

VYUŽITÍ PROGRAMOVACÍHO JAZYKA R PRO EFEKTIVNÍ ZPRACOVÁNÍ GEOLOGICKÝCH DAT

D. Burger

Ústav geologických věd, Masarykova univerzita, Kotlářská 2, 611 37 Brno, xburger@sci.muni.cz

R-language je volně šiřitelný multiplatformní software s kvalitním výstupem vektorové a bitmapové grafiky určený pro analýzu dat, matematické a statistické výpočty (Ihaka a Gentleman 1996; Bates et al. 2002). Je vhodnou alternativou běžně dostupného kancelářského i obvykle používaného komerčního softwaru speciálně určeného pro geochemické výpočty (Janoušek 2000) a je užitečným nástrojem aplikovatelným v geologických vědách (např. Janoušek 2000, Grunsky 2002). Mezi výhody programu patří snadná dostupnost (volně stažitelný z www.r-project.org), poměrně dobrá přenositelnost zdrojového kódu mezi jednotlivými platformami, neustálý vývoj a snadný update, množství zabudovaných funkcí pro manipulaci s daty, jejich statistické zpracování a grafický výstup, možnost tvorby vlastních funkcí i celých programových balíků. Součástí R-language je grafické rozhraní s „run-time“ prostředím a debuggerem, vlastní jazyk má syntaxi podobnou jiným skriptovacím programovacím jazykům postavených na C. Uživatelská znalost programování není nutná v případě použití některých komplexních programových balíků jako je např. Geochemical Data Toolkit (=GCDkit – viz Janoušek et al. 2003), nicméně naučit se vytvářet vlastní programy v R není nic složitého, k dispozici je celá řada manuálu a ukázkových skriptů. Velkou předností je podpora OOP (objektově orientovaného programování), regulárních výrazů POSIX standardu i syntaxe jazyka Perl, podpora nerelačních i relačních typů databází v rámci programových balíků.

Kvality tohoto programovacího jazyka lze demonstrovat na aplikaci sloužící pro přepočty minerálních analýz a geothermobarometrické výpočty, vykazující rysy modulově postaveného softwaru, kombinujícího uživatelský komfort s mnoha užitečnými funkcemi, zahrnujícími například vylepšený import/export dat, možnost korekce parametrů kalkulací, klasifikaci některých minerálních skupin, alokací pozic ve vzorcích, dopočítávání vakancí, ukládání, načítání a editaci vypočtených výsledků a diagramů, výběr skupin dat přímo z diagramu apod. Předností je rychlé zpracování velkého množství dat, řádově ve vteřinách v závislosti na počtu úkonů a počítačové agilitě konkrétního uživatele.

Bates, D., Chambers J., Dalgaard, P., Gentleman, R., Hornik, K., Iacus, S., Ihaka, R., Leisch, F., Lumley, T., Maechler, M., Masarotto, G., Murrell, P., Ripley, B., Temple, L. D. and Tierney, L. [The R Development Core Team] (2002): The R Environment for Statistical Computing and Graphics. Version 1.6.1 (2002-11-01)

Grunsky, E.C. (2002): R: a data analysis and statistical programming environment an emerging tool for the geosciences. *Comput. and Geosci.*, 28, 1219-1222.

Ihaka, R. and Gentleman, R. (1996): R: A language for data analysis and graphics. *Journal of Computational and Graphical Statistics*, 5, 299-344.

Janoušek, V., Farrow, C. M. and Erban, V. (2003): GCDkit: new PC software for interpretation of whole-rock geochemical data from igneous rocks. *Geochimica et Cosmochimica Acta* 67, A186.

Janoušek, V. (2000): R an alternative to spreadsheets and special software for geochemical calculations and plotting. *Geolines*, 10, 34-35.