

L'Intelligence Artificielle au service de l'ingénierie de l'information documentaire dans les projets Sonatrach

Artificial Intelligence at the Service of Information Engineering for Documentary Management in Sonatrach Projects

Brahim Djouadi^{1*}

¹ Sonatrach, Direction Engineering & Project Management (Algérie)

Reçu le : 31/07/2025

Publié le : 31/12/2025

Résumé

À l'ère du numérique, l'ingénierie documentaire renforcée par l'intelligence artificielle (IA) devient un levier stratégique pour la gestion des documents industriels. Cet article étudie un cas chez Sonatrach avec plus de 148 000 documents d'un centre de traitement du gaz. Il propose une méthode pour exploiter ces documents techniques, juridiques et de retour d'expérience. L'usage d'outils IA open source permet d'automatiser le traitement, la classification et l'analyse sémantique. L'article met en avant les défis et les opportunités de l'IA pour optimiser la gestion documentaire dans les projets industriels.

Mots-clés : Ingénierie documentaire ; Intelligence artificielle ; Gestion de l'information ; Projets industriels ; Sonatrach ; Patrimoine informationnel

Abstract

In the digital age, documentary engineering enhanced by artificial intelligence (AI) has become a strategic lever for managing industrial

project documents. This article examines a case at Sonatrach involving over 148,000 documents from a gas processing facility. It proposes a method to exploit these technical, legal, and feedback documents. The use of open-source AI tools enables automated processing, classification, and semantic analysis. The article highlights the challenges and opportunities AI presents to optimize document management in industrial projects.

Keywords: Documentary Engineering; Artificial Intelligence; Information Management; Industrial Projects; Sonatrach; Information Assets

1. Introduction

À l'ère de l'économie numérique, l'information est devenue une ressource stratégique de premier ordre, conditionnant non seulement la qualité des décisions mais aussi l'efficacité opérationnelle, la conformité réglementaire et, plus largement, la compétitivité des organisations. De nos jours, l'information, qu'elle soit technique, administrative, juridique ou financière, constitue un capital immatériel crucial pour les entreprises. Sa gestion est devenue un enjeu majeur dans un environnement dynamique, où l'accès rapide à des informations pertinentes et fiables peut faire la différence entre la réussite et l'échec. La maîtrise de l'information documentaire, c'est-à-dire l'ensemble des documents produits ou reçus lors des activités organisationnelles, est dès lors un élément clé de la performance globale des entreprises.

Dans ce contexte, la gestion de l'information documentaire ne se limite plus à un simple archivage ou à la conservation des documents. Elle englobe l'ensemble des processus visant à collecter, structurer, sécuriser, diffuser et valoriser l'information tout au long de son cycle de vie. Ce corpus documentaire, loin d'être un simple résidu des activités d'une organisation, constitue un patrimoine informationnel

précieux. Sa gestion efficace permet non seulement de réduire les coûts, les erreurs et les risques juridiques, mais aussi de renforcer la prise de décision, d'améliorer la réactivité et de capitaliser sur l'expérience acquise. À l'inverse, une gestion inefficace peut entraîner des pertes de temps considérables, des retards dans les projets et des risques importants en matière de conformité.

Cette problématique revêt une dimension particulièrement cruciale dans les entreprises industrielles de grande envergure, où la complexité technique, la multiplicité des intervenants et l'intensité des réglementations génèrent des volumes considérables de documents. Sonatrach, acteur majeur du secteur énergétique, en est un exemple frappant. Ses projets, souvent stratégiques, à haute intensité capitaliste et fortement réglementés, engendrent une quantité massive d'informations à gérer. La gestion de ces flux documentaires — qu'il s'agisse de plans techniques, de rapports d'étude, de contrats, de permis, de spécifications ou de retours d'expérience — devient ainsi un facteur décisif pour la maîtrise des délais, des coûts, de la qualité et de la conformité. Dans ce cadre, l'ingénierie documentaire émerge comme une discipline transverse visant à organiser, structurer et piloter l'information documentaire tout au long du cycle de vie des projets. Cette approche va au-delà de la simple gestion archivistique pour s'inscrire dans une démarche stratégique intégrant les dimensions organisationnelles, technologiques et métiers. Elle repose sur des outils spécifiques, des méthodologies rigoureuses et une culture de la gestion de l'information partagée par tous les acteurs du projet.

