Your Boss May Be Ambivalent About Hybrid Work, but Enterprise Search Is Not

Article contributed by dtSearch®

Recently, there's been a number of stories about bosses incentivizing employees to return to work or forcing workers to return to the office on the theory that remote work is less productive. Enterprise search offers a counterargument to that. Intranet-based enterprise search instantly combs through terabytes of organizational data for maximal employee productivity, whether workers log in from the office or Tahiti.
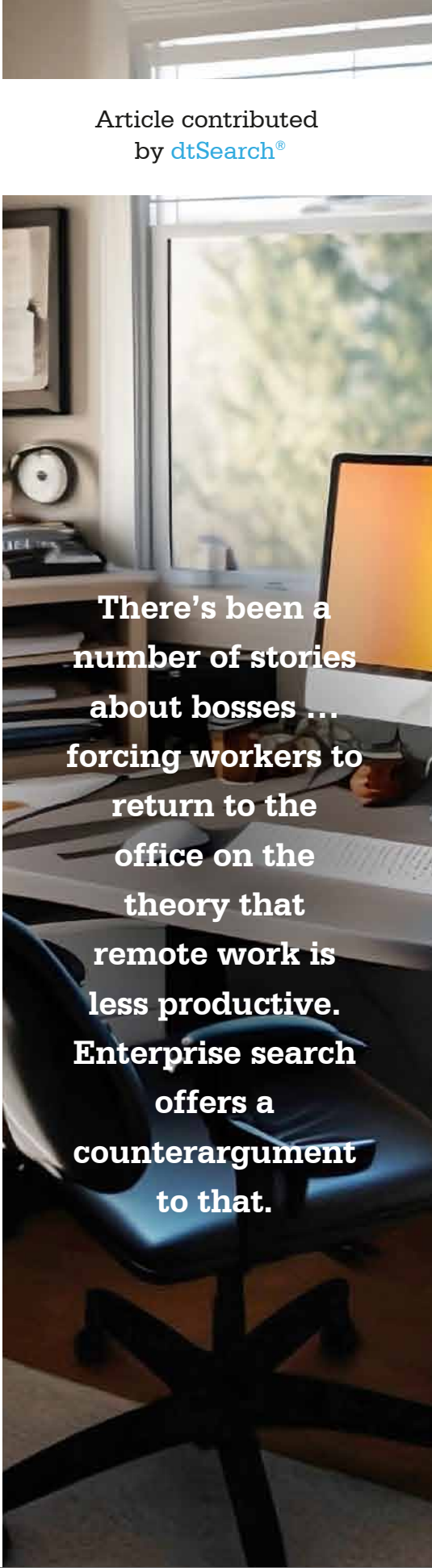
How does enterprise search work?

An example of enterprise search would be dtSearch® which 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. But enterprise search can only instantly search terabytes after first indexing the data. Indexing is easy. Just tell the indexer what folders, email archives and the like to cover, and the search engine can take it from there. Files can be local or cloud-based like Office365 or SharePoint, so long as these appear as part of the Windows folder system. For efficiency, the indexer works with the binary format of files rather than pulling up each in its associated application.

What are the binary formats like?

Binary formats typically present as a jumble of binary codes. Looking at the binary format of most current document and email files, you'd be hard-pressed to discern any text at all, much less read complete sentences. The first step for enterprise search is to identify the relevant file format to apply the exact right parsing specification. As each parsing specification can be hundreds of pages long, applying the correct one is critical for the accurate identification of all text and metadata.

> There's been a number of stories about bosses … forcing workers to return to the office on the theory that remote work is less productive. Enterprise search offers a counterargument to that.

How does enterprise search identify the correct file format?

dtSearch will use the binary format itself to identify the file format rather than the file extension, as it is all too possible to have a PDF with a .DOCX extension or an Access database with a .ONE extension. Enterprise search also needs to work with recursively nested items, like an email with a RAR or ZIP attachment containing a PowerPoint with an Excel spreadsheet embedded inside. Again, the binary format tells the whole story. After identifying the text and metadata, enterprise search can start indexing. The final index holds each unique word and number across all data and the location of each in the data.

What happens after indexing?

