف: در زبان R، تعاریف متغیرها می‌توانند صریح (Explicit) یا ضمنی (Implicit) باشند.

تعریف صریح

متغیرها به وضوح توسط کاربر تعریف می‌شوند. از عملگر تخصیص (<- یا =) برای این کار استفاده می‌شود.

Listing 3.3: Explicit Variable

```
my_var <- 20
```

تعریف ضمنی

برخی اشیاء در R به صورت ضمنی در زمان استفاده از آنها تعریف می‌شوند، مثلاً در توابع.

Listing 3.4: Implicit Variable

```
for (i in 1:5) {
   print(i)
}
```

در اینجا متغیر i بدون تعریف صریح، به طور خودکار در حلقه for تعریف می‌شود.

۴.۱.۳ تاثیر پویا بودن انقیاد نوع و مقدار

انعطاف‌پذیری زبان R به دلیل انقیاد پویا امکان تعریف متغیرها بدون در نظرگرفتن نوع را فراهم می‌کند، اما ممکن است خطاهای نوع (type error) را در زمان اجرا افزایش دهد.

Listing 3.5: Type error

```
x <- 10
x <- x + "5" # Error: Unable to add number and string.
```

۵.۱.۳ کاربرد و مزایا

- انعطاف‌پذیری: مناسب برای تحلیل داده و کدنویسی سریع.
- سادگی: نیاز به تعریف دقیق نوع و مقدار ندارد.
- کاهش پیچیدگی: مناسب برای کاربران غیر برنامه‌نویس (مانند تحلیل‌گران داده).

۶.۱.۳ نتیجه‌گیری

در زبان R:

۱. انقیاد نوع و مقدار به صورت پویا و در زمان اجرا (Runtime) انجام می‌شود.

۲. تعاریف متغیرها می‌توانند صریح (با تخصیص مقدار) یا ضمنی (در توابع یا ساختارهای کنترلی) باشند.

۳. انعطاف‌پذیری انقیاد نوع و مقدار، R را به زبانی ساده و مناسب برای تحلیل داده و آزمایش سریع ایده‌ها تبدیل کرده است.

۲.۳ بررسی انواع متغیرها

۱.۲.۳ متغیرهای ایستا (Static Variables)

در زبان R، متغیرهای ایستا به صورت صریح وجود ندارند. با این حال، متغیرهایی که خارج از بدنه توابع تعریف می‌شوند (مانند متغیرهای سراسری)، ممکن است به صورت ایستا در نظر گرفته شوند. این متغیرها در طول اجرای برنامه در حافظه باقی می‌مانند. مثال:

Listing 3.6: Static variables

```
global_var <- 100 # Global variable

my_function <- function() {
  print(global_var) # Access to a global variable
}

my_function()
```

- پیاده‌سازی

این متغیرها در فضایی به نام محیط جهانی (Global Environment) ذخیره می‌شوند. محیط جهانی در طول اجرای برنامه فعال است و متغیرها تا زمانی که برنامه در حال اجرا باشد، باقی می‌مانند.

- سرعت تخصیص

متغیرهای ایستا سریع‌تر از متغیرهای پویا تخصیص داده می‌شوند، زیرا در آغاز اجرا تخصیص می‌یابند.

۲.۲.۳ متغیرهای پویا در پشته (Dynamic Stack Variables)

متغیرهای تعریف‌شده داخل توابع، معمولاً در پشته ذخیره می‌شوند. این متغیرها فقط در طول اجرای تابع معتبر هستند. مثال:

Listing 3.7: Dynamic stack variables

```
my_function <- function() {
  local_var <- 10 # Local variable
  print(local_var)
}

my_function()
# Unreachable error outside the function
# print(local_var)
```

- پیاده‌سازی

این متغیرها در محیط تابع (Function Environment) ذخیره می‌شوند که در پشته مدیریت می‌شود. زمانی که تابع تمام می‌شود، محیط آن نیز از بین می‌رود.

