Instantly Search Terabytes of Text

dtSearch®

Desktop with Spider

Network with Spider

Publish (portable media)

Web with Spider

Engine for Linux

Engine for Win & .NET

includes 64-bit versions

dtSearch® Features

The Smart Choice for Text Retrieval® since 1991

dtSearch Product Line Overview

dtSearch®

Desktop with Spider

Network with Spider

Publish (portable media)

Web with Spider

Engine for Linux

Engine for Win & .NET

includes 64-bit versions

The dtSearch product line can instantly search terabytes of text across a desktop, network, Internet or Intranet site.

dtSearch products also serve as tools for publishing, with instant text searching, large document collections to Web sites or CD/DVDs.

dtSearch products generally share the following feature set:

- over two dozen indexed, unindexed, full-text and fielded data search options
- display of HTML, XML and PDF files with highlighted hits and with embedded images, links and formatting intact

- built-in HTML converters for hit-highlighted display of MS Office, OpenOffice, ZIP, .MSG, Unicode and other popular file types
- extensive distributed search capabilities, including integrated searching of local and remote content, and Spider support for static and dynamic content (ASP.NET, MS CMS, SharePoint,® etc.)
- multiple developer options for full-text and fielded data classification, as well as special options for searching XML, SQL and other database formats

dtSearch Desktop with Spider, dtSearch Network with Spider

dtSearch Desktop with Spider and dtSearch Network with Spider instantly search popular file types on a PC or a network.

Both products index and search (with highlighted hits) a wide variety of local and network content, including PDF, HTML and XML files; word processor, database, spreadsheet, presentation and similar "Office" files; ZIP files; and email messages (Outlook Express, Outlook, Exchange, Eudora, .MSG and other email

Instantly Search Terabytes of Text

dtSearch Product Line Overview

formats) along with the full-text of attachments.

Through the dtSearch Spider, both applications also support adding Web-based content to a local or networked search.

dtSearch Publish

dtSearch Publish offers easy publishing of an instantly searchable document collection to CD, DVD or other portable media. The product can also mirror an existing Web site on CD/DVD. The resulting CD/DVD application runs with "zero footprint" on the end-user's computer.

dtSearch Web with Spider

dtSearch Web with Spider quickly publishes a large volume of instantly searchable data to an Internet or Intranet site.

Visitors accessing the site can instantly search with hit-highlighted display of Web-ready content (HTML, PDF, XML), including WYSIWYG display of all images, formatting and links.

dtSearch Web automatically converts other file types ("Office," Unicode, ZIP, etc.) to HTML for browser display with highlighted hits, making non-Web-based files "Web-ready."

dtSearch Web installs through an easy wizard-based setup. The application offers numerous customization options, so all aspects of the dtSearch Web display can match the style of the rest of the site.

Operating through dtSearch Web, the Spider can expand the scope of the searchable database beyond a site's own data to static and dynamic content on third-party sites.

dtSearch Text Retrieval Engine for Win & .NET, dtSearch Text Retrieval Engine for Linux

The dtSearch Engine lets developers add dtSearch's built-in file format support and searching to Web-based and other applications.

The dtSearch Engine for Linux offers APIs in C++ and Java. The dtSearch Engine for Win & .NET includes sample code and API support for C++, Delphi,® Java, and .NET—including C#, VB.NET, ASP.NET and ADO.NET. (See over for special database options for SQL and XML, and other fielded data information.) The dtSearch Engine also includes a .NET Spider API.

dtSearch Spider

The dtSearch Spider embedded in multiple dtSearch products provides integrated searching of remote Web site content, along with local data. In addition to support for the file formats above, the dtSearch Spider can also index and search dynamically-generated content, such as ASP/ASP.NET, MS CMS, SharePoint,® etc.

Over two dozen search options

The dtSearch product line generally offers over two dozen indexed, unindexed, fielded and full-text search options.

dtSearch also offers Unicode support for hundreds of international languages, special forensics options, and advanced database search options.

The Smart Choice for Text Retrieval® since 1991

dtSearch® Search Features

How dtSearch Works

- dtSearch can instantly search terabytes of text because it builds a search index that stores the location of words in documents.
- Indexing is easy—simply select folders or entire drives to index and dtSearch does the rest.
- dtSearch automatically recognizes and supports all popular file formats, and never alters original files.
- A single index can hold over a terabyte of text (up to billions of documents).
- dtSearch can also create—and search with a single search request—an unlimited number of indexes.
- Since you may sometimes want to search files that dtSearch has not indexed, dtSearch also does unindexed as well as "combination" searching.
- See over for details on file format support and hit-highlighted display.

