Intelligent Sentinel Agent

Eduardo Leiva Correa
Pontifícia Universidade Católica do Rio Grande do Sul (PUC-RS)
ecorreo@ig.com.br

Abstract

The target of this text is present the project of a Software Agent capable of managing data received through the Internet. The paradigm of the agent is limited in a identification of unwanted subjects as pornography, occultism, drugs and so one specified by the supervisor. To execute this work, a Knowledge Base is created by means of words connected with the mentioned subjects. The decisions of the agent are supported by recognition of these words received through the Internet and later analysis with a decision tree. More details can be found at www.inf.unisinos.br/pos-redes/seguranca/edicao3/isa/isa.html

1 INTRODUÇÃO

1.1 Apresentação

O objetivo deste texto é apresentar um Agente de Software capaz de monitorar informações provenientes da Rede Mundial. Elaborei este agente como trabalho de conclusão de grau em dezembro de 1999 e desde aquela época venho aperfeiçoando-o em diversos aspectos.

O paradigma do agente limita-se à identificação de textos que abordam assuntos de qualquer natureza como por exemplo: pornografia, drogas, terrorismo, jogos em geral, etc. Para cumprir com essa tarefa, existe uma base de conhecimento cujo conteúdo são palavras relacionadas, em maior ou menor grau, com os assuntos acima citados. As decisões a serem tomadas pelo agente basear-se-ão no reconhecimento dessas palavras-chave que venham a ser encontradas nos textos recebidos através da WWW (Word Wide Web).

Esse Agente poderá analisar o fluxo de texto que ocorre num computador que utilize um modem como meio de comunicação com a Internet, assim como, àqueles que sejam providos de uma placa de rede para tal comunicação. Observe o leitor que este agente também pode ser utilizado por provedores de Internet para filtrar informação vinda da Rede Mundial (firewall).

Neste software, duas classes de usuários são bem definidos: um supervisor, que poderá treinar o agente, configurá-lo e receber mensagens de alerta do mesmo, e o usuário final, cuja atuação será monitorada pelo agente aqui exposto.

Para o agente é indiferente o navegador utilizado pelo usuário final, o qual pode ser o Netscape, Explorer ou similar.

Este produto visa oferecer a pais de família, escolas, universidades e a sociedade em geral, uma barreira capaz de propiciar ambientes controlados para adolescentes e crianças. Assim sendo, chegamos à conclusão de que este agente é de interesse de todos nós.

2 CONCEITOS BÁSICOS

2.1 WWW

O agente, objeto deste projeto, deverá ser capaz de acessar determinados endereços dentro da Word Wide Web, requisitar e transportar frames para posterior análise. Sendo assim, apresento, brevemente, conceitos e tecnologias envolvidas nesse processo.

O conceito World Wide Web relaciona-se com objetos, denominados objetos WWW, que podem ser hipertextos os quais estão armazenados em servidores espalhados por toda a inter-rede. Esses servidores são denominados servidores WWW e atendem solicitações de programas clientes denominados clientes WWW.

[SOARES97] define hipertexto como um documento composto por um conjunto de nós (fragmentos de informação em diversas mídias, por exemplo, imagem, texto, som, vídeo) interligados por elos definidos por um par de âncoras. As âncoras podem ser um nó ou uma região dentro de um nó.

Para que toda essa comunicação seja possível, existem protocolos e meios físicos capazes de executar dita tarefa. Esse hardware é relacionado com o termo *Internet*, tema do próximo assunto.

2.2 Internet

Podemos definir o termo como: conjunto de redes e roteadores que abarca muitos países e utiliza os protocolos TCP/IP para formar uma só rede virtual cooperativa [COMER96].

