

Leveraging Online Data With Full-Text Search

A web portal can include terabytes of data. But posting data and enabling users to harness it are two separate things. A full-text search engine lets users effectively leverage the data.

To implement full-text search online, you'll need a search engine like dtSearch. dtSearch Web can run as an "off the shelf" (no-programming required) application on a Windows IIS web server. The server itself can reside "on premises" or on a hosting platform like Azure or AWS. The dtSearch Engine SDK can run not only on a Windows web server but also a Linux web server—or even a macOS server.

A search option can fit seamlessly into the "look and feel" of a site. dtSearch Web has sample HTML5 forms that are easily customizable. For developers, the dtSearch Engine is an option as well, particularly if the developer wants to integrate search into a broader application or add on a backend database like SharePoint or SQL/NoSQL. The dtSearch Engine can index both database data and any referenced files or embedded BLOB data files.

dtSearch instantly searches terabytes by first building an index across the dataset. Indexing is easy. Just point to the relevant data and dtSearch will do the rest. No need to even tell dtSearch what type of content it is indexing; dtSearch will figure that out for itself. An index can hold up to a terabyte, and there are no limits on the number of indexes that dtSearch can build and simultaneously search.

Both dtSearch Web and the dtSearch Engine enable concurrent search across the indexed dataset, with each search request running in a stateless manner on a different thread. That way, each search thread can deliver instant searching without affecting other search threads. For international-oriented sites, searching works with not only English text, but also any of the hundreds of Unicode languages, including right-to-left languages like Hebrew/Arabic and double-byte Chinese/Japanese/Korean ("CJK") text.

Following a search, dtSearch provides multiple options for sorting and instantly re-sorting search results. Users can browse a complete copy of each retrieved item with highlighted hits. For online search, a caching option can store a full copy of the original data along with the index. Caching will make display of the data with highlighted hits a lot "snappier" versus having dtSearch return to the original data if it resides in a different remote location.

Article contributed
by [dtSearch®](#)



**Posting data
and enabling
users to harness
it are two
separate things.**

And now the really important part: dtSearch Corp. will not see any search requests and the like. dtSearch products do not send information back to the "Mother Ship." For that reason, dtSearch customers use dtSearch products not only with publicly available online data repositories, but in high-security situations too. (You can see some of the different vertical markets dtSearch use spans at dtSearch.com.)

If you want to take searching offline, dtSearch Desktop and dtSearch Network work on individual PCs or in a classic Windows network environment to instantly search terabytes. All dtSearch products support popular file types like PDF, Microsoft Word, Access, Excel, PowerPoint and OneNote. The products work with email formats like PST, OST, MSG and EML including the full text of all attachments. And of course the products work with online formats like HTML, XML, etc.

Multilayer nested files are not a problem. dtSearch can automatically go through compressed archives like ZIP or RAR, and index all files in the archive. If an email has a ZIP attachment containing a Word document and embedded in the Word document is an Excel file, dtSearch can nonetheless index and search everything.

To accommodate data updates, dtSearch can reindex just those files that have been added, deleted or modified since the last index build. In fact, dtSearch lets you update indexes automatically as often as you want using the Windows Task Scheduler. Updating an index does not lock out searching so even concurrent searching can proceed while an index updates.

dtSearch products have over 25 different full-text and metadata search features. Use Boolean and proximity searching for precision search requests. Add on concept searching to find synonyms. Or add on fuzzy searching to sift through minor typographical and OCR errors as often appear in emails and PDFs. Advanced users can check out the Features Map at dtSearch.com to find even more search options like user-defined metadata-specific variable term relevancy-ranking, generating and searching for hash values, or the flagging of any credit card numbers in data.

dtSearch has enterprise and developer products that run "on premises" or on cloud platforms to instantly search terabytes of "Office" files, PDFs, emails along with nested attachments, databases and online data. Because dtSearch can instantly search terabytes with over 25 precision search options, many dtSearch customers are Fortune 100 companies and government agencies. But anyone with lots of data to sift through can download a fully-functional 30-day evaluation copy from dtSearch.com

Article contributed
by [dtSearch®](http://dtSearch.com)



**dtSearch Web
has sample
HTML5 forms
that are easily
customizable.**