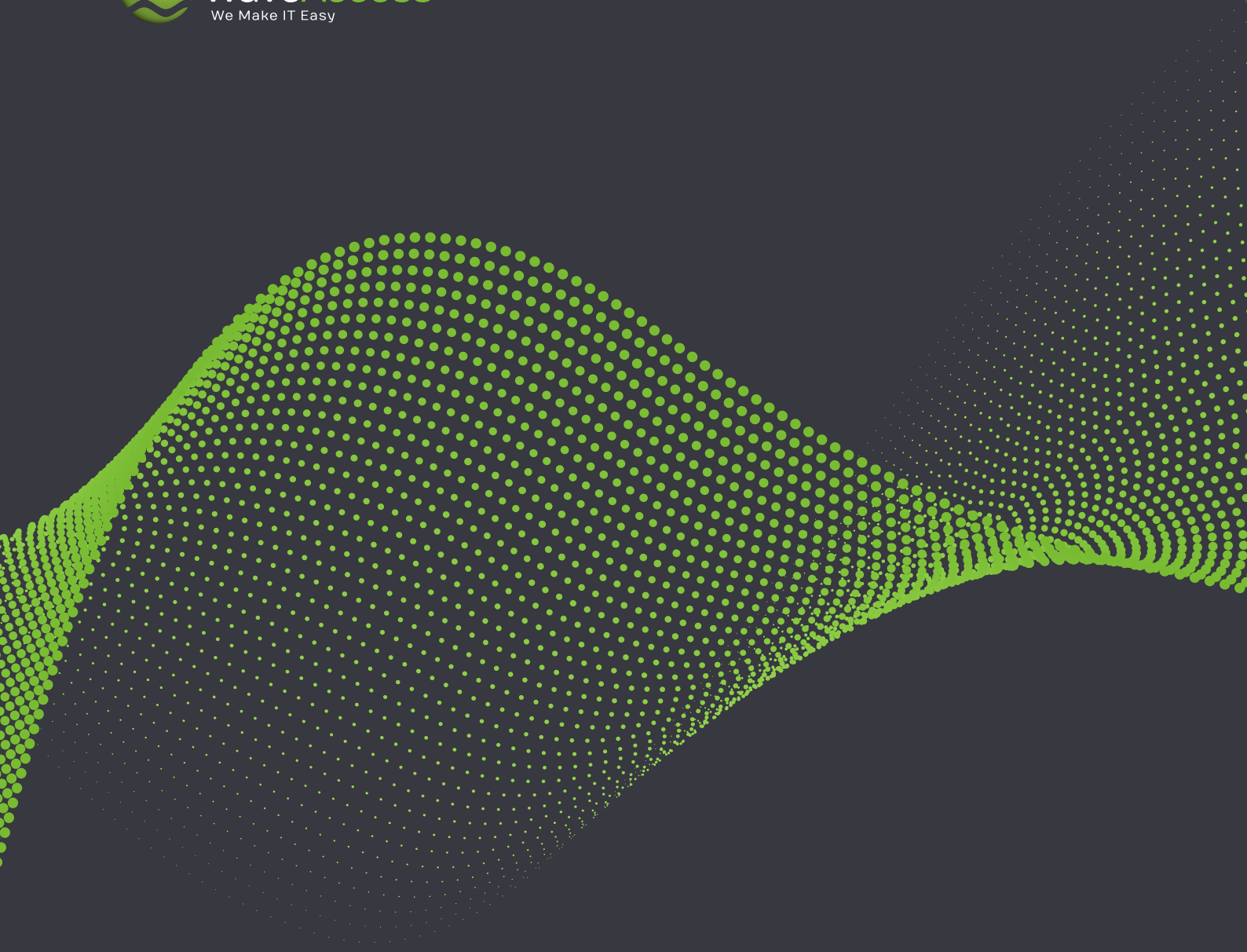
WaveAccess
We Make IT Easy



Text Mining:

Semantic Text Analysis based on Azure Machine Learning algorithms

Content

Text mining is an automated process of deriving high-quality information from text. What makes it different from other types of data analysis is that the input data is not formalized in any way, which means it cannot be described with a simple mathematical function.

WaveAccess offers a Text Mining solution based on Azure ML algorithms to analyze structured and unstructured text data. The solution is a Text Mining module that we embed in your current business application (your CRM, ERP, doc management system, knowledge DB etc).