- سرعت تخصیص

تخصیص در پشته سریع است، زیرا فقط به تغییر اشاره‌گر پشته نیاز دارد. آزادسازی حافظه نیز خودکار و سریع انجام می‌شود.

۳.۲.۳ متغیرهای پویا در هیپ به طور صریح

در R، هرگاه متغیری مانند یک بردار، لیست یا داده‌ای بزرگ تعریف شود، این متغیرها در هیپ ذخیره می‌شوند. کاربر به طور مستقیم از تخصیص حافظه آگاه نیست، اما تخصیص هیپ توسط مفسر R انجام می‌شود. مثال:

Listing 3.8: Dynamic variables on heap explicitly

```
large_vector <- rep(1, 1e6) # Large vector
print(object.size(large_vector)) # Check the size at the heap
```

- پیاده‌سازی

هیپ در R برای تخصیص حافظه اشیاء بزرگ‌تر و پیچیده‌تر (مانند بردارها و لیست‌ها) استفاده می‌شود. مدیریت حافظه از طریق جمع‌آوری زباله (Garbage Collection) انجام می‌شود.

- سرعت تخصیص

تخصیص هیپ نسبت به پشته کندتر است، زیرا شامل درخواست‌های پیچیده‌تر از سیستم‌عامل و جمع‌آوری زباله است.

۴.۲.۳ متغیرهای پویا در هیپ به طور ضمنی

این نوع تخصیص زمانی رخ می‌دهد که R به صورت خودکار برای متغیرهای کوچک یا موقتی نیز از هیپ استفاده کند. برای مثال، مقادیری که در عملیات‌های موقتی ایجاد می‌شوند. مثال:

Listing 3.9: Dynamic variables on heap implicitly

```
result <- sum(1:100) # Create a temporary value for the range
print(result)
```

- پیاده‌سازی

R ممکن است برای مقادیر موقتی نیز حافظه‌ای در هیپ تخصیص دهد، اما این حافظه به سرعت پس از استفاده آزاد می‌شود.

- سرعت تخصیص

سرعت این تخصیص نسبت به تخصیص مستقیم در پشته کمتر است، اما R از بهینه‌سازی برای کاهش تاثیر عملکرد استفاده می‌کند.

۵.۲.۳ مقایسه انواع متغیر

نوع تخصیص	سرعت تخصیص	دلایل
ایستا (Static)	بسیار سریع	تخصیص در زمان شروع برنامه انجام می‌شود.
پویا در پشته (Stack)	سریع	مدیریت ساده با تغییر اشاره‌گر پشته.
پویا در هیپ (Heap)	متوسط تا کند	نیاز به درخواست‌های سیستم‌عامل و مدیریت زباله.
هیپ ضمنی (Implicit Heap)	کندتر	پیچیدگی در تخصیص موقت و آزادسازی سریع.

جدول ۱.۳: مقایسه انواع تخصیص و سرعت آن‌ها

۶.۲.۳ نتیجه‌گیری

زبان R بیشتر از تخصیص پویا در هیپ استفاده می‌کند، اما تخصیص پشته نیز برای متغیرهای محلی وجود دارد. سرعت تخصیص متغیرها به نوع تخصیص و پیچیدگی اشیاء بستگی دارد. برای داده‌های کوچک و ساده، پشته سریع‌ترین گزینه است، در حالی که برای داده‌های بزرگ یا پیچیده، هیپ استفاده می‌شود که کندتر است. R با استفاده از جمع‌آوری زباله و بهینه‌سازی داخلی، تاثیر کندی تخصیص در هیپ را کاهش می‌دهد.

۳.۳ حوزه‌های تعریف زبان

۱.۳.۳ حوزه تعریف ایستا (Static Scope)

در حوزه ایستا، متغیرها بر اساس محل تعریف‌شان در کد (نه محل فراخوانی) تعیین می‌شوند. R از این نوع حوزه تعریف پشتیبانی می‌کند. وقتی یک متغیر در داخل یک تابع تعریف نشده باشد، R به محیط

لغوی تابع نگاه می‌کند (محیطی که تابع در آن تعریف شده است) و متغیر را از آنجا می‌گیرد. مثال حوزه ایستا:

Listing 3.10: Static scope

```
x <- 10 # variable in the global environment

outer_function <- function() {
  inner_function <- function() {
    return(x) # x takes from the lexical environment
  }
  return(inner_function())
}

print(outer_function()) # output: 10
```

در این مثال، متغیر x در محیط جهانی تعریف شده است و تابع inner_function آن را بر اساس محیط لغوی تابع پیدا می‌کند.

