

Program szkolenia

Linux oraz Wolne i Otwarte Oprogramowanie dla Data Science

Prowadzący: Damian Melniczuk

1. Instalacja systemu Debian Wheezy
 - Przeprowadzenie instalacji systemu
 - Omówienie zalet i wad systemu Debian
 - Charakterystyka systemu partycji i plików
 - Wyjaśnienie idei szyfrowania
2. Środowisko graficzne Gnome 3
 - Demonstracja możliwości
 - Przedstawienie wskazówek odnośnie użytkowania
3. Terminal tekstowy
 - Omówienie podstawowych komend
 - Tips and Tricks
4. Repozytoria, pakiety i instalacja
 - Aktualizacja systemu
 - Instalowanie programów z repozytoriów
 - Zarządzanie repozytoriami
5. Skrypty BASH i podstawowe narzędzia
 - Dlaczego warto pisać skrypty
 - awk, grep, sed, czyli bardzo proste i przydatne programiki
 - Przykład pracy z surowymi danymi, czyli użycie podstawowych narzędzi do przygotowania danych
6. System kontroli wersji
 - Dlaczego powinno używać się systemów kontroli wersji
 - Przykład użycia
7. Backupy
 - Omówienie zagadnienia
 - Konfiguracja programu
8. Praca zdalna
 - SSH jako standardowe narzędzie pracy każdego mobilnego linuxowca
 - MOSH jako lepszy SSH
 - SCP - czyli jak skopiować dane z "komputera biurkowego" na "laptopa terenowego"
 - Omówienie programu Byobu

9. Selekcja oprogramowania z oficjalnych repozytoriów Debiana

- Omówienie poszczególnych programów – GNU R, Python i SciPy/Sickit, Octave, Weka
- Wskazanie wersji dostępnych w repozytorium

10. Ręczna kompilacja i instalacja oprogramowania spoza repozytoriów

- Dodatkowe pakiety dla R i moduły dla Pythona
- Repozytorium Cloudera i Hadoop
- Przykład ręcznej kompilacji