

DeepFinder

IT solution for searching and retrieving information from all types of structured or unstructured documents

Use Cases

Audits of industrial facilities

Knowledge capitalisation

Management of industrial assets throughout their life cycle

Maintenance of installations

Programme closure

Renovation of installations

Dismantling of installations

Assistance with diagnostics and audits

Project Benefits

Exploiting archived data

Quick access to data

Pertinent analysis of the data

Easy to use

Decision support

Reduction of uncertainties

CONTEXT & STAKES

Throughout the life cycle of industrial and nuclear installations, operators produce a large quantity of documents in a variety of formats and media.

This information concerns different parameters, and the consultation of this data is essential for maintenance, audits, upgrades, extensions, dismantling operations, etc. In order to process, structure and exploit this information, the text analysis tools (EDM) currently used are insufficient.

New tools allow the aggregation and structuring of all forms of textual data, in paper or digital form, which are then exploited as an aid to decision making at operational and commercial levels.

THE SOLUTION

DeepFinder is an open-source NLP AI (artificial intelligence applied to natural language) solution developed by Assystem. It allows the extraction of relevant data from documents (texts, images, plans) accumulated during the life of a complex industrial installation.

Thus, DeepFinder allows:

- Classification of documents by domain
- Syntactic and semantic indexing of documents for rapid data retrieval
- Natural language interrogation of documents for accurate response
- Creation of domain-specific ontologies
- Secure access to sensitive documents

DEMATERIALISE AND TRANSFORM paper or digital documents into machine-readable text and data

CLASSIFY documents by subject area and key terms

SEARCHING AND STRUCTURING information and data contained in documents by their semantics due to Question Answering systems (search engine)

PROVIDE A SINGLE-ENTRY POINT to this mass of documents

SHARE information between actors

MANAGE several file catalogues

SECURE the storage and access to documents and data

CHARACTERISTICS OF THE SOLUTION

DeepFinder

CHARACTERISTICS OF THE SOLUTION

The DeepFinder solution mixes traditional techniques for capturing and classifying textual documents (EDM, search engine) with the use of ontologies and new Deep Learning and Artificial Intelligence (AI) technologies applied to natural language processing (NLP).

The solution uses an innovative indexing engine that allows the search of millions of documents in a few seconds. A Question Answering model, trained on more than 10,000 nuclear question-answer pairs, can find an exact answer to a user's natural language request regardless of the chosen domain.

DeepFinder can be deployed across multiple data catalogues with user management allowing access to be restricted to one or the other. Once deployed, DeepFinder synchronises the data at a defined frequency.



TECHNICAL CHARACTERISTICS

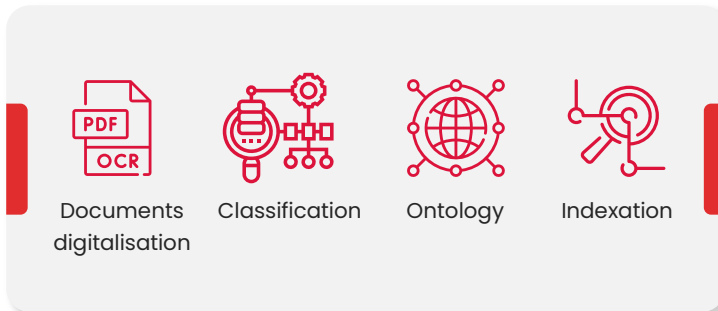
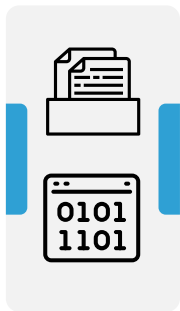
- Operating system: Linux
- Open-Source Modules
- User platform: Web
- Deployment platform: On premise, Azure, AWS
- GPU acceleration: Yes (optional)
- Security: SSO, CAS



PERFORMANCE

- Multi format file supported
- Indexing of several data catalogues
- Deployment in 3 days by catalogue

Unstructured input data



Users



ANSWERS

CASE STUDY



50 GB of data
/ 20,000 files



QA engine trained on
nuclear questions and
answers



Customised
classification of
documents

CEA - Dismantling a nuclear facility

The project concerns the dismantling of a nuclear facility.

Issue. A mass of unstructured documents (PDF, Word, Excel, etc.) from non-centralised archives.

Solution. The DeepFinder solution made it possible to carry out an inventory of the documents and to collect all the infrastructure data relating to dismantling.

Client benefits. Highlighting of archived data - Transformation of non-usable data in an automated way (scanned documents, images, etc.) - Time saving on data research.

business@assystem.com

www.assystem.com