The solution can be used for analytical purposes like

- Data search in unstructured arrays
- Text topics extraction
- Facts extraction
- Text semantic (meanings) extraction
  - summarization as extracting
  - summarization as abstracting
- Keywords and named entities extraction
- Categorization
- Multiple text metrics evaluation
- Sentiments analysis (Empathy, Dissatisfaction, Satisfaction categories etc)

Business value of text mining

- Saves time and labor costs by automating manual work
- Provides new analytical insights: what was said (written) and how
- Ensures services based on analyzing a large amount of text data
- Mitigates risks of not receiving crucial information at the right time
- Helps to make smart decisions on what is fraud, what is risky, and what is relevant for business

Use cases

A knowledge base for the company can be created based on the accumulated text data

Example

Our client, a large pharmaceutical company from the US, used our text mining solution which dramatically improved the efforts of finding the right medicine feedback and reviews in the sea of unnecessary information.

Also, the text mining solution was used to analyze decades of research in medicine science, volumes of clinical patient data – it helped to extract the essence from huge amounts of unstructured data for better analytical insights.

Quick acquisition of relevant data from databases, to speed up workflows and management

Example

The customer gets a great number of questions regarding their products, that are redirected to the specific support subdivisions (Gold Members support, Support for individuals, etc.). Sorting and redirection were formerly done manually by an entire department. We developed a text mining module to speed up and automate up to 90% of customer support operations.

Automatic customer request sorting by type, complexity, priority, or profitability, and further passing them on to agent

Example

for our client, an air ticket reseller, we developed and delivered a lead scoring system that predicts the “quality” of a given lead based on its marginality and chance of deal closing. The most “quality” leads get high scoring and are processed by the sales team first. it took just six months for the company to get 17% growth in the highly competitive market of air ticket sales.

Example

A service company gets up to 3000 repair orders daily. Most orders are categorized as warranty cases, which are low-margin. But, in fact, up to 20% of warranty cases end up being the non-warranty type, which has a higher margin. WaveAccess has developed and delivered a text mining module that analyzes the order text and detects the case type and its category right at the input.

Risk management

Text mining technology enables complete management of thousands of sources of text documents, and provides the ability to link together information and be able to access the right information at the right time

Fraud detection through claims investigation

The majority of information in jurisprudence is collected as text. Insurance companies use text mining technologies by combining the results of text analysis with structured data to prevent fraud and swiftly process claims

Spam and unwanted content filtering

Spam impacts business productivity and safety due to viruses. Text mining techniques can be implemented to improve the effectiveness of statistical-based filtering methods.

How it works

Data Collection

Text

pdf, xls, doc, plain text

API

Data Processing

Keywords extracting, topic mining, sentiment analysis, categorization

Text Analytics API:
Sentiment, Key Phrases Named Entity Recognition

Azure Search:
Lucene, Microsoft — full-text query parsers

Azure Machine Learning —
training a custom model for text summarization using azureml.PyTorch.

