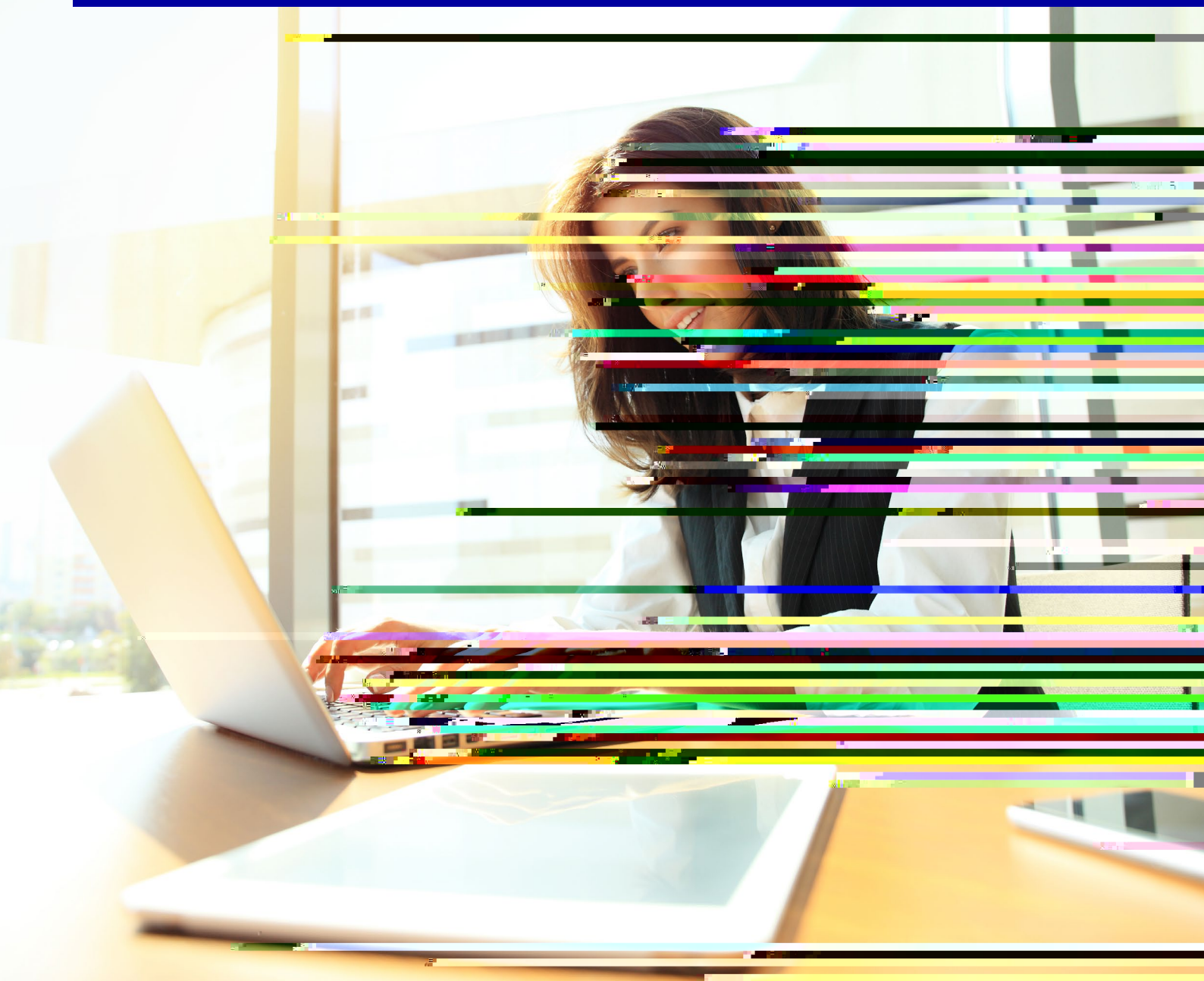
opentext™

WHITE PAPER

OpenText Magellan Text Mining helps users gain insight from unstructured content

How to uncover insights and information that optimize organizations' content.



opentext™

The following discussion of how Magellan Text Mining works uses terms that are common in the field of content analytics but may not be familiar to all readers. For example, "text mining" and "text analytics" are sometimes used interchangeably. Although they are both methods of extracting valuable information from written text, they are not quite the same thing. Other definitions are in the box below.

