

PageRank a SRank v Python-u

Obsah

PageRank a SRank v Python-u.....	1
PageRank.....	1
SRank.....	1
Závěrem.....	1
Appendix A: Kód na získání PRank.....	2
Appendix B: Kód na získání SRank.....	5

Minulý týden nás Google poctil svým updatováním PageRank-u. Někomu se zvýšil, někomu snížil, ale tak to už v životě chodí. Každý, kdo si chtěl zjistit svůj aktuální PRank, musel využít jednu z nespočet a nástrojů (kontrola PageRank [se nachází i na Webtrh-u](#)). Proč ale zadávat pro někoho citlivá data do nástrojů někoho jiného? Vždyť skript na kontrolu PageRanku a Sranku si můžeme jednoduše vytvořit i doma na koleně. A my si ukážeme, jak to jde v Python-u.

PageRank

Základem získání PageRank-u (PRank) je použití featury z Google Toolbar-u. PageRank se získá z URL

```
http://www.google.com/search?client=navclient-  
auto&features=Rank:&q=info:%s&ch=%s
```

Kde za parametr q doplním URL adresy, které chceme získat PR a za parametr ch doplníme hash vypočtený z dané URL adresy. Samotný výpočet je dost komplikovaný a celý kód najdete v příloženém PDF dokumentu.

SRank

SRank neboli SeznamRank se dá zjistit v Python-u úplně jednoduše. Python má v sebe už implicitně podporu pro XML-RPC, takže výsledný kód je jen na pár řádků:

```
rv = xmlrpclib.ServerProxy('http://srank.seznam.cz').getRank('0',  
url, 0)
```

Samozřejmě, je nutno zkontrolovat návratovou hodnotu a výsledné číslo vydělit hodnotou 2.55 a zaokrouhlit.

Závěrem

Snad budou zmíněné utility k něčemu užitečnému, například k hromadnému ověřování URL adres bez pomoci třetích stran. A dávat cenné data třetím stranám (jako jsou soukromé web projekty) je někdy nechcené.

Appendix A: Kód na získání PRank

""""

Google Toolbar 3.0.x/4.0.x Pagerank Checksum Algorithm

Autorova stránka: pagerank.gamesaga.net

""""

```
def IntStr(String, Integer, Factor):
    for i in range(len(String)) :
        Integer *= Factor
        Integer &= 0xFFFFFFFF
        Integer += ord(String[i])
    return Integer

def HashURL(Str):
    C1 = IntStr(Str, 0x1505, 0x21)
    C2 = IntStr(Str, 0, 0x1003F)

    C1 >>= 2
    C1 = ((C1 >> 4) & 0x3FFFC0) | (C1 & 0x3F)
    C1 = ((C1 >> 4) & 0x3FFC00) | (C1 & 0x3FF)
    C1 = ((C1 >> 4) & 0x3C000) | (C1 & 0x3FFF)

    T1 = (C1 & 0x3C0) << 4
    T1 |= C1 & 0x3C
    T1 = (T1 << 2) | (C2 & 0xF0F)

    T2 = (C1 & 0xFFFFC000) << 4
    T2 |= C1 & 0x3C00
    T2 = (T2 << 0xA) | (C2 & 0xF0F0000)

    return (T1 | T2)

def CheckHash(HashInt):
    HashStr = "%u" % (HashInt)
    Flag = 0
    CheckByte = 0
```

Víc informací na <http://www.kiwwisk.cz/>

```

i = len(HashStr) - 1
while i >= 0:
    Byte = int(HashStr[i])
    if 1 == (Flag % 2):
        Byte *= 2;
        Byte = Byte / 10 + Byte % 10
    CheckByte += Byte
    Flag += 1
    i -= 1

CheckByte %= 10
if 0 != CheckByte:
    CheckByte = 10 - CheckByte
    if 1 == Flag % 2:
        if 1 == CheckByte % 2:
            CheckByte += 9
        CheckByte >>= 1

return '7' + str(CheckByte) + HashStr

def RetPr(URL):
    import urllib
    hsh = CheckHash(HashURL(URL))
    gurl = 'http://www.google.com/search?client=navclient-
auto&features=Rank:&q=info:%s&ch=%s' % (urllib.quote(URL), hsh)

    f = urllib.urlopen(gurl)
    st = f.read()
    st = st.lstrip().rstrip()
    st = st[9:]
    try:
        st = int(st)
    except:
        st = 0

    return st

if __name__ == '__main__':
    import sys
    import urllib
    if len(sys.argv) < 3:

```

Víc informací na <http://www.kiwwisk.cz/>

```
print 'No URL specified.'  
sys.exit()  
  
if sys.argv[1] == 'PR':  
    print RetPr(sys.argv[2])
```

Appendix B: Kód na získání SRank

```
def srank( url ):
    import xmlrpclib

    s = xmlrpclib.ServerProxy('http://srank.seznam.cz')
    rv = s.getRank('0', url, 0)

    if rv['status'] is not 200:
        raise Exception, "Error getting srank for domain %s, error %i" % (url,
rv['status'])

    return int( round( rv['rank'] / 2.55 ) )

if __name__ == '__main__':
    import sys

    if len(sys.argv) < 3:
        print 'No URL specified.'
        sys.exit()

    if sys.argv[1] == 'SR':
        print srank(sys.argv[2])
```