L'apparition de l'intelligence artificielle (IA) marque une avancée majeure dans cette dynamique. Grâce à ses capacités d'automatisation, d'apprentissage et d'analyse prédictive, l'IA offre de nouvelles perspectives pour la transformation de la gestion documentaire. En automatisant l'identification des métadonnées, en extrayant des informations clés des documents non structurés, en améliorant la

recherche documentaire grâce à des moteurs sémantiques et en détectant les incohérences ou anomalies, l'IA permet de rendre la gestion documentaire plus agile et plus intelligente. En outre, l'IA permet également d'anticiper les risques en analysant les données des projets passés, ouvrant ainsi de nouvelles perspectives pour la planification et la prise de décision. L'intégration de ces technologies dans les systèmes documentaires d'entreprise représente un levier stratégique d'optimisation, de sécurisation et de valorisation du patrimoine informationnel, contribuant ainsi à l'efficacité et à la compétitivité des entreprises.

Cet article explore cette convergence entre l'ingénierie de l'information documentaire et les apports de l'intelligence artificielle, en particulier dans le contexte des projets industriels de Sonatrach. Il analyse comment ces technologies peuvent transformer la gestion documentaire, quels sont les bénéfices concrets de cette transformation, ainsi que les défis à relever pour une mise en œuvre réussie. En abordant ces enjeux, l'article met en lumière les nouvelles dynamiques qui façonnent la gestion de l'information dans le secteur industriel et l'importance de cette évolution pour les entreprises contemporaines.

1- L'information documentaire dans l'entreprise

1.1 Définition de l'information documentaire

L'information documentaire désigne l'ensemble des processus liés à la gestion, la collecte, l'organisation et la diffusion de l'information dans un environnement donné. Ce domaine englobe la manière dont les entreprises et les organisations gèrent et exploitent leurs documents et leurs données pour optimiser la prise de décision, l'efficacité opérationnelle et la gestion des connaissances. Une définition plus précise pourrait être : « *L'information documentaire regroupe les activités de gestion des documents, de recherche documentaire et de*

gestion de l'information, à travers des systèmes et des processus qui permettent de collecter, structurer, conserver, et restituer des informations de manière efficace » (De Jean, F., 2009). Cette fonction peut se décomposer en plusieurs sous-activités comme l'archivage, la gestion électronique des documents (GED), la gestion des connaissances, et la veille documentaire.

1.2 L'évolution historique de l'information documentaire

L'évolution de l'information documentaire a suivi le développement des sociétés humaines et la nécessité croissante de gérer l'information. Historiquement, les premières formes d'information documentaire étaient basées sur des supports physiques, tels que des tablettes d'argile, des parchemins ou des papyrus. Les grandes civilisations comme les Égyptiens, les Grecs et les Romains ont mis en place des systèmes pour préserver et transmettre leurs savoirs, souvent dans des bibliothèques. Avec l'arrivée de l'imprimerie au 15^{ème} siècle, la gestion de l'information a connu une révolution majeure. L'imprimerie permettait de diffuser rapidement des connaissances, mais aussi de les archiver sous forme de livres et de revues. À cette époque, la gestion de l'information s'organisait principalement autour de bibliothèques et d'archives, où l'accent était mis sur la conservation physique des documents. Au 20^{ème} siècle, avec l'avènement des technologies de l'information (TI) et des ordinateurs, la gestion documentaire a pris un tournant important. Les entreprises ont progressivement intégré des systèmes informatiques pour gérer et stocker leurs informations. Dans les années 1960 et 1970, des systèmes comme les bases de données relationnelles ont été développés, permettant une gestion plus structurée et plus rapide de l'information. C'est à ce moment-là que l'on parle d'automatisation et de numérisation des processus documentaires.