After indexing, anyone, whether on-premises or in Tahiti, can instantly search across terabytes. Intranet-based queries can run in a completely stateless manner, with each search thread proceeding independently. This makes concurrent search easily scalable, with no built-in limits on the number of simultaneously executing, instant search threads. Following a search, for convenient browsing of retrieved items, search results can display a complete copy of each item with highlighted hits, regardless of whether the person searching is in-office or remote.
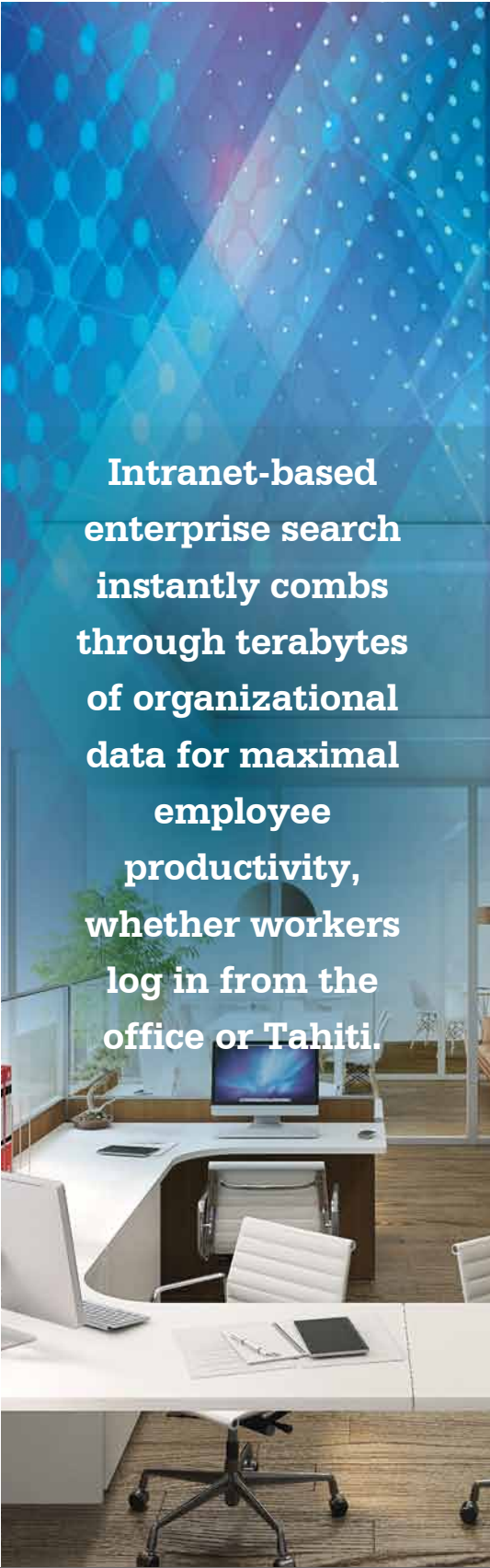
What types of search options are there?

Searchers can pick and choose from over 25 different search options. A basic natural language "all words" search for *flextime productivity matrix* would look for any document, email or similar item that contains all 3 of these words. An "any words" search request would retrieve any item matching even one of these terms.

What about more precision, structured search requests?

Boolean search looks for specific and/or/not combinations of words or phrases, like *(summer flextime or productivity matrix)* and *(final memo and not draft)*. Proximity searching requires a word or phrase within X words of another word or phrase in either direction or just before the other word or phrase. Search requests can also specify that certain terms

Article contributed by dtSearch®

Intranet-based enterprise search instantly combs through terabytes of organizational data for maximal employee productivity, whether workers log in from the office or Tahiti.

appear in particular metadata, like limiting a search to only emails with *productivity matrix* in subject metadata.

Any other search options?

Stemming looks for different endings on the same root word, like *studying* and *studied* for *study*. Wildcards can cover one or more missing characters so *covid** would find *covid* as well as *covid19*. Concept searching finds general synonyms like *plague* for *disease*. Concept search can also work with custom synonym rings. If *matrixABC* is renamed *matrixDEF*, these could be custom synonyms, with a search for one extending automatically to the other. Fuzzy searching adjusts from 1 to 10 to look for typographical deviations, like a mistype of *productimity* for *productivity* in an email or a similar OCR error.

What about relevancy-ranking?

By default, dtSearch will apply vector-space relevancy ranking. If *memo* is in millions of documents but *productivity* appears in only a few, then *productivity* would get a higher relevancy ranking than *memo*, with denser *productivity* mentions getting the highest ranking. Or a search can apply custom positive or negative variable term weighting regardless of the prevalence of indexed data terms, with an optional greater weight for specific terms in certain metadata or positionally at the top or bottom of a file.

Any other search options?

dtSearch also supports date and date range queries across different date formats. A date range of *June 20, 2023* to *July 15, 3023* would pick up both *July 6, 2023* and *7/6/23* in the full text and metadata. Search can further specify numbers or numeric ranges, and can even identify certain number sequences like flagging any credit card numbers that may appear in data. For an international hybrid work world, enterprise search supports Unicode, covering not only European-origin languages but also right-to-left text like Arabic and Hebrew and double-byte Asian characters like Chinese, Japanese and Korean.

Article contributed
by dtSearch®

For an international hybrid work world, enterprise search supports Unicode, covering not only European-origin languages but also right-to-left text like Arabic and Hebrew and double-byte Asian characters like Chinese, Japanese and Korean.