Uma Rede de Redes pode ser imaginada como uma rede unificada que engloba muitas redes diferentes. Entretanto, a arquitetura subjacente da rede de redes fica oculta. Isto é, o usuário ou programador de aplicação não é obrigado a entender os detalhes da interconexão do hardware para utilizar a internet. Por outro lado não se quer impor uma nova topologia de interconexão de redes. A intenção da Rede de Redes é enviar dados através de redes intermediárias mesmo que não estejam interligadas de forma direta às máquinas de origem ou destino. Quer-se, ainda, que todas as máquinas, participantes da Internet, sejam identificadas através de um nome e um endereço.

Desde o ponto de vista do usuário, a Rede de Redes é uma única rede na qual todas as máquinas conectam-se e onde não é necessário conhecer as ligações físicas. Talvez o usuário saiba que para que toda essa estrutura funcione coerentemente, deve seguir normas e padrões preestabelecidos, ou seja, deve possuir um protocolo de comunicação. O protocolo que a Rede de Redes utiliza é conhecido como TCP/IP.

2.3 Protocolo TCP/IP

Protocolo pode ser pensado como uma descrição formal de formatos de mensagens e regras que dois ou mais máquinas devem seguir para intercambiar dados. Os protocolos podem descrever detalhes de baixo nível das interfaces de máquina a máquina ou do intercâmbio entre programas aplicativos [COMER96].

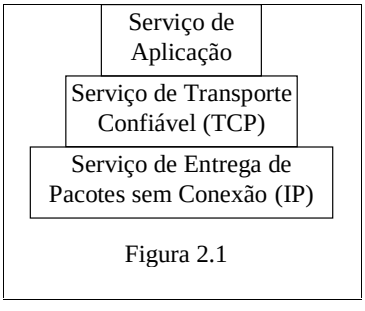

Figura 2.1

O software de Internet está desenhado em torno a três conceitos de serviço de rede organizados hierarquicamente; muitos dos êxitos obtidos devem-se a essa arquitetura surpreendentemente robusta e adaptável. Imaginando uma pirâmide (ver figura 2.1), onde existe dependência entre os degraus, na base encontramos um serviço de entrega de pacotes *sem conexão*, conhecido como *Protocolo Internet* ou simplesmente, IP. No seguinte degrau,

podemos imaginar um serviço de transporte *confiáve*l, conhecido como *TCP Transfer Control Protocol*. E na ponta da pirâmide residem os *serviços de alto nível* da rede de redes os quais determinam como os usuários percebem a Internet e demonstram o poder dessa tecnologia.

2.4 Frames

O agente recebe informações da Rede Mundial na forma de quadros ou frames. Esses quadros são compostos por uma seqüência de centenas ou milhares de bytes.

O mecanismo através do qual o agente captura esses frames, pode ser apreciado na figura 2.2.

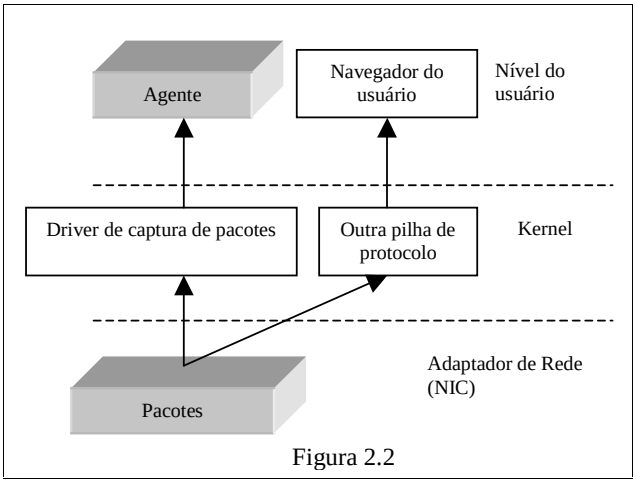

Figura 2.2

No nível inferior encontramos um adaptador de rede ou NIC (Network Interface Card). Essa placa de rede captura os pacotes destinados ao usuário. Nos casos em que o computador não estiver em ambiente de rede, é utilizado o *dial-up* para cumprir com dita tarefa. O driver recebe os pacotes e os encaminha para o agente. Nesse processo são utilizados buffers com o intuito de não criar congestionamento e dessa forma, perda de pacotes.