À partir des années 1990, l'essor d'Internet et de la dématérialisation des documents a bouleversé la gestion de l'information. Les entreprises ont commencé à adopter des logiciels de gestion électronique des documents (GED), permettant de stocker, indexer et retrouver des informations plus facilement. La gestion de l'information est devenue un élément stratégique pour les entreprises, contribuant à leur compétitivité et à leur efficacité opérationnelle.

1.3 L'importance de la fonction information documentaire dans les entreprises

Aujourd'hui, l'information documentaire joue un rôle essentiel dans la compétitivité des entreprises. La gestion de l'information devient un levier stratégique dans plusieurs domaines :

A- **Optimisation de la prise de décision** : L'information est un bien précieux, et sa gestion efficace permet aux dirigeants de prendre des décisions éclairées, basées sur des données précises et actuelles. Une gestion documentaire bien organisée permet de réduire les risques d'erreur liés à la méconnaissance des informations pertinentes.

B- **Gestion des connaissances** : La fonction information documentaire ne se limite pas à la gestion de l'information statique. Elle inclut également la gestion des connaissances et des savoir-faire au sein de l'entreprise. Dans un contexte où l'innovation et la réactivité sont des facteurs clés de succès, une bonne gestion des connaissances internes permet à l'entreprise de capitaliser sur l'expérience et les compétences de ses collaborateurs.

C- **Amélioration de l'efficacité opérationnelle** : La mise en place de systèmes de gestion documentaire bien conçus permet de réduire le temps de recherche d'information, d'optimiser les processus et de limiter les erreurs. Cela augmente la productivité des employés et améliore la qualité des services rendus.

D- Conformité et sécurité : Les entreprises sont souvent soumises à des exigences légales et réglementaires strictes concernant la conservation des documents. La fonction information documentaire veille à ce que les informations sensibles soient protégées et archivées conformément à ces exigences.

E- Collaboration et communication : L'information documentaire permet également d'améliorer la collaboration et la communication au sein des équipes. Les systèmes de gestion documentaire permettent de partager des informations de manière sécurisée et en temps réel, facilitant ainsi le travail d'équipe.

Au fur et à mesure que la technologie évolue, la fonction information documentaire subit une transformation profonde. De simples processus de gestion et d'archivage de l'information, cette fonction s'oriente progressivement vers une approche plus intégrée et stratégique de l'information. L'introduction des nouvelles technologies, telles que l'intelligence artificielle, le big data et la gestion des données massives (data science), transforme radicalement la manière dont les informations sont collectées, traitées, et utilisées au sein des entreprises.

Dans ce contexte, la fonction information documentaire évolue pour devenir une véritable *ingénierie de l'information documentaire*. Ce passage de la gestion documentaire traditionnelle à l'ingénierie documentaire implique la mise en place de systèmes de gestion plus complexes et intelligents, capables de gérer des volumes d'informations de plus en plus importants, tout en répondant aux enjeux de qualité, de sécurité, et d'accessibilité.

2 - L'ingénierie de l'information documentaire

2.1 Définition et objectifs de l'ingénierie de l'information documentaire

L'ingénierie de l'information documentaire (IID) est une discipline qui vise à optimiser la gestion, la structuration, le traitement et l'exploitation de l'information dans une organisation. Contrairement à la gestion documentaire traditionnelle, qui se concentre principalement sur le stockage et l'archivage des documents, l'ingénierie de l'information documentaire s'intéresse à la conception de systèmes et de processus permettant de traiter l'information de manière intelligente et stratégique (Fouquet, P. 2008). Elle inclut des outils et des méthodes avancées pour structurer l'information, favoriser son partage, et en maximiser la valeur.

L'ingénierie de l'information documentaire repose sur plusieurs axes fondamentaux :

- **Conception de systèmes d'information documentaires** : l'intégration de logiciels et d'outils pour la gestion et le traitement de l'information (ex. systèmes de gestion de contenu, outils d'indexation, bases de données).
- **Automatisation et optimisation des processus** : mise en place de solutions pour automatiser la collecte, l'indexation, le stockage et la diffusion de l'information.
- **Valorisation de l'information** : transformation de l'information en connaissance utile pour l'organisation.
- **Sécurité et conformité** : gestion des risques liés à la confidentialité, à la sécurité des données et à la conformité avec les normes réglementaires.

