



GSA 1001: hardware fabricado pela Dell, software do Google

A CAIXA MÁGICA DO GOOGLE

O Google Search Appliance coloca a agilidade do site de busca a serviço de grandes empresas

→ Já imaginou ter o Google dentro de sua empresa, para que internautas e funcionários localizem as páginas que desejam? Para isso, basta conectar à rede o Google Search Appliance 1001 (GSA). Dentro de um servidor fabricado pela Dell está o motor de busca do Google. O sistema inclui os robôs que analisam o conteúdo e criam o índice a ser usado nas buscas (crawlers) e os mecanismos para avaliar a relevância das páginas, com controles que permitem favorecer determinados resultados. Ferramentas de

administração completam o pacote. Conectamos o GSA à rede do INFOLAB para indexar o site da **INFO** e os documentos do laboratório. Em poucas horas, reconheceu 3 milhões de arquivos, que, depois, puderam ser localizados numa busca simples, com a agilidade habitual do Google.

A facilidade de implantação é um dos pontos fortes. Para começar, basta conectar um cabo Ethernet à máquina e acionar o sistema de gerenciamento via web. Um assistente guia o administrador pelo processo de configuração.

Robôs bisbilhoteiros

Na interface web do GSA há uma área dedicada à configuração dos robôs de busca. Eles indexam pastas compartilhadas na rede, páginas da internet e informações contidas em bancos de dados. O GSA é compatível com os bancos de dados IBM DB2, MySQL, Oracle 10g, Microsoft SQL Server e Sybase. O buscador reconhece 220 tipos de arquivos, de planilhas do Excel a desenhos do AutoCAD. O robô lê os arquivos e gera, para cada um deles, uma página HTML, que é armazenada para acelerar as buscas. É possível controlar a frequência de indexação para que o robô vasculhe mais vezes determinada fonte de dados. O GSA conta, também, com uma interface de entrada de dados chamada Feed, para arquivos ou sistemas não suportados. O formato de entrada é o XML. Há, ainda, o mecanismo OneBox, que permite incluir módulos externos na busca, como um que exhibe informações do Cognos Business Intelligence, da IBM.

Após a indexação dos dados, é a hora de programar os resultados da busca. O algoritmo de relevância do GSA usa regras semelhantes às do Google, inclusive o PageRank, o método de classificação de páginas criado pela empresa. É possível tornar um arquivo ou página mais ou menos relevante, priorizando palavras-chave ou URLs. A empresa também pode remover URLs específicas para que não apareçam nos resultados da busca. Além disso, no caso de documentos armazenados em servidores redundantes, um registro específico impede que os arquivos sejam indexados em duplicata.

Palavras similares

O GSA ajuda os consumidores a encontrar o que procuram mesmo digitando palavras erradas ou similares. Para isso, usa uma biblioteca de sinônimos e corretores ortográficos. Por exemplo, um site de e-commerce pode apresentar

como resultado as páginas contendo a palavra refrigerador, mesmo que o internauta digite geladeira ou "refrigerador". A interface de busca é personalizável. Pode-se, por exemplo, substituir o logotipo do Google pelo da empresa usuária. É possível, ainda, configurar a busca como pública ou privada. Neste último modo, o usuário só enxerga, na página de resultados, os arquivos que ele tem autorização para acessar. O GSA trabalha integrado aos diretórios da rede, usando autenticação centralizada ou senha para acesso.

Tamanho é documento

Um aspecto negativo do GSA é a inexistência de acordo de garantia de serviço (SLA). Se o equipamento quebrar, a empresa ficará nas mãos do Google e da Westcon Brasil, que faz o atendimento no país. O Google sugere que o cliente adquira outro GSA como backup. O preço do GSA varia conforme a capacidade contratada. O modelo básico indexa até 500 mil documentos e custa 220 mil dólares. O que foi testado pelo INFOLAB tem capacidade para 3 milhões de documentos e custa 220 mil dólares. Há, ainda, uma versão que já vem com backup e balanceamento de carga entre servidores incluídos. O servidor de backup avulso custa de 36 mil dólares (para 500 mil documentos) a 53 mil dólares (3 milhões de documentos). São preços que não assustaram clientes como Americanas.com, Pão de Açúcar, Magazine Luiza, Webmotors e o banco Santander. ☒

LUIZ CRUZ E MAX ALBERTO GONZALES



THUNDERSTONE TAMBÉM BUSCA

A caixa amarela do Google não está sozinha no mercado. Há outros equipamentos para buscas como o Thunderstone Search Appliance. Essa caixa azul pode fazer até mil pesquisas por minuto, em intranet ou extranet, reconhecendo cerca de 100 tipos de documentos. Está disponível em modelos com capacidade para indexar de 250 mil até 15 milhões de páginas.

Google Search Appliance 1001	
FABRICANTE	Google
BUSCA	9,5
GERENCIAMENTO	8,5
CONFIGURAÇÃO	7,5
EXTRAS	8,5
PREÇO (R\$)	526 240
AVALIAÇÃO TÉCNICA	9,0
CUSTO/BENEFÍCIO	7,4