Name	Definition
Text mining	The process of deriving high-quality information and relationships from textual

opentext™

Creating and using semantic metadata

Assigning semantic metadata throughout a text or group of texts is one of the fundamental techniques Magellan Text Mining uses to derive useful insights and help users connect with the content that's most relevant to them. For easier reading, we will simply say "metadata" here, keeping in mind that non-semantic types of metadata are also relevant from a broader analytics perspective. The metadata can be classified and searched, quickly highlighting aspects of the text that humans could only find by reading carefully. Once metadata is added to a document, useful relationships can be built. For example, a user can create the list of all published documents for a given author and, in turn, retrieve all documents by that author.

Users routinely create metadata manually during the process of authoring, editing and sharing content. A basic example is when an author creates a document and gives it a title. Or a user can manually add key terms associated with a document in an Enterprise Content Management system, perhaps for search engine optimization (SEO) or easier retrieval. However, manually creating metadata is very time-consuming and not always on-target, so organizations seek text analysis tools that can automate the process to make it faster and smarter.

Magellan Text Mining can assign metadata to a document either semi-automatically or completely automatically, depending on which mode the user prefers.

Automated metadata assignment

In completely automated mode, Magellan Text Mining can automatically process textual data, storing the metadata so it can be used as is, without necessary revision.

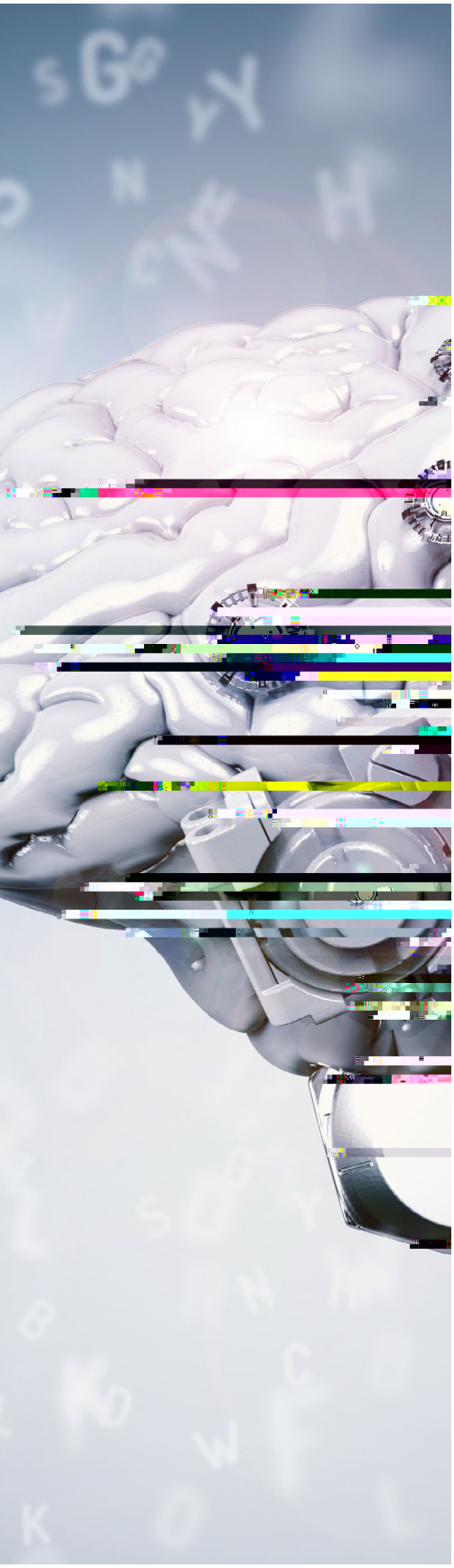
For instance, it could verify that documents or emails do not contain specific personally identifiable information (PII), such as full legal names and credit card purchase histories.

Combining different semantic metadata is an excellent way to improve search efficiency. For example, a user may only be interested in content that was authored by Mary J. Jones, contains a credit card number and mentions OpenText. Having all this metadata ahead of time reduces the number of documents searched, delivering the information that users need much faster.

Semi-automated metadata assignment

In semi-automatic assignment, Magellan Text Mining can spot and recommend important keywords for each document, which can then be reviewed by the author or a content manager. For example, Magellan Text Mining could suggest what key phrases constitute variations on entity names, such as New York City, the Big Apple, Manhattan or New York. Then, a reviewer can select what is or is not a relevant organization to extract.

During the installation of Magellan Text Mining, an organization can configure its preferences for fully or semi-automated metadata assignment depending on the use case. Fully automated metadata assignment is better for tasks that require highly



Concept Extraction

Concept Extraction identifies meaningful information from documents using special grammatical, pattern-based algorithms to extract key concepts.

Concepts: A keyword (simple concept) or key phrase (complex concept) found within text. For example, 'financial statement' would be a key phrase, while 'goal'

Conclusion