۲.۳.۳ حوزه تعریف پویا (Dynamic Scope)

در حوزه پویا، متغیرها بر اساس محل فراخوانی تابع (نه محل تعریف) تعیین می‌شوند. R به طور پیش‌فرض این نوع حوزه را پشتیبانی نمی‌کند، اما می‌توان آن را با استفاده از ویژگی‌های مدیریت محیط‌ها (مانند parent.frame() یا assign()) شبیه‌سازی کرد. مثال:

Listing 3.11: Dynamic scope

```
dynamic_function <- function() {
  print(x) # x takes from the calling environment
}

caller_function <- function() {
  x <- 20 # Define x in this environment
  dynamic_function()
}

caller_function() # output: 20
```

در این مثال، با استفاده از ویژگی‌های محیط، رفتار حوزه پویا شبیه‌سازی شده است.

۳.۳.۳ افزودن حوزه پویا به زبان R

اگر بخواهیم حوزه پویا را به R اضافه کنیم، باید تغییرات زیر اعمال شود:

۱. مدیریت محیط اجرا (Execution Environment)

به جای استفاده از محیط لغوی، تابع باید به محیط فراخوانی دسترسی داشته باشد. این کار با تغییر نحوه ذخیره و جستجوی متغیرها در محیط‌ها ممکن می‌شود.

۲. پیاده‌سازی نمونه

برای شبیه‌سازی این تغییر، می‌توان از یک مفسر اصلاح‌شده یا تابعی کمکی استفاده کرد.

نمونه کد برای حوزه پویا:

Listing 3.12: Adding a dynamic scope to R

```
dynamic_scope <- function() {
  print(x) # Gets x from the calling environment
}

caller_function <- function() {
  x <- 30 # Define x in the calling environment
  evalq(dynamic_scope(), envir = environment()) # Function
      execution in the calling environment
}

caller_function() # output: 30
```

در اینجا، با استفاده از evalq() و انتقال محیط، رفتار حوزه پویا شبیه‌سازی شده است.

۴.۳.۳ پشتیبانی از هر دو نوع حوزه در R

R به طور مستقیم فقط حوزه ایستا دارد. حوزه پویا را می‌توان با استفاده از محیط‌های پویا (مانند parent.frame() یا evalq()) شبیه‌سازی کرد.

۵.۳.۳ بلوک‌ها و کلمات کلیدی حوزه در R

در R، بلوک‌ها با استفاده از تعریف می‌شوند. هیچ کلمه کلیدی خاصی برای تغییر مستقیم حوزه تعریف وجود ندارد، اما می‌توان با استفاده از توابع مدیریتی مانند assign()، environment()، parent.frame() و eval() به صورت ضمنی حوزه را تغییر داد. مثال بلوک:

Listing 3.13: Blocks in R

```
{
  y <- 50
  print(y)
}
# y is not available outside the block
# print(y) # Error: object 'y' not found
```

۶.۳.۳ نتیجه‌گیری

- حوزه ایستا: R به طور پیش‌فرض از این نوع استفاده می‌کند و متغیرها را از محیط تعریف پیدا می‌کند.
- حوزه پویا: پشتیبانی مستقیم وجود ندارد، اما می‌توان آن را شبیه‌سازی کرد.

- بلوک‌ها: ر R با تعریف می‌شوند و هیچ کلمه کلیدی خاصی برای مدیریت حوزه وجود ندارد، اما توابع محیطی می‌توانند حوزه را به طور غیرمستقیم تغییر دهند.

