

# USING PYTHON TO ACCESS WEB DATA

**By : Rahul Bajaj**

WHAT IS THIS  
PRESENTATION ALL  
ABOUT?

# SOCKETS IN PYTHON

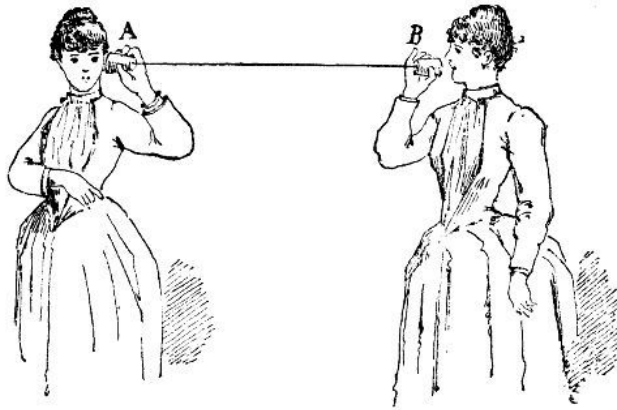


FIG. 76. Trådtelefon.



# SOCKETS IN PYTHON

- Python has built-in support for TCP sockets.

```
import socket
mysock = socket.socket(socket.AF_INET, socket.SOCK_STREAM)
mysock.connect( ('www.py4inf.com', 80) )
```

Host



Port

# AN HTTP REQUEST IN PYTHON

```
import socket
mysock = socket.socket(socket.AF_INET, socket.SOCK_STREAM)
mysock.connect(('www.py4inf.com', 80))

mysock.send('GET http://www.py4inf.com/code/romeo.txt HTTP/1.0\n\n')
```

while True:

```
    data = mysock.recv(512)
    if ( len(data) < 1 ) :
        Break
    print data
mysock.close()
```

URL DOCUMENT PROTOCOL

# OUTPUT

**HTTP/1.1 200 OK**

**Date: Sun, 14 Mar 2010 23:52:41 GMT**

**Server: Apache**

**Last-Modified: Tue, 29 Dec 2009 01:31:22 GMT**

**ETag: "143c1b33-a7-4b395bea"**

**Accept-Ranges: bytes**

**Content-Length: 167**

**Connection: close**

**Content-Type: text/plain**

**But soft what light through yonder window breaks**

**It is the east and Juliet is the sun**

**Arise fair sun and kill the envious moon**

**Who is already sick and pale with grief**

MAKING HTTP EASIER  
WITH  
URLLIB!

# URLLIB IN PYTHON

- Since HTTP is so common, we have a library that does all the socket work for us and makes web pages look like a file

```
import urllib
fhand = urllib.urlopen('http://www.py4inf.com/code/romeo.txt')
for line in fhand:
    print line.strip()
```

But soft what light through yonder window breaks  
It is the east and Juliet is the sun  
Arise fair sun and kill the envious moon  
Who is already sick and pale with grief



SO, NOW WHAT IS

BETTER ?

SOCKET..?

URLLIB...?

OR

MAYBE SOMETHIGN

ELSE :P

THANKS