

**ANTG**

# **ZDOBYWANIE DANYCH W SIECI INTERNET**

## **ZA POMOCĄ PYTHONA**

## **PARSER / GRABBER — TO**

NARZĘDZIE / SCRIPT / BOT DO WYLICZENIA DANYCH  
ZE ŹRÓDŁA, GDZIE ŹRÓDŁO TO STRONA,  
PLIK, KOD, POŁĄCZENIE I T.D.

## **PO CO PARSOWAC DANE?**

1. MASZ ZAWSZE AKTUALNĄ INFORMACJĘ.
2. NIE MUSISZ RĘCZNIE ZARZĄDZAĆ PROCESEM.
3. MOZESZ ZDOBYWAĆ DOWOLNĄ INFORMACJĘ W SIECI.
4. MOŻESZ AUTOMATYZOWAĆ CAŁĄ SWOJĄ PRACĘ.

JASNE, ŻE TO NIE WSZYSTKO...

## **GDZIE UZYWAJA PARSERY?**

1. KURSY WALUT (GIEŁDY).
2. NOWOŚCI W CZASIE RZECZYWISTYM (PORTALE).
3. SKLEPY INTERNETOWE (NAPEŁNIENIE, POBIERANIE OPISÓW I T.D).
4. MONITOROWANIE OPINII, GUSTÓW UŻYTKOWNIKÓW, AKTYWNOŚCI.
5. GDZIE JEST TREŚĆ DYNAMICZNA.

JASNE, ŻE TO NIE WSZYSTKO...

## **CZY TO JEST LEGALNE?**

1. JEŻELI JEST INTERFEJS API, TO TAK.
2. JEŻELI NIE MA INTERFEJSU, ALE JEST COPYRIGHT ZABRONIAJĄCY KOPIOWANIE MATERIAŁÓW BEZ ZEZWOLENIA, TO NIE.
3. JEŻELI NIE MA ANI INTERFEJSU ALBO COPYRIGHT'U, TO TAK.  
(TO ZNACZY ŻE NIE JEST ZABRONIONE)

## **JAKIE NARZEDZIA DO GRABBINGU MA PYTHON?**

1. STANDARDOWA BIBLIOTEKA — URLLIB (Py2 — urllib2).
2. FRAMEWORK DO POBIERANIA DANYCH — SCRAPY.
3. BIBLIOTEKA DO PARSINGU HTML DOM — BEAUTIFULSOUP.
4. FRAMEWORK DO TESTÓW AUTOMATYCZNYCH — SELENIUM.  
+ MODUL XLRD DO PRACY Z XLS-PLIKAMI.

JASNE, ŻE TO NIE WSZYSTKO...

## **JAKIE NARZEDZIA MA PRZEGLADARKA?**

KONSOLA DEVELOPERA.  
DOSTĘP DO NARZĘDZIA — KLAWISZ F12.

NO I PARĘ PRZYKŁADÓW NA KONIEC...  
**DZIENKI!**