

BUZZWORD DEMYSTIFIER

Text mining

Information often lies, unstructured and inaccessible, in disparate document formats. Data mining can open up this valuable seam of data and deliver valuable business intelligence from it.



Raoul Jetley
ABB Corporate Research
Bangalore, India
raoul.jetley@in.abb.com

It is estimated that up to 80 percent of all information in organizations is stored in an unstructured text format. This information includes customer requirements, sales dossiers, technical specifications, maintenance reports and stakeholder feedback. It is difficult to extract business intelligence from such disparate data using traditional data analysis methods so, instead, text-based data mining, or text mining, is used →1.

Simply put, text mining is the set of processes required to transform unstructured text documents or resources into meaningful, structured information. The structured information can then be used to automatically discover hidden patterns and predict future outcomes using a combination of statistical, linguistic and pattern-recognition techniques.

01 Large amounts of valuable data can be hidden in company documentation. Data mining can bring it to light.

02 Data mining can deliver important, hitherto unavailable, insights for sales and marketing, business decision making, investment or purchase decisions, customer relationships, etc.

02



01

Text mining is an interdisciplinary field that draws on information retrieval, data mining, machine learning, statistics and computational linguistics. These techniques are used to discover and present knowledge – facts, business rules and relationships – that is otherwise locked in textual form, impenetrable to automated processing.





searching through the tokenized text and storing the results in a more structured, organized manner that is amenable to further analyses.

- Select important features to build concept and category models. The number of concepts present in unstructured data is typically very large. The key to this step is to identify the most relevant features and use these to build meaningful models based on data categories and relationships.
- Analyze the structured data to discover relationships between the concepts. At this point, the text mining process merges with the traditional data mining process. Classic data mining techniques, such as clustering, prediction and classification can be used on the structured data resulting from the previous steps.

Text mining is the set of processes required to transform unstructured text documents or resources into meaningful, structured information.

Common applications resulting from these analyses include recognition of named entities, automatic summarization, categorization based on relevant features, and mining for customer sentiments and opinions expressed within the text →2. ●

A typical text mining process includes the following steps:

- Identify and preprocess the text to be mined. This step involves text clean-up to remove unnecessary information from the text, splitting the text into individual tokens (ie, smaller components) and identifying parts-of-speech based on the grammar of the language used.
- Extract relevant information and transform it into structured data. Information is retrieved by

Imprint

Editorial Board

Bazmi Husain
Chief Technology Officer
Group R&D and Technology

Adrienne Williams
Senior Sustainability
Advisor

Christoph Sieder
Head of Corporate
Communications

Reiner Schoenrock
Technology and Innovation
Communications

Roland Weiss
R&D Strategy Manager
Group R&D and Technology

Andreas Moglestue
Chief Editor, ABB Review
andreas.moglestue@
ch.abb.com

Publisher

ABB Review is published
by ABB Group R&D and
Technology.

ABB Switzerland Ltd.
ABB Review
Segelhofstrasse 1K
CH-5405 Baden-Daettwil
Switzerland
abb.review@ch.abb.com

ABB Review is published
four times a year in
English, French,
German and Spanish.
ABB Review is free
of charge to those
with an interest in
ABB's technology
and objectives.

For a subscription, please
contact your nearest
ABB representative or
subscribe online at
www.abb.com/abbreview

Partial reprints or
reproductions are
permitted subject to
full acknowledgement.
Complete reprints
require the publisher's
written consent.

Publisher and
copyright ©2018
ABB Switzerland Ltd.
Baden/Switzerland

Printer

Vorarlberger
Verlagsanstalt GmbH
6850 Dornbirn/Austria

Layout

DAVILLA AG
Zurich/Switzerland

Artwork

Konica Minolta
Marketing Services
WC1V 7PB London,
United Kingdom

Disclaimer

The information
contained herein
reflects the views of
the authors and is for
informational purposes
only. Readers should
not act upon the
information contained
herein without seeking
professional advice.
We make publications
available with the
understanding that the
authors are not rendering
technical or other
professional advice or
opinions on specific facts
or matters and assume
no liability whatsoever in
connection with their use.

The companies of the ABB
Group do not make any
warranty or guarantee,
or promise, expressed or
implied, concerning the
content or accuracy of the
views expressed herein.

ISSN: 1013-3119

abb.com/abbreview



Tablet version

Production of the tablet version of ABB Review (for both iOS and Android) is being discontinued at the end of 2018. Readers of the tablet versions are recommended to use the pdf or web versions instead.

abb.com/abbreview

Stay informed...

Have you ever missed a copy of ABB Review?
Sign up for the e-mail alert at abb.com/abbreview
and never miss another edition.

Please note that when you register for this alert, you will receive an e-mail with a confirmation link. Please ensure that you have confirmed your registration.

