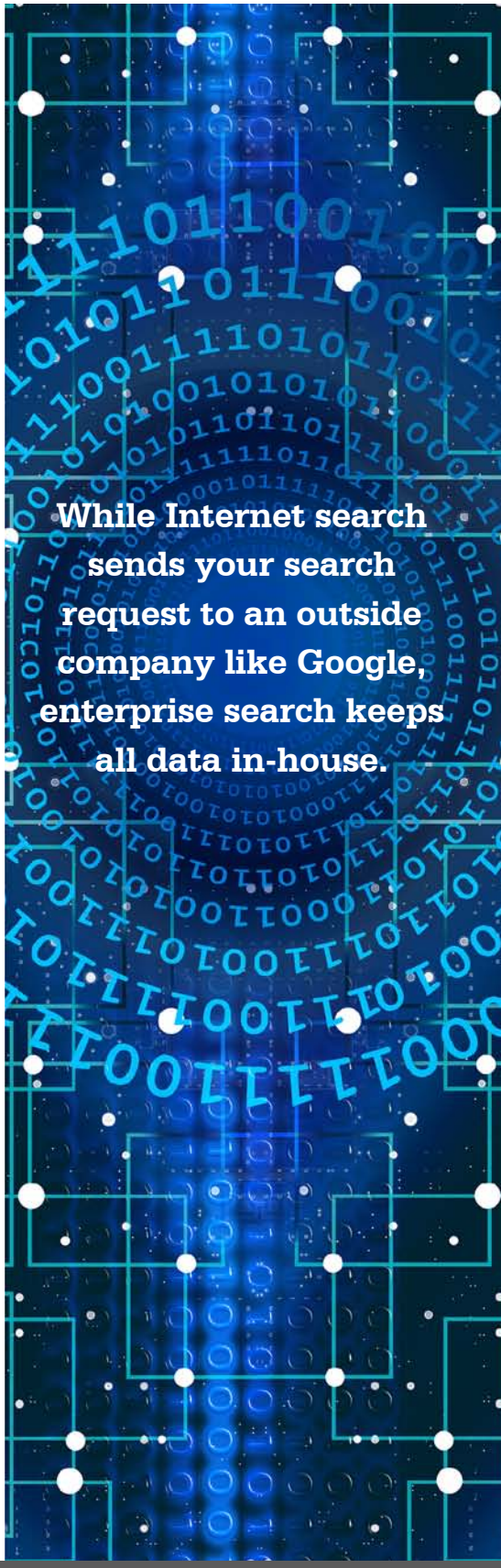
A Cornucopia Of Data To Sift Through?

Article contributed by dtSearch®

Let's say you want to know whether it's likely to rain during your afternoon outing. A simple yes/no or even a less determinate answer like a 40% probability can get you going with appropriate protection. But now let's say instead of a straightforward weather forecast, you get terabytes of historical and present-day weather data and analysis dumped on you. Not so helpful as you are running out the door late.

Same thing with enterprise data. It's great that your office has terabytes of records sitting on its local and remote servers, but how are you going to sift through that cornucopia to find what you need? That's where enterprise search comes in, letting you find the needle in the haystack. (Doesn't autumn offer a cornucopia of data analogies?)

Enterprise search provides a similar search mechanism to Internet search across your own office content. But while Internet search sends your search request to an outside company like Google, enterprise search keeps all data in-house. (Note: this article's descriptions use dtSearch® for its discussion of functionality. If you are using a different search engine, please first verify how it works.)



Enterprise search instantly searches terabytes after first building one or more search indexes. Each index can hold up to a terabyte of data, and there are no limits on the number of indexes that enterprise search can build and simultaneously search. So is indexing a lot of work? Not for you; all you need to do is point to the folders and the like to index. Enterprise search automatically recognizes popular file types like PDF, Microsoft Word, Excel, Access, PowerPoint, OneNote, web-based formats, email formats, etc.

Indexing aims to be foolproof. The indexer can take a PDF saved with a .DOCX file extension or a Word document saved with a .PDF file extension and handle that correctly. Indexing can further work with emails plus multilayer nested attachments, like a ZIP or RAR attachment holding a Word document that itself embeds an Excel spreadsheet. Remote data like Office 365 files or SharePoint files are fine so long as these present as part of the Windows file system. And content that may be invisible in its native application, like black text against a black background, is just ordinary text to the indexer.

Reprinted with permission of Young Upstarts, youngupstarts.com

Multilanguage data is also not an issue. Enterprise search supports Unicode covering hundreds of international languages, including European languages, right-to-left Hebrew and Arabic, and double-byte character Chinese, Japanese and Korean. A single file or email can have multiple different Unicode encodings, and enterprise search will track that progression. Enterprise search can even find Unicode emojis 🐸.