Testes realizados em computadores isolados, apontam para uma degradação de performance mínima, por não dizer imperceptível. Já nos casos em que o agente cumpra as tarefas de um firewall, o assunto degradação é de extrema importância. Com o auxílio de buffers, essa degradação pode ser minimizada até um ponto aceitável.

2.5 Agentes de Software

Conceituar o termo “agente”, não é tarefa fácil, muitas definições podem ser apreciadas na bibliografia existente. Apresentamos, à continuação, a definição de Russell,a qual utilizaremos, ao longo deste trabalho

Segundo [RUSSELL 95] um agente é algo que pode ser visto como percebendo o seu meio através de sensores e atuando naquele meio. O comportamento de um agente baseia-se nas informações recebidas do meio e da base de conhecimento, construída para o ambiente particular no qual atua.

O comportamento de um agente baseia-se na sua experiência e na base de conhecimento construída para um determinado ambiente onde atua. Desta forma podemos afirmar que o agente é autônomo o que significa que o seu comportamento não depende de atitudes humanas ou de fatores externos.

Podemos pensar em nosso agente específico como percebendo palavras chaves dentro de um meio, qual seja um grupo de frames contendo texto, e em posse de uma base de conhecimento, decide por uma das seguintes ações : cancelar a comunicação com a rede de redes ou continuar monitorando a navegação. Previamente o agente deve ser submetido ao processo de aprendizagem. Este processo será examinado mais adiante.

2.6 Aprendizagem

A idéia do aprendizado é de que a percepção seja utilizada não somente para decidir qual ação executar, mas, para fornecer ao agente a habilidade de atuar no futuro. A profundidade do aprendizado pode variar desde uma memorização da experiência até a elaboração de teorias científicas.

[RUSSELL 95] propôs um modelo geral de agente de aprendizado, composto de quatro módulos principais: *elemento de aprendizado, elemento de performance*, *módulo crítico* e *gerador de problema.*

O agente proposto neste trabalho, não apresenta toda a complexidade que demandaria a construção de um agente como o acima apresentado. Entretanto, o *elemento de aprendizado* está presente no nosso agente.

O agente constrói um elemento de aprendizado utilizando exemplos que contenham uma lista formada por duplas (percepções, ação). Quando o elemento de aprendizado depara-se com uma percepção existente em tal lista, executa a ação correspondente. Do contrário, chama ao algoritmo de aprendizado *induzido* quem retorna uma hipóteses h a ser utilizada pelo agente para escolher a ação a ser executada. Uma das abordagens utilizadas na construção desse algoritmo de aprendizado é conhecida como *arvores de decisão*.

Indução por árvores de decisão é uma das formas mais simples de implementar algoritmos de aprendizagem. A árvore toma como entrada um objeto ou situação caracterizada por um grupo de atributos e como resultado, é obtido um sim ou não. Cada nó interno da árvore, corresponde a um teste a respeito de um dos atributos do objeto em questão. Os galhos que saem do nó simbolizam os possíveis resultados do mencionado teste. Cada folha da árvore especifica um valor booleano a ser retornado como resultado da pesquisa.

2.7 Base de Conhecimento

Podemos definir Base de Conhecimento (B.C) como um depósito de dados dinâmico utilizado como fundamento para decidir qual ação executar. Essa B.C pode ser interpretada como sendo a "experiência" do agente.

Com o intuito de identificar texto duvidoso, é utilizada uma base de conhecimento composta por palavras comprovadamente existentes nesse tipo de texto. A cada uma dessas palavras atribui-se um valor que indica a frequência de ocorrência da mesma, nos textos analisados pelo agente. Quando o agente executa o processo de aprendizado, essa base de conhecimento sofre mutações ocasionadas pela inclusão de novas palavras, como veremos em seguida.