2.2 L'évolution de l'ingénierie de l'information documentaire

L'ingénierie de l'information documentaire a émergé avec l'essor des nouvelles technologies et l'augmentation exponentielle de la quantité d'information générée par les organisations. Cette évolution s'est construite autour de plusieurs grandes étapes (Caron, J. & Lambert, C. 2015) :

A- Les premières formes de gestion documentaire (avant 1960) : L'information était principalement gérée sous forme papier, avec des processus manuels pour la collecte, le classement, et l'archivage. À cette époque, les bibliothèques et les archives étaient les principales structures pour organiser l'information.

B- L'automatisation et la numérisation (1960-1990) : L'introduction des premiers ordinateurs et des bases de données a permis de numériser l'information, de faciliter la recherche documentaire et d'optimiser les processus de gestion documentaire. Les systèmes de gestion électronique des documents (GED) sont apparus à cette époque.

C- Les systèmes d'information intelligents (1990-2000) : La gestion de l'information est devenue plus complexe avec l'essor des technologies comme les bases de données relationnelles et les systèmes de gestion de contenu d'entreprise (ECM). L'automatisation a permis de mieux structurer et indexer l'information.

D- L'émergence des technologies avancées (2000-présent) : Aujourd'hui, l'ingénierie de l'information documentaire s'appuie sur des technologies de pointe comme l'intelligence artificielle (IA), le machine learning, la gestion de l'information dans le cloud, et la gestion des données massives (big data). Ces outils permettent une gestion plus dynamique et proactive de l'information, facilitant la prise de décision, l'analyse prédictive et l'automatisation des processus.

2.3 Les composants clés de l'ingénierie de l'information documentaire

L'ingénierie de l'information documentaire repose sur plusieurs composants qui interagissent pour garantir une gestion efficace de l'information :

A- La structuration de l'information :

○ **Taxonomie et ontologies** : Il s'agit de la création de modèles hiérarchiques et de schémas conceptuels pour organiser l'information et faciliter sa recherche. Les ontologies permettent de mieux comprendre les relations entre les différents types d'information et de la catégoriser de manière logique et intelligible.

○ **Métadonnées** : La gestion des métadonnées permet de décrire, indexer et organiser l'information de manière standardisée, afin d'en faciliter l'accès et l'exploitation.

B- L'automatisation des processus documentaires :

○ **Workflow documentaire** : Ce processus permet d'automatiser la gestion des flux d'information au sein de l'entreprise, en optimisant la circulation et l'approbation des documents (par exemple, automatisation des circuits de validation).

○ **OCR (Reconnaissance Optique de Caractères)** : La numérisation des documents papier permet d'extraire des informations de texte à partir d'images ou de documents scannés.

C- La gestion du cycle de vie de l'information (records management) :

○ Cela comprend toutes les étapes, depuis la création du document jusqu'à son archivage et sa destruction. La gestion du cycle de vie vise à garantir que l'information est conservée pendant la période nécessaire à son utilisation et qu'elle est éliminée de manière sécurisée lorsqu'elle n'est plus utile.

D- L'intégration des données et l'interopérabilité :

○ L'ingénierie documentaire s'appuie sur des systèmes qui permettent d'intégrer l'information provenant de diverses sources (bases de données, emails, documents divers) pour créer une vue unifiée et centralisée de l'information.

E- L'intelligence documentaire :

○ L'intelligence documentaire combine les outils d'analyse de l'information avec les technologies d'intelligence

artificielle. Cela permet d'extraire des tendances, des insights ou de réaliser des analyses prédictives à partir des documents et des données traitées.

2.4 Les technologies au cœur de l'ingénierie de l'information documentaire

L'ingénierie de l'information documentaire s'appuie sur un ensemble de technologies avancées qui permettent de gérer l'information de manière plus précise, plus rapide et plus automatisée (Bataille, M. & Gagnon, C. (2021) :

A- **Systèmes de gestion électronique des documents (GED)** : Ces systèmes permettent de numériser, organiser et rechercher des documents électroniques à l'aide de métadonnées et d'indexation.