Enriched Data

Metadata: searchable and categorized.
Keywords, sentiments, topics

Enriched archives, knowledge DB

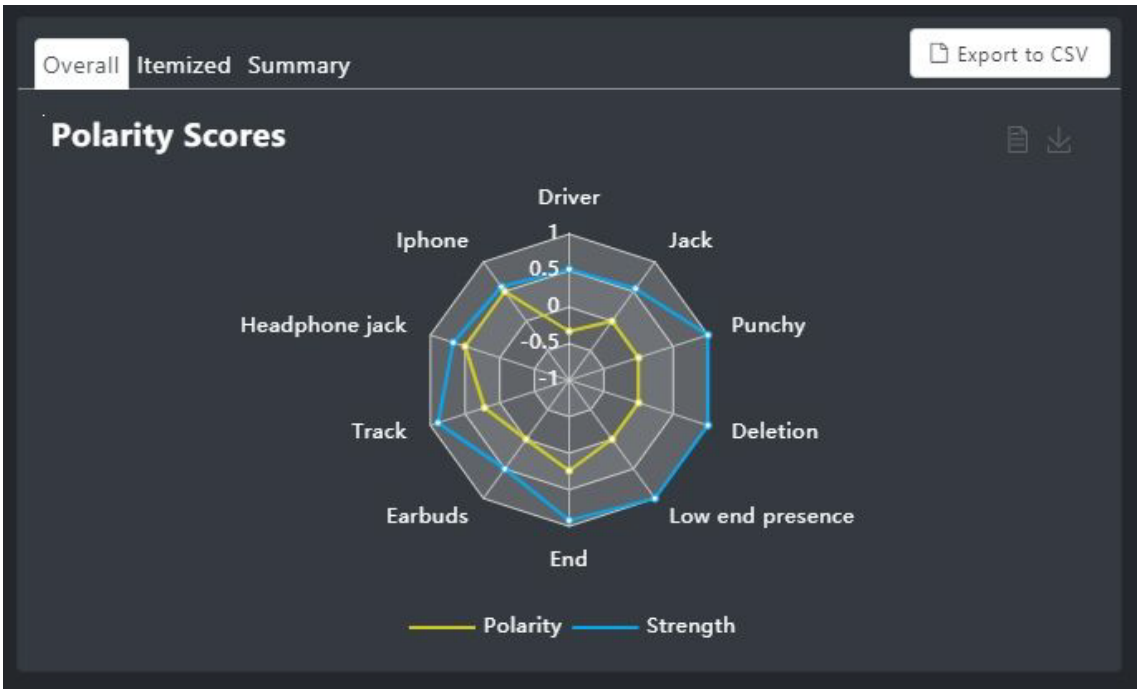
Vizualization:
metrics, scores, reports

Dashboards:
Bizapps/web service

Text mining solution features

Polarity scores for sentiment analysis

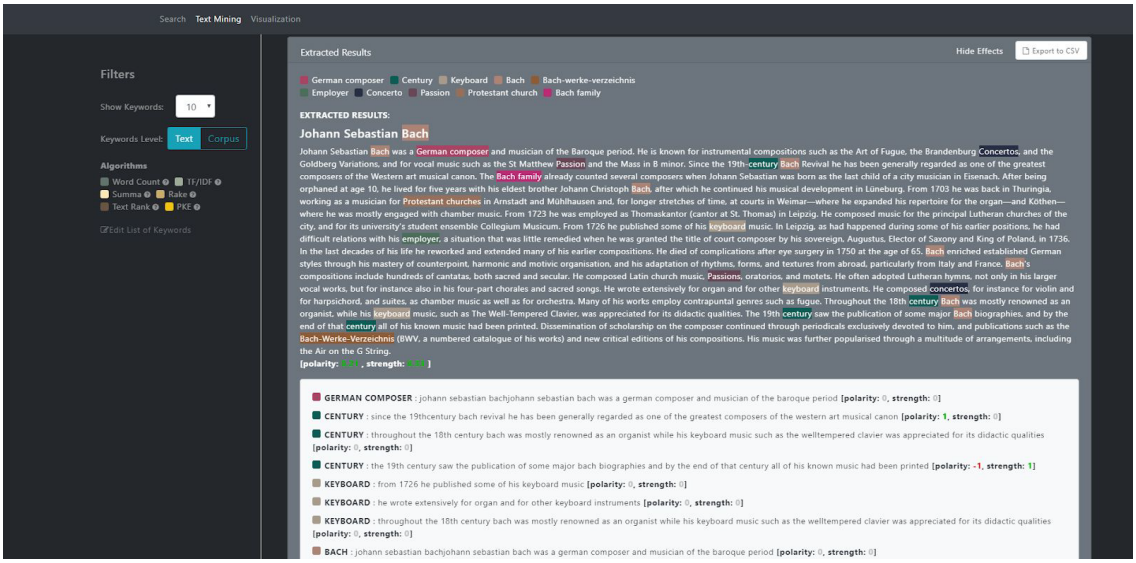
Polarity score refers to identifying sentiment orientation (positive, neutral, and negative) in text. Polarity is float which lies in the range of [-1,1] where 1 means positive statement and -1 means a negative statement.



The grid displays a list of documents, the number of keywords found in the document, and the total rank counted by each algorithm.