3- ARQUITETURA GERAL

Utilizando a perspectiva do Fluxo de Dados, podemos imaginar o seguinte fluxograma do agente:

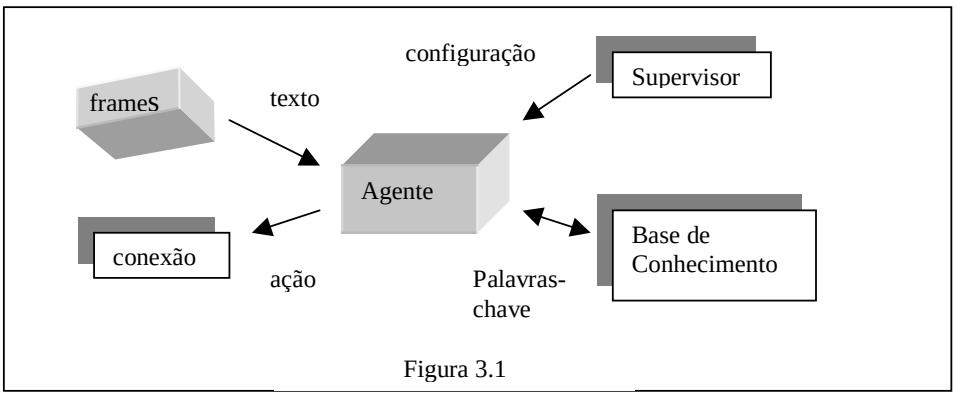

Figura 3.1

Na figura 3.1 o Agente recebe os frames os quais são analisados; nessa análise é utilizada a Base de Conhecimento. Se o texto em análise for considerado indesejado, a classe Conexão executa uma ação. A escolha da ação a ser executada é dependente da configuração especificada pelo Supervisor; a qual pode ser enviar um mail de aviso ao Supervisor e/ou retirar o navegador da memória do computador e encerrar a comunicação.

Desde o ponto de vista do conceito de Agente e o meio de atuação, podemos verificar o seguinte esquema:

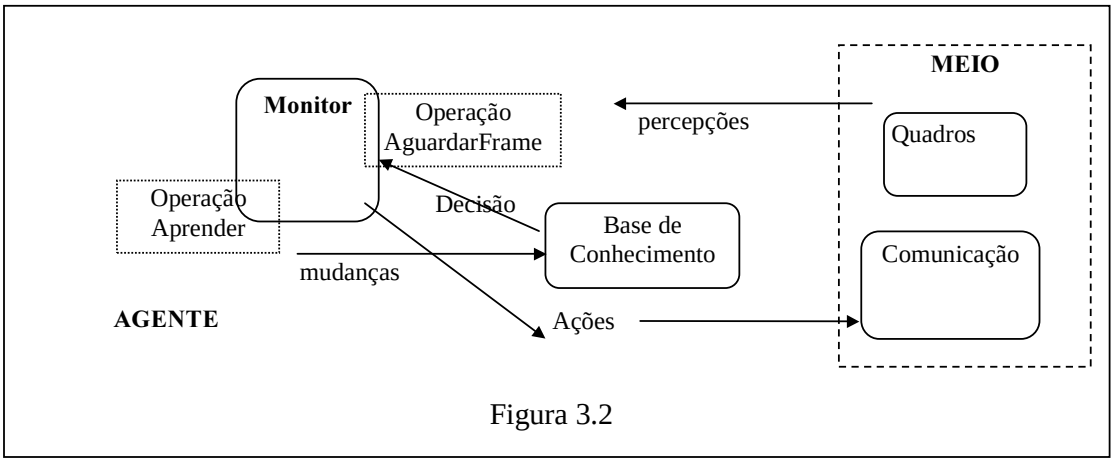

Figura 3.2

4 ESTUDO DE CASO

4.1 Treinamento e Base de Conhecimento

Entre os diversos assuntos possíveis a escolher, acredito que pornografia tenha uma conotação especial, sendo assim, selecionei três endereços para treinar o agente: www.playboy.com , www.erotica.com.br , e www.sexo.com.br .Basta digitar o endereço na *tela treinamento* do agente e ele realiza a conexão, faz a requisição daquela url, armazena os frames no disco local, retira dos frames palavras irrelevantes como aquelas que fazem parte da linguagem html, java script, artigos, numerais, preposições, etc. Por último, registra as palavras relevantes na B.C junto com a frequência de ocorrências e a data. Observe a B.C obtida:

