

# Web Indexing Overview

## Introduction

The World Wide Web is a part of the Internet that allows clients to publish and consume information quickly, efficiently and easily. The Web defines and assumes a simple, standard communications protocol that allows clients to share a wide variety of data types, such as text, sounds, pictures and even video.

The Web is composed of millions of documents that contain special links which point to other documents. These links are known as *uniform resource locators* or URLs and can point to documents stored on the local computer or any other computer connected to the Internet. There are two types of programs that make the Web work, *web servers* and *web browsers*. A web server runs on a computer and accepts and services requests from web browsers. Web servers don't provide any user interfaces for users, they are the plumbing code that loads documents and hands them off to web browsers. A web browser is a program run by users on their local computer. The web browser interprets URLs within a web document and contacts the web server that stores that document. The browser asks the server for the document and then draws it on the screen for the user.

Because of the ease of use and the rich content the Web provides, traffic on the Internet has increased dramatically over the last several years. It is estimated that almost half the traffic on the entire Internet is now Web related. At the same time the amount of information available on the Web has grown just as fast if not faster..

## Internet Indexing

To address this information explosion and help users find the information they need, a number of *Internet indexing sites* have appeared. Some examples of these sites are <http://www.yahoo.com>, <http://altavista.digital.com> and <http://www.excite.com>. These sites are powerful computer systems that find information on the Internet and index the textual contents. These indexing sites allow users to post queries over the indexes to find sites and documents that discuss the topics they are interested in.

For example, a user might be interested in epilepsy. A query for epilepsy and neural pathways against <http://www.lycos.com> returns the following document from the Web (in addition to over 13,000 others!):

**EPILEPSY RESEARCH, Volume 22, Issue 1 [0.3057, 2 of 3 terms]**

**Outline:** Tables of Contents service for EPILEPSY RESEARCH

**Abstract:** EPILEPSY RESEARCH, Volume 22, Issue 1 Tables of Contents service for EPILEPSY RESEARCH Volume 22, Issue 1, September 1995 Effects of anticonvulsant drug gabapentin on the enzymes in metabolic pathways of glutamate and GABA Goldlust,

<http://www.elsevier.nl/cas/estoc/contents/SA1/09201211/SZ954747.html> (2k)

This type of indexing and search service makes the World Wide Web a far more useful resource by allowing users to find the information they need across all the Web documents. Without an indexing and searching mechanism, users would be forced to make guesses about useful and interesting sites, and then browse the sites, hoping to find references to the information they need.

## Intranet Indexing

As these Internet indexing sites became more popular, other individual sites saw the value of content indexing and searching and began offering the same services over their own local information. The first

services were custom implementations, but soon web server software began providing indexing and search services as another server feature.

Many corporations are now creating internal webs to provide documents to their employees with the same ease of use as the World Wide Web. Since corporations produce huge amounts of information, they also need to take advantage of indexing and searching technology to help their users find the right documents quickly.

## **Index Server**

Index Server is Microsoft's content indexing and searching solution for Microsoft Internet Information Server (IIS). It indexes the full text and properties of documents stored on an IIS server. It allows the administrator (or any user familiar with web documents) to create simple query forms. Index Server can be used to provide indexing services for both corporate intranets, and for sites on the Internet.