

Servidor Proxy Web com Controle de Acesso

Você fará um servidor Proxy HTTP usando sockets em Python. Qualquer requisição HTTP feita pelo navegador deve ser direcionada ao Proxy que se encarregará de retornar a resposta apropriada. No geral, esta resposta conterá o documento solicitado. O Proxy também implementará controle de acesso com base no endereço dos servidores Web da seguinte maneira: para qualquer domínio contido numa “lista negra” o Proxy negará a contactar o respectivo host. Por exemplo, caso o nome ufmt.br esteja na lista, qualquer endereço terminado com ufmt.br (ex.: comp.cua.ufmt.br) será bloqueado.

Cada requisição processada pelo Proxy cai em um dos 3 casos abaixo:

1. O documento solicitado não existe no banco de dados (BD) local – O proxy contacta o servidor Web contido no campo Host da requisição HTTP, obtém o documento desejado, retorna-o ao navegador e por fim o armazena na base de dados local juntamente com informações extra como o endereço web do documento e data de acesso;
2. O documento solicitado existe no BD local – O Proxy verifica junto ao servidor Web se a cópia local corresponde ao documento original. Em caso afirmativo, o Proxy simplesmente retorna o documento ao navegador. Caso contrário a cópia é primeiro atualizada e depois retornada ao navegador;
3. Não é permitido o acesso ao servidor Web contido na requisição – O Proxy simplesmente retorna ao navegador uma página padrão informando que não é permitido o acesso ao referido servidor.

A qualquer momento o Proxy permite que a lista de controle de acesso seja alterada. Para isso o navegador deve contactar diretamente o proxy (digitando o endereço IP do Proxy na barra de endereço), que retornará uma tabela contendo a lista negra e um formulário onde é possível adicionar novos domínios ou remover algum existente.

Utilize o SQLite3 para o banco de dados. Como o Python oferecer suporte nativo a este SGBD, não é necessário instalar nenhuma ferramenta extra para utilizá-lo.

Para efeitos de simplificação você pode considerar apenas sites que utilizam o protocolo HTTP (não é necessário se preocupar com o HTTPS).

O Proxy deve ser multithread para poder atender várias requisições simultaneamente.