palavra	frequência	palavra	frequência	Palavra	frequência	palavra	frequência
uol	8	Pixel	8	gostosa	4	anal	4
playboy	183	Storeslides	5	Sexo	10	teens	2
entertainment	9	leftnav	14	Sex	20	mustbe	4
Script	7	&	6	Girls	3	ffc	4
Creatives	8	>	9	shop	2	menos	2
imx	57	gatas	54	Oral	3	gatinha	4
video	2	xxx	3	Teenage	3	nude	3
porno	3	porn	2	Pussy	2	Adult	3
Pictures	4	Nudity	9	Babes	2	Ass	2
Club	10	Blockquote	2	Normal	2	Aceito	2

Observe que diversas palavras irrelevantes são inseridas na B.C. Essas palavras serão retiradas , com o transcorrer do tempo, pelo Agente em virtude de que não serão utilizadas; em outras palavras, na medida em que mais páginas pornográficas sejam visitadas, mais apurada será a B.C

.4.2 Monitoração e Aprendizagem

Uma vez treinado o agente, como visto acima, este fica residente na memória do computador apresentando apenas um ícone ao lado do relógio do computador (tray icon). Quando um ou diversos frames

sejam identificados como indesejados, o agente armazena esses frames em local previamente indicado, e executa uma ação qual seja enviar um mail ao supervisor e/ou retirar o navegador da memória. Em posse desses frames armazenados em local especial, o agente aprende novas palavras e descarta aquelas cuja frequência seja inferior ao indicado pelo supervisor. Dessa forma, a B.C. é refinada com o transcorrer do tempo.

O usuário Supervisor pode tornar o agente mais ou menos rigoroso em sua análise. Para tal, é utilizada uma escala de zero a duzentos. Um valor vinte indica certa tolerância no que se refere à presença de palavras indesejadas no texto apresentado pelo navegador. Um valor zero indica rigorosidade máxima.

Em nossa análise, o agente foi capaz de identificar sites pornográficos, que utilizam a linguagem portuguesa, com um grau de certeza de 92%. Esse valor percentual tende a aumentar na medida em que o agente aprende novas palavras.

5 CONCLUSÕES

O trabalho aqui exposto pretende apresentar uma reunião de conhecimentos multidisciplinares com um único objetivo: detecção de assuntos indesejados vindos através da Rede Mundial. Desta forma, se faz possível brindar, à sociedade, uma solução eficaz no combate a temas polêmicos presentes em escolas, universidades e residências.

Com respeito ao software, salientamos que está longe de esgotar o leque de ferramentas e recursos que poderiam ser aplicados objetivando o aprimoramento do mesmo. Podemos mencionar que fazer a análise das imagens visualizadas nos navegadores aumentaria sobremaneira o grau de certeza do agente. Por outro lado, manter uma Base de Conhecimento armazenada em um Provedor de Acesso à Internet diminuiria significativamente o tempo de treinamento e aprendizagem do agente.

6 BIBLIOGRAFIA

1. [COMER96] COMER, Douglas, E. Redes Globales de Información con Internet y TCP/IP 3 ed. México 1996. 621p
2. RUSSELL95] RUSSELL, Stuart, NORVIG, Peter. Artificial Intelligence A Modern Approach. EUA, 1995. 932p.
3. [SOARES97] SOARES, Luiz Fernando, LEMOS, Guido, COLCHER, Sérgio. Redes de Computadores: das Lans, Mans e Wans às redes ATM. Rio de Janeiro, 1997. 705p