۴.۳ نوع‌های داده‌ای

۱.۴.۳ نوع‌های داده‌ای پایه (Basic Data Types)

زبان R از انواع داده‌ای مختلفی پشتیبانی می‌کند که برای تحلیل داده‌ها، محاسبات آماری، و کار با داده‌های پیچیده طراحی شده‌اند. در اینجا، تمامی انواع داده‌ای در R، نحوه تخصیص آنها در حافظه، پیاده‌سازی، و ویژگی‌های هرکدام توضیح می‌دهیم:

منطقی (Logical)

- مقادیر: TRUE، FALSE، یا NA (عدم مقدار)
- کاربرد: برای شرط‌ها و عملگرهای بولی.
- عملگرها: & (AND)، | (OR)، ! (NOT).
- تخصیص در حافظه: معمولاً به صورت ۱ بیت یا بیشتر برای ذخیره مقدار منطقی.

مثال:

Listing 3.14: Logical data type

```
x <- TRUE
y <- FALSE
z <- x & y # FALSE
```

عددی (Numeric)

- مقادیر: .مقادیر عددی (شامل مقادیر اعشاری)
- کاربرد: محاسبات ریاضی و آماری.
- عملگرها: +، -، *، /، توان، %% (باقیمانده).
- تخصیص در حافظه: به صورت پیش‌فرض ۸ بایت (۶۴ بیت) برای هر مقدار.

مثال:

Listing 3.15: Numeric data type

```
a <- 3.14
b <- 2
result <- a + b # 5.14
```

عدد صحیح (Integer)

- مقادیر: شامل مقادیر عددی صحیح.
- کاربرد: شمارنده‌ها و محاسبات عدد صحیح.
- ایجاد مقدار: با استفاده از L پس از عدد.
- تخصیص در حافظه: ۴ بایت (۳۲ بیت).

مثال:

Listing 3.16: Integer data type

```
int_val <- 5L
typeof(int_val) # "integer"
```

رشته‌ای (Character)

- مقادیر: شامل مقادیر متنی.
- کاربرد: پردازش رشته‌ها، نام‌گذاری، و تحلیل متن.
- عملگرها: الحاق (paste و paste0).
- تخصیص در حافظه: به صورت آرایه‌ای از کاراکترها با طول متغیر.

مثال:

Listing 3.17: Character data type

```
str <- "Hello"
full_str <- paste(str, "World") # "Hello World"
```

مختلط (Complex)

- مقادیر: شامل مقادیر مختلط a) + .bi).
- کاربرد: محاسبات ریاضی پیشرفته.
- عملگرها: جمع، تفریق، ضرب، تقسیم.
- تخصیص در حافظه: شامل دو مقدار عددی برای بخش حقیقی و موهومی.

مثال:

Listing 3.18: Complex data type

```
comp <- 3 + 2i
Im(comp) # 2
```

۲.۴.۳ ساختارهای داده‌ای پیچیده (Advanced Data Structures)

بردار (Vector)

- مقادیر: مجموعه‌ای از مقادیر همگن.
- کاربرد: تحلیل داده و آرایه‌های یک‌بعدی.
- عملگرها: length و [ ] برای دسترسی.
- پیاده‌سازی: به صورت آرایه در حافظه.

مثال:

Listing 3.19: Vector data type

```
vec <- c(1, 2, 3)
vec[1] # 1
```

ماتریس (Matrix)

- مقادیر: آرایه‌ای دوبعدی از مقادیر همگن.
- کاربرد: محاسبات ماتریسی.
- پیاده‌سازی: به صورت آرایه دوبعدی.

مثال:

Listing 3.20: Matrix data type

```
mat <- matrix(1:6, nrow=2)
mat[1, 2] # 3
```

لیست (List)

- مقادیر: شامل مقادیر ناهمگن.
- کاربرد: ذخیره ساختارهای پیچیده.
- پیاده‌سازی: اشاره‌گرهایی به مقادیر مختلف در حافظه.

مثال:

Listing 3.21: List data type

```
lst <- list(num=1, txt="Hello", vec=c(1,2,3))
lst$num # 1
```

چارچوب داده (Data Frame)

- مقادیر: ساختاری جدولی برای داده.
- کاربرد: ساختاری جدولی برای داده.
- پیاده‌سازی: به صورت لیست با ستون‌های هم‌طول.