B- **Big Data et Data Analytics** : L'exploitation des grandes quantités de données (structured et non structured) permet d'en extraire des informations pertinentes pour la stratégie de l'entreprise. Ces technologies sont particulièrement utiles dans l'analyse prédictive et l'intelligence d'affaires.

C- **Intelligence artificielle (IA) et machine learning** : Ces technologies aident à automatiser l'indexation des documents, à améliorer la recherche documentaire et à extraire des informations cachées dans les textes grâce à l'analyse sémantique et au traitement du langage naturel (NLP).

D- **Cloud Computing** : Les plateformes cloud permettent une gestion de l'information à distance et partagée, facilitant l'accès aux documents depuis n'importe quel endroit tout en garantissant leur sécurité.

E- **Blockchain** : Dans certains contextes, la blockchain est utilisée pour garantir l'intégrité des informations et des documents, notamment dans les secteurs où la traçabilité est essentielle (juridique, financier).

3 - Impact de l'IA sur l'ingénierie documentaire des projets de Sonatrach

Après avoir posé les bases théoriques de l'information documentaire et de l'ingénierie documentaire, ce chapitre se concentre sur l'application concrète de ces concepts, augmentés par l'intelligence artificielle (IA), dans le contexte spécifique des projets de Sonatrach. Nous examinerons comment l'immense fonds documentaire généré par la construction de complexes de traitement de gaz (CPF) existants peut être transformé en un atout stratégique pour la réalisation de nouveaux projets de "boosting" adjacents, en mettant l'accent sur les retours d'expérience, les documents techniques et les aspects juridiques.

3.1 Valorisation Stratégique du Fonds Documentaire Existant de Sonatrach

La construction d'un CPF (Central Processing Facility) de gaz est une entreprise colossale, générant une quantité pléthorique d'informations. Chez Sonatrach, un tel projet a pu produire plus de 148 000 documents, incluant des échanges avec les autorités, des spécifications techniques, des rapports d'ingénierie, des manuels d'opération, des contrats juridiques, des plans de sécurité, et bien d'autres. Historiquement, ce volume est une mine d'or mal exploitée, dans des formats hétérogènes et des systèmes de recherche limités.

L'ingénierie de l'information documentaire, appuyée par l'IA, offre une voie pour déverrouiller cette valeur. Plutôt que de repartir de zéro pour chaque nouveau projet de "boosting", il devient possible d'exploiter activement les connaissances encapsulées dans ces documents historiques.

3.1.1 Le Défi du "Brownfield" et l'Opportunité du Passé

Les projets d'extension ou de modification sur un site existant (dits "brownfield") sont intrinsèquement complexes. Ils nécessitent une compréhension approfondie de l'infrastructure en place, de son histoire opérationnelle, de ses contraintes et de ses performances passées. Le fonds documentaire d'un CPF est le reflet de cette histoire.

Problématique sans IA : Sans une approche structurée et intelligente, retrouver une information cruciale – par exemple, la spécification exacte d'une vanne installée il y a 5 ans ou un rapport d'incident lié à un type de compresseur spécifique – peut prendre des jours, voire des semaines. Cela engendre des retards, des coûts supplémentaires, des erreurs de conception et des risques accrus.

L'opportunité avec l'IA : L'IA permet de transformer ces 148 000+ documents d'un simple dépôt d'archives en une base de connaissances vivante et intelligente. Elle rend possible une recherche sémantique, l'extraction automatique d'informations, la détection de relations et la génération de retours d'expérience automatisés.

3.2 Cas Pratique : Exploitation des Données du CPF pour les Projets de "Boosting"

Considérons un scénario où Sonatrach entreprend la construction d'une unité de "boosting" adjacente à l'un de ses CPF de gaz existants.

3.2.1 Retour d'Expérience (REX) via l'IA

Le REX est capital pour l'amélioration continue. Les 148 000+ documents du CPF existant contiennent des milliers d'années-homme d'expérience.

