



## Unleashing the Power of Words – Our WordsWorth Text-Mining Solution

Our WordsWorth is a state-of-the-art, Cloud-capable text-mining offering from scanning through named-entity recognition, document mark-up, semantic search, text translation & summarization, table & invoice extraction.

### The Need

Today's digital world has resulted a Cambrian explosion of documents, ever easier to produce and to transmit. While offices worldwide have indeed become increasingly "paperless", paper reports have only partially been superseded by electronic data exchanges of structured, tabular data. Instead, unstructured narratives have migrated from paper to PDF, or equivalents. The volume of textual documents has soared, lifted by improved editor tools, automated text generation, and a dramatic increase in the options an author has to disseminate a message: emails, chats,

blogs, social media, cloud drives, collaborative document-sharing suites, to name a few. Furthermore, multiple providers in each of these formats compete to best serve the need for humans to tell a story, to instruct or explain, or simply to express views.

Proliferation of documents, types and formats poses a significant challenge to the reader. It is increasingly difficult to discern useful signals from spurious noise, or even facts from opinion. Digitization of processes and businesses, the "always-on" reachability through mobile devices has

raised expectations for quick results. There is simply not enough time to wade through the flood of documents, to discern which is important or which is reliable. Fortunately, machines armed with Natural Language Processing (NLP) algorithms can help. NLP promises efficiency, quality and exhaustive coverage in working with unstructured, textual documents. ➤

### Our Solution

Pre-trained on universally applicable language models and enhanced with case-specific vocabulary, the text-mining solution WordsWorth excels in accurate interpretation of text documents. It achieves this by combining the most advanced underlying methods from multiple cloud providers (AWS, Azure, GCP) with the flexibility of multiple, dedicated open-source algorithms. It offers users two means to interact with the functionality, either through the intuitive graphical user interface (GUI) or through dedicated Python libraries, which may be invoked via the command line or embedded within custom applications.

WordsWorth performs a wide spectrum of text-mining services:

- OCR – converts scanned text into machine-readable flowing text
- language detection & translation – covers all European languages, plus Chinese and Japanese
- topic modeling – discerns whether a document is relevant to the reader's subject of interest
- named entity recognition – identifies and extracts names, places..., classifying them into personally identifiable information (or not)
- table recognition – identifies tables within the document, exportable into standard formats (.xlsx, .csv)
- invoice extraction – finds all relevant invoice elements and maps them to their text content and coordinates in the original document.
- document mark-up – color-codes identified words, or blacks-out / anonymizes sensitive passages
- version comparison – highlights differences between multiple versions of a document

- semantic search – finds relevant passages associated with the meaning of input search words, beyond exact matches of key-word search
- summarization – paraphrases documents to maximally condense while preserving most relevant topics
- export – converts into a variety of popular editable text files (format selected depending on content)

### Advantages/Benefits

- Speed: quickly determine whether entire documents are relevant, quickly scan for passages of interest
- Quality: find the best available cloud API or open-source Python library, rivaling human performance
- Cost: process everything, avoiding rework or errors common to sample-selection or fatigue
- Integration: work with documents as you would with other files

### Example Use Cases

- Tagging documents (metadata) according to contents – in order to route them to the right recipient
- Providing advanced concept search (or document similarity) to document management systems
- Redacting (blacking out) of sensitive information from confidential / legal documents
- Generally structuring concepts (& associated quantities) from narratives into tables / databases
- Populating systems with data automatically read from invoices / forms
- Providing abstracts / summaries of large documents

### Contacts

**David Thogmartin**

Leader aiStudio

Tel: +49 211 8772 2336

dthogmartin@deloitte.de



Deloitte refers to one or more of Deloitte Touche Tohmatsu Limited (“DTTL”), its global network of member firms, and their related entities (collectively, the “Deloitte organization”). DTTL (also referred to as “Deloitte Global”) and each of its member firms and related entities are legally separate and independent entities, which cannot obligate or bind each other in respect of third parties. DTTL and each DTTL member firm and related entity is liable only for its own acts and omissions, and not those of each other. DTTL does not provide services to clients. Please see [www.deloitte.com/de/ueberUns](http://www.deloitte.com/de/ueberUns) to learn more.

Deloitte provides industry-leading audit and assurance, tax and legal, consulting, financial advisory, and risk advisory services to nearly 90% of the Fortune Global 500® and thousands of private companies. Legal advisory services in Germany are provided by Deloitte Legal. Our professionals deliver measurable and lasting results that help reinforce public trust in capital markets, enable clients to transform and thrive, and lead the way toward a stronger economy, a more equitable society and a sustainable world. Building on its 175-plus year history, Deloitte spans more than 150 countries and territories. Learn how Deloitte’s more than 345,000 people worldwide make an impact that matters at [www.deloitte.com/de](http://www.deloitte.com/de).

This communication contains general information only, and none of Deloitte GmbH Wirtschaftsprüfungsgesellschaft or Deloitte Touche Tohmatsu Limited (“DTTL”), its global network of member firms or their related entities (collectively, the “Deloitte organization”) is, by means of this communication, rendering professional advice or services. Before making any decision or taking any action that may affect your finances or your business, you should consult a qualified professional adviser.

No representations, warranties or undertakings (express or implied) are given as to the accuracy or completeness of the information in this communication, and none of DTTL, its member firms, related entities, employees or agents shall be liable or responsible for any loss or damage whatsoever arising directly or indirectly in connection with any person relying on this communication. DTTL and each of its member firms, and their related entities, are legally separate and independent entities.