Basic Search Types

- **Phrase** searching finds phrases like: *due process of law.*
- **Boolean** operators like and/or/not can join words and phrases: *due process of law and not (equal protection or civil rights).*
- **Proximity** searching finds a word or phrase within "n" words of another word or phrase: *apple pie w/38 peach cobbler.*

- **Directed proximity** searching finds a word or phrase "n" words before another word or phrase: *apple pie pre/38 peach cobbler.*
- **Phonic** searching finds words that sound alike, like *Smythe* in a search for *Smith.*
- **Stemming** finds variations on endings, like *applies, applied, applying* in a search for *apply.*
- **Numeric range** searching finds any number between two numbers, such as between *6* and *36.*
- **Macro** capabilities make it easy to include frequently used items in a search request.
- **Wildcard** support allows *?* to hold a single letter place, and *** to hold multiple letter places: *apple* and not appl?sauce.*

Fuzzy Searching

- Fuzzy searching uses a proprietary algorithm to find search terms even if they are misspelled.
- **Fuzziness adjusts from 0 to 10** so you can fine-tune fuzziness to the level of OCR or typographical errors in your files.
- A search for *alphabet* with a fuzziness of 1 would find *alphaqet*; with a fuzziness of 3, it would find both *alphaqet* and *alpkaqet.*
- Fuzziness is not built into the index, so you can vary fuzziness at the time of each search.

Concept / Synonym / Thesaurus Searching

- Concept searching lets you look for *fast* and find *quick, speedy,* etc.
- dtSearch offers variable levels of **automatic synonym expansion** based on a comprehensive semantic network of the English language.
- You can also add your own thesaurus terms.

Relevancy Ranking and Natural Language Searching

- dtSearch can sort and instantly re-sort searches by relevancy with respect to number of hits, file name, file date, etc.
- **Natural language** algorithms let you enter a "plain English" or unstructured indexed search request.
- **Relevancy ranking** in a natural language search is based on the frequency and density of hits in your files.
- For example, in the search request *get me Sam's memo on the 1999 CorpX takeover,* if *1999* appeared in 3,000 files, and *Sam* appeared in only two files, then *Sam* would get a much higher relevancy rating, taking you straight to the most relevant files.
- A **positional scoring** option ranks documents more highly when hits are near the top of a file, or otherwise clustered in a file.

Instantly Search Terabytes of Text

dtSearch® Search Features

- **Variable term weighting** works with both natural language searching and structured search requests to provide additional positive or negative weighting to user-specified terms.
- Variable term weighting can also apply to document fields.

OCR and Imaging

- dtSearch supports the PDF "image with hidden text" format, and highlights hits right on the scanned image in this format.
- dtSearch also supports combined text and image display in HTML and other Web-based formats.
- dtSearch recommends using fuzzy searching for sifting through possible OCR errors.

Forensics

Special forensically-oriented features include:

- Automatic parsing of text segments in large data blocks, such as those recovered through an "undelete" process, from unallocated computer space, or from partially recovered file fragments
- Language recognition algorithms for detecting text in a variety of languages (Western European, Middle-Eastern, etc.)
- Automatic recognition of dates, email addresses, and credit card numbers
- A proprietary tool for converting Outlook and Exchange message stores, including the full-text of all attachments, to .MSG, for convenient access without requiring Outlook and Exchange running.

International Languages

dtSearch includes support for **hundreds of international languages**.

- Unicode support built into all dtSearch products allows for indexing and searching of non-English text, including every character set, in hundreds of languages supported by the Unicode standard.
- Search options that automatically work on text in any language include: fuzzy (adjustable from 0 to 10); natural language with automatic relevancy-ranking; variable term weighting; phrase; boolean (and/or/not); proximity and directed proximity; wildcard; macro; numeric range; and fielded data (alone or combined with full-text searching).

Language Extension Pack

- The dtSearch product line includes an English noise word list and stemming rules (to find words such as *learn, learned, learns, learning*, etc. that are linguistically related).
- dtSearch's UK distributor offers pre-packaged sets of noise word lists and stemming rules covering over 25 European languages.

- For more information on the Language Extension Pack, please contact dtSearch, or visit www.dtsearch.co.uk.

FindPlus® Distributed Searching

Through FindPlus, a single search request can span multiple local and remote locations.

- For example, a single search request can span multiple indexes residing on a hard drive, a local area network, an Intranet server, and even a publicly-available Web server.
- The same search request can then return comprehensive search results from all locations, ranking all retrieved files by relevance, and instantly re-sorting, for example, by file date.
- FindPlus will display all retrieved files with highlighted hits (as well as HTML, XML and PDF links, images and formatting), even if the files reside on remote servers.
- FindPlus distributed searching uses an XML-based protocol for streaming search results, making it easy for developers to integrate this feature into applications.

Using related technology, the **dtSearch Spider** embedded in multiple dtSearch products supports indexing with instant searching of combined remote Web site and locally available data.

- In addition to PDF, HTML, XML, "Office" documents, ZIP

The Smart Choice for Text Retrieval® since 1991

dtSearch® Search Features

repositories, and the like, the dtSearch Spider can also index and search dynamically-generated content, such as ASP/ASP.NET, MS CMS and SharePoint.®

- The Spider can follow links vertically within a URL, or horizontally across URLs, to any specified level of depth.
- The Spider supports public sites, secure content HTTPS sites, password-accessible sites and forms-based authentication.
- After a search, the Spider provides integrated relevancy-ranking and display of local and Spidered content, including WYSWYG display with highlighted hits of Web-ready content.