A. Extraction Automatisée des Leçons Apprises :

L'IA (via des techniques de modélisation thématique et de résumé automatique) peut parcourir tous les rapports d'incidents, de

maintenance, les comptes rendus de réunions de projet et les non-conformités. Elle peut identifier des schémas récurrents de défaillance pour des équipements spécifiques.

Quelques Logiciels « Open Source » pour l'extraction automatisée des leçons apprises :

- **SpaCy ou NLTK** : pour la « tokenisation, la lemmatisation » (1), et la reconnaissance d'entités nommées.
- **Gensim** : pour la modélisation thématique (Latent Dirichlet Allocation - LDA).
- **Hugging Face Transformers** : pour des modèles de résumé automatique plus avancés.

Comment ça marche : Un modèle d'apprentissage non supervisé analyse les textes pour y détecter des groupes de mots co-occurents (2), formant des "sujets". Des règles sémantiques ou des modèles d'apprentissage supervisé, entraînés sur des exemples de "leçons apprises" manuelles, peuvent ensuite extraire les conclusions et recommandations.

B. Prédiction des Risques Projet :

Par exemple, En analysant les déviations de budget et de planning des projets précédents du CPF, ainsi que les causes profondes identifiées dans les rapports de clôture, l'IA peut fournir une estimation des risques pour le nouveau projet de "boosting" (ex : "risque élevé de dépassement de 10% du budget pour les travaux de raccordement en raison de complexités imprévues déjà rencontrées").

Quelques Logiciels « Open Source » pour la prédiction des Risques Projet :

Scikit-learn : Pour les modèles de régression ou de classification prédictive.

Pandas : Pour la manipulation de données structurées extraites des rapports de projet. (3)

Comment ça marche : Après l'extraction des données numériques et Textuelles pertinentes, des modèles d'apprentissage automatique sont entraînés sur les performances passées pour prédire les issues futures.

Activité REX	Sans IA (Moyenne)	Avec IA (Cible)	Bénéfice Qualitatif/Quantitatif	Outils IA / Logiciels Open Source
Recherche de REX pertinents	Plusieurs jours/semaines	Quelques minutes/heures	Réduction drastique du temps, exhaustivité	Recherche sémantique, graphes de connaissances (Neo4j, ArangoDB)
Synthèse des leçons apprises	Manuelle, fragmentée	Automatisée, structurée	Identification des schémas récurrents, prévention des erreurs	Modélisation thématique (Gensim), Résumé automatique (Hugging Face)
Capitalisation de la connaissance	Difficile, perdue	Structurée, interrogeable	Base de connaissances vivante, accès facilité pour les nouveaux projets	Graphes de connaissances (Neo4j), Outils NLP (SpaCy)
Prédiction des risques projet	Intuitive, basée sur l'expérience	Basée sur données historiques, objective	Réduction des dépassements de coût et de délai	Scikit-learn, Pandas

Tableau 1 : Impact de l'IA sur le Processus de Retour d'Expérience

3.2.2 Exploitation des Documents Techniques

Les schémas P&ID, PFD, les manuels d'équipements, les plans de fabrication, les certifications sont des actifs techniques d'une valeur inestimable.

A. Extraction de Données Techniques Structurées :

Exemple Concret : Pour le projet de "boosting", il est crucial de connaître la pression maximale de service (PMS) de chaque ligne existante sur le CPF adjacent, le matériau de chaque tuyauterie

interconnectée, ou les dimensions exactes d'un skid pour la pré-fabrication.

L'IA peut extraire ces données directement des P&ID (via Computer Vision pour l'analyse des symboles et du texte) et des fiches techniques (via Extraction d'Information).

Quelques Logiciel Open Source Données Techniques Structurées :

OpenCV : Pour l'analyse d'images et la reconnaissance de formes sur les P&ID.

Tesseract OCR : Pour l'extraction de texte à partir d'images/scans (souvent intégré dans des pipelines).

PyPDF2 ou fitz (PyMuPDF) : Pour l'extraction de texte et l'analyse de structure des PDF techniques.

LangChain / LlamaIndex : avec des modèