

BRAINWARE GLOBALBRAIN™

Desktop and Enterprise Search

More than a decade ago, a group of scientists with advanced degrees in physics, mathematics and chaos theory set their sights on the mountains of unstructured data they saw being produced and reproduced by enterprises around the world. Convinced there had to be a better way to access full document text than through keyword and metadata indexes, they invented the Globalbrain search engine.

SIMPLE ENTERPRISE FINDABILITY

Globalbrain is highly scalable and enables access to vast amounts of structured and unstructured data (including e-mails, e-mail attachments, OCR-ed images, PDF files, word processing documents, spreadsheets and hundreds of other formats). Simply search for information by using queries comprised of phrases, sentences, paragraphs or even entire documents of text rather than complex Boolean logic or complicated taxonomies.

It's simple enough to run searches on your personal laptop and yet powerful enough to execute queries against disparate, multi-terabyte repositories of data located around the globe in a matter of seconds.

Access vast amounts of structured and unstructured data including:

- | e-mails
- | e-mail attachments
- | OCR-ed images
- | PDF files
- | word processing documents
- | spreadsheets
- | and hundreds of other formats

CAPABILITY	GLOBALBRAIN	OTHER SEARCH
Language independent - no dictionaries	<input checked="" type="checkbox"/>	
Returns relevant results even with spelling errors	<input checked="" type="checkbox"/>	
Query words, phrases, multiple paragraphs of text	<input checked="" type="checkbox"/>	
Processes extremely high volumes at very high speeds	<input checked="" type="checkbox"/>	
Based on statistical engine	<input checked="" type="checkbox"/>	
No taxonomies required	<input checked="" type="checkbox"/>	
Natural language searching	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>

IT'S ABOUT MATH

Globalbrain employs an associative access or "n-gram" approach to search. This approach takes out the human element, simply turning words into mathematical representations. For example, when tri-grams (three-letter components) are used to index text, words are parsed into three-letter parts. For example, the word "sample" would be parsed as "sam", "amp", "mpl", etc.

The search query is then parsed in the same manner and Globalbrain performs comparisons of its three-letter snippets to those in the index using a matrix of ones and zeros. Finally, it rates the possible hits based on matches to those three-letter groups.

This practice is also known as "fuzzy search" or "fuzzy logic" because it's not a rifle shot to an exact match but instead returns results that allow users to choose which ones are most relevant to them. Globalbrain does not try to derive meaning, nor does it care who, what or why you are searching... it just helps you find what you are looking for as quickly as possible.

Globalbrain is extremely successful at handling misspellings, uncommon terms and very large amounts of text in a search query and results are returned with lightning speed. In the case of the tri-gram, the knowledge base you are searching against has already been indexed into a matrix, so your query string simply needs to be turned into ones and zeroes and quickly matched against the existing data. This is much faster than checking against multi-lingual and multi-variant dictionaries or keywords.

BALANCING PRECISION AND RECALL

Because Globalbrain determines both how close your request is to the indexed data and how far away it is, you have the option of refining the relevance of what gets returned. If you are looking for as near a match to the query you entered as possible, keep the relevancy high at 90%. If you need to find as much information as possible about the subject matter, lower the relevancy so you can also return similar but not exact matches (for example, as in the case of e-discovery).

Unlike traditional keyword search, Globalbrain performs best when it's presented with more information in the query string rather than less. A 20-page search query would normally paralyze a typical search solution, yet Globalbrain can handle it with great speed and accuracy. Because the search query is converted from words to n-grams, your searches are language independent. There is no need to filter for language at any level.

LARGE REPOSITORIES - SMALL INDEXES

Globalbrain finds indexed information - whether it resides on corporate file servers, email servers or websites. Because the data indexed is now binary text, the index footprint is very small - usually only 10% of the original document repository. For example, an indexed 200GB repository results in a knowledge base of about 20GB of searchable text.

EASILY EMBEDS IN THIRD-PARTY APPLICATIONS

Although Globalbrain can be run out of the box with its own Web interface, there are Java, COM, SOAP and .NET APIs available so that the search and classification functionality of the application can be added to existing applications or Web portals.

EASE OF ADMINISTRATION

Globalbrain can be implemented in a matter of days. Updates (i.e., document additions, deletions and alterations) are easily accommodated without having to re-index data.

www.brainware.com | info@brainware.com | US +1.703.948.5800 | UK +44 (0)1159 079 300 | © 2009 Brainware, Inc. All rights reserved. Brainware, Brainware Distiller and Globalbrain are trademarks of Brainware, Inc. and/or its licensors. All other trademarks are the property of their respective owners.