After indexing, enterprise search lets you instantly search across the indexed data cornucopia. Or be the OFFICE SUPERHERO and extend searching, operating in an Intranet or a classic network environment, to all of your co-workers, letting everyone instantly search at once. While indexing is resource-intensive, searching is resource-light, allowing search threads to operate instantly and concurrently without interfering with each other. When data updates, enterprise search can update its indexes to reflect files that have been added, deleted or modified without affecting continuing instant concurrent search.

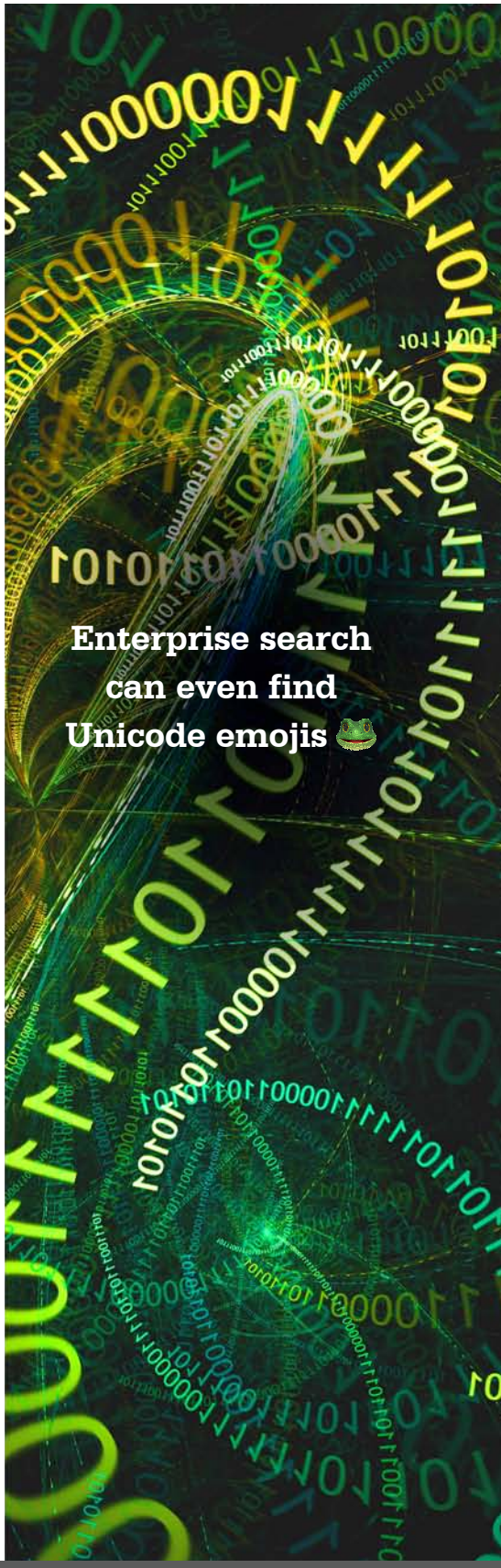
Enterprise search offers over 25 different full-text and metadata search types so everyone can get to the right information quickly. Enter a simple unstructured natural language query: *holiday party beverage order.* Or leverage precision word and phrase Boolean (and/or/not), proximity, data range, etc. search elements: *"holiday party" and (beverage w/7 soda) and not (coffee w/18 tea) and date(10/31/23 to 1/31/24)*.

Date range search can automatically pick up different date formats like *11/15/23* and *Nov 15, 2023*. Concept searching can find *drink* for *beverage*. If *holiday party* is mistyped as *holimay party*, a low-level fuzzy search can still flag that. Or add metadata-specific search elements to your search request. Advanced search options can even find stray credit card numbers that may have snuck their way into open office data.

After a search, view retrieved items with highlighted hits for convenient navigation. "Vector-space" relevancy ranking takes you right to the key files. Take a natural language *beverage order* query. If *order* is in 2 billion documents but *beverage* is in just a few dozen, *beverage* files would get a higher ranking, with densest mentions getting the highest rankings. You can also add your own customer variable term weighting, giving a higher positive or negative weight to full-text and/or metadata content or positionally to content at the top or bottom of a file. Or instantly re-sort by some unrelated criterion like file date or file location.

Drowning in a cornucopia of data no more. Now you can get out and enjoy the holidays.

Article contributed by dtSearch®



Reprinted with permission of Young Upstarts, youngupstarts.com