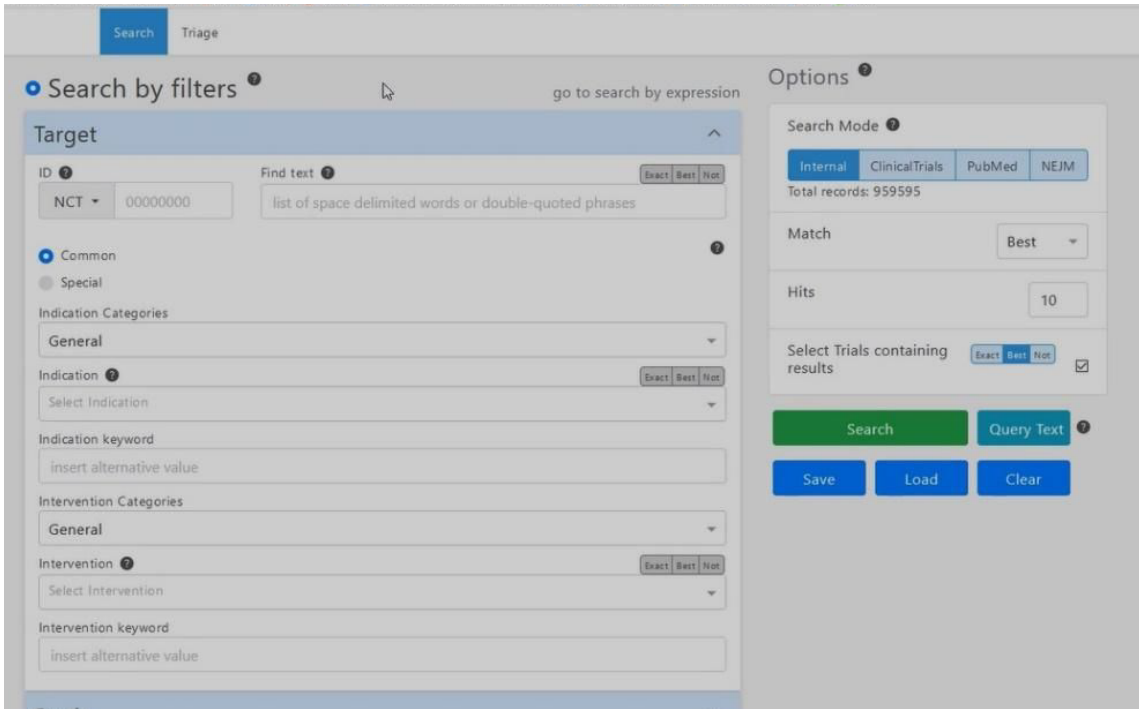
#	Title	Keywords Count	Word Count	TF/IDF	Summa	Rake	Text Rank	PKE	Total
1	Finally audiophile earbuds for iPhone users!	79	25.25	10.128	11.969	10.956	10.857	15.989	85.149
2	Cheap easily tangled cables. Fussy controller. Not comfortable. Headphones &...	82	13.84	7.864	12.928	10.882	15.386	10.951	71.851
3	An audiophile bargain	67	15.514	9.01	10.455	11.196	9.5	10.923	66.598
4	Great sound Great Price	47	16.986	9.969	9	5.662	6.5	14.031	62.148
5	Decent earbuds, clear audio, lacking fullness, not for me.	49	10.346	7.567	7.691	15.541	6	8.953	56.098
6	NoMore	29	18.5	9.643	4	5.287	5.5	7.855	50.785
7	Excellent headphones	45	12.2	11.62	5.5	9.647	6.5	5.161	50.628
8	To enjoy this headset it test your patience as this could be the only headset I ...	43	11	7.448	4.5	11.216	5.5	8.082	47.746
9	Really uncomfortable in my ears	46	8.865	9.818	7.5	6.731	5.5	8.892	47.306
10	Poor build quality on lightning connector	34	13.325	9.719	6	6.899	3.5	4.249	43.692

– **The source text** is displayed with keywords extracted according to selected algorithms (see left)



– **Search by filters**

Advanced search forms (more than 50 filters can be applied) can be customized according to customer text metadata



Input data sources supported

- PDF, XLS, DOC files (API available)
- plain text (API available)

Technologies applied

- Machine Learning
- NLTK
- Sentence Fusion
- GATE
- custom text reading rules and algorithms

Microsoft platform

- MS Azure
- MS Azure ElasticSearch (full text search, custom search)
- MS SQL server
- Azure Machine Learning (azureml.PyTorch)

Languages supported

- English

About

19
years of delivering successful outcomes for customers

350+
talented & passionate professionals in 4 countries

4
global R&D centers and almost any technology

20+
industry verticals from banking to healthcare

300+
successful projects delivered and counting

96%
customer satisfaction index

Las Vegas
headquarters

USA, Denmark and Easten Europe
sales offices

Microsoft Partner
Microsoft

2019 Partner of the Year
Media & Communications Award
2018 Partner of the Year
Artificial Intelligence Award
2017 Partner of the Year
Business Analytics Award

Microsoft Partner
Microsoft

Gold Application Development
Silver DevOps
Silver Cloud Platform
Silver Datacenter

ISO 9001 CERTIFIED

WaveAccess
We Make IT Easy

Contacts

USA

10161 Park Run Drive,
Suite 159 Las Vegas,
Nevada 89145

+1 866 311 24 67

www.wave-access.com

Russia

ул. Большая Морская,
19, офис 4д, 91186
Санкт-Петербург

+7 812 407 2350

www.waveaccess.ru

Denmark

Automatikvej 1,
3.floor,
2860 Søborg

+45 20 55 6222

www.waveaccess.dk

Germany

Gablonzer Str. 11,
Karlsruhe
76185

+49 721 957 3177

www.wave-access.de