مثال:

Listing 3.22: Data Frame data type

```
df <- data.frame(A=c(1,2), B=c("X", "Y"))
df$A #    A
```

آرایه (Array)

- مقادیر: آرایه چندبعدی از مقادیر همگن.
- کاربرد: محاسبات چندبعدی.

مثال:

Listing 3.23: Array data type

```
arr <- array(1:8, dim=c(2,2,2))
arr[1,1,1] # 1
```

۳.۴.۳ اشاره‌گرها و متغیرهای مرجع

R به طور مستقیم اشاره‌گرها را در اختیار کاربر قرار نمی‌دهد. تمامی اشیاء در R به صورت مرجع‌محور (Reference-Based) مدیریت می‌شوند. عملگر address() برای بررسی مکان حافظه استفاده می‌شود.

۴.۴.۳ بازیافت حافظه (Garbage Collection)

R از یک جمع‌آوری زباله (Garbage Collector) استفاده می‌کند.

- پیاده‌سازی: از روش‌های ردیابی و شمارش مرجع برای آزادسازی حافظه اشیاء استفاده می‌شود.
- عملگر: gc() برای اجرای دستی جمع‌آوری زباله.

۵.۴.۳ رفع مشکلات نشتی حافظه و اشاره‌گر معلق

R به دلیل مدیریت خودکار حافظه، از مشکلاتی مانند اشاره‌گر معلق (Dangling Pointer) جلوگیری می‌کند. مدیریت: اشیاء استفاده‌نشده توسط Garbage Collector شناسایی و آزاد می‌شوند.

۶.۴.۳ نتیجه‌گیری

- R طیف وسیعی از نوع‌های داده‌ای ساده و پیچیده را ارائه می‌دهد.
- حافظه به صورت خودکار و بهینه تخصیص می‌یابد.
- بازیافت حافظه به جلوگیری از مشکلات نشتی حافظه کمک می‌کند، و این ویژگی R را برای تحلیل داده‌های پیچیده ایده‌آل می‌کند.

منابع

[۱] The R Project for Statistical Computing: وب‌سایت رسمی زبان R که مستندات رسمی و جامع در مورد زبان، ساختارها، و پیاده‌سازی آن را ارائه می‌دهد.
https://www.r-project.org

[۲] R Documentation: پایگاه داده‌ای گسترده از مستندات R برای توابع و کتابخانه‌های مختلف.
https://www.rdocumentation.org

[۳] Advanced R by Hadley Wickham: کتابی معتبر درباره مفاهیم پیشرفته در R، شامل مدیریت حافظه، پیاده‌سازی داده‌ها، و مفاهیم مرتبط با حوزه تعریف.
https://adv-r.hadley.nz

[۴] R for Data Science: کتابی از Hadley Wickham و Garrett Grolemund که اصول برنامه‌نویسی و تحلیل داده با R را پوشش می‌دهد.
https://r4ds.had.co.nz

[۵] R: in Collection Garbage مستنداتی درباره جمع‌آوری زباله و مدیریت حافظه در R.
https://stat.ethz.ch/R-manual/R-devel/library/base/html/gc.html

[۶] R Programming for Data Science by Roger D. Peng: کتابی با تمرکز بر اصول برنامه‌نویسی R برای دانش داده و تحلیل آماری.
https://leanpub.com/rprogramming

[۷] Stack Overflow (R Tag): انجمنی فعال برای پرسش و پاسخ در مورد مسائل برنامه‌نویسی R.
https://stackoverflow.com/questions/tagged/r

[۸] CRAN (Comprehensive R Archive Network): مخزن رسمی R برای دانلود بسته‌ها و مستندات مرتبط.
https://cran.r-project.org

[۹] W3Schools: وب‌سایتی آموزشی برای یادگیری زبان R و دیگر تکنولوژی‌های برنامه‌نویسی.
https://www.w3schools.com/r/