Databases and Fielded Data

- dtSearch indexes and searches fields in all supported file types, supporting both full-text searching encompassing fields and search requests limited to specific fields.
- After a search, dtSearch can display field values in search results.
- dtSearch can also index and search popular database formats, including ODBC support and support for BLOB data.
- For XML, dtSearch supports hierarchical field structures in XML data, covering both fields and attributes, enabling highly refined nested field queries.
- For SQL, the dtSearch Engine supports precision fielded data searching through ADO.NET and COM.

Classification and Filters

- dtSearch developer products include extensive support for adding on-the-fly classification information and other fields to documents during indexing.
- dtSearch developer products can filter results based on document full-text contents, document fielded data, database content, or data attributes attached during document indexing.
- Search filter objects allow multiple users with different security classifications to search the same document collection, without having to maintain separate indexes corresponding to each classification level.
- A single query can include an "exact phrase" full-text end-user search request; second-level Boolean search expressions (such as one or more field or metadata search criteria); and developer-added filtering expressions (such as a filtering expression to filter out documents that do not match a user's security classification).

Instantly Search Terabytes of Text

Popular uses of dtSearch products include …

information management: "dtSearch took its file indexing and searching expertise, moved it to the Web, and now offers a variety of tools to help users and companies avoid ever losing a file again on a desktop, a network or the Web" – *Network World*. "dtSearch will become the centerpiece of our digital library" – *case study*

news and real-time data: "The dtSearch APIs were easy to use and enabled us to integrate the product in a real-time environment, where a fraction of a second matters" – *case study*

forensics and intelligence: dtSearch is "so good that even criminal investigators use it for forensics" – *The London Morning Paper.* "The FBI and local law enforcement agencies … have used dtSearch and [its customer's] software to sift through massive amounts of computer data." "Dramatic examples where the product has helped catch criminals" include "a money laundering and black market gem operation" and "a police shooting in California" – *Washington Business Journal*

other government: in the past two years alone, **4 out of 5 of Fortune Magazine's top Aerospace and Defense companies** have purchased dtSearch developer or multi-user licenses. dtSearch provides "the portability, speed and accessibility that our defense security customers require" – *case study*. "The speed at which results are returned is blazing. Thank you for making an excellent product that does an exceptional job at helping our citizens discover content in our document repository" – *case study*

international government: "The multi-language support was also a great benefit, given the international character of our client base" – *case study*. "dtSearch's multilingual support was key for the European Parliament project" – *case study*

technical documentation: "We chose dtSearch to create an advanced search portal for the DOD document library" – *case study.* "Aviation firm earns wings with new text search," using "dtSearch Web, to serve the information up in an easily indexed manner online" – *Computerworld Canada*

financial and accounting: in the past two years alone, dtSearch has sold multi-user or developer licenses to **3 out of 4 of the "Big 4" accounting firms**

email filtering and archiving: searching of inbound, outbound, and archived emails (along with the full text of attachments). "Fishing expeditions are easier with features such as fuzzy searches, which can find words even when they are misspelled, and a built-in thesaurus … Functioning equally well against both email folders and text files, these are very powerful and useful capabilities for an investigator" – *Computer Forensics: Incident Response Essentials*

medical: "PET data are extracted from the NUC-NET database using dtSearch" – *Newsline, The Journal of Nuclear Medicine*
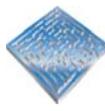
legal and compliance: "There is no quicker way to find a needle in a digital haystack" – *The Lawyer's PC*. dtSearch "raises the bar in text retrieval, providing an industrial strength search engine that is an absolute must-have for Web-based data information management … dtSearch outmaneuvers the competition. It's a robust, sophisticated and powerful search tool … If you want speed and accuracy, dtSearch is the only way to go" – *Law Office Computing*

recruiting and staffing: "The result is very fast online searching of our resume database" – *case study*

other software: "Our users have been begging for this functionality for years" – *case study*. "The search features … have been well-received and widely praised by customers, partners and pundits" – *case study*. "Customer response has been phenomenal" – *case study.* "The door of my office is plastered with printouts of positive feedback" – *case study*

For hundreds of developer case studies and press reviews, please see www.dtsearch.com

Instantly Search Terabytes of Text

The Smart Choice for Text Retrieval® since 1991, dtSearch® offers more than a decade and a half in text search and retrieval.

To download **fully-functional evaluations** of any products, or for additional information, please visit www.dtsearch.com, email sales@dtsearch.com, or call 1-800-IT-FINDS (US/Canada) or (301) 263-0731.

To visit the dtSearch **online store**, please go to www.dtsearch.com/dtStore.html. For a complete developer price list, please email sales@dtsearch.com.

dtSearch also has **distributors worldwide**, including coverage on six continents. Please see www.dtsearch.com/distributors.html.

dtSearch®

www.dtsearch.com • 1-800-IT-FINDS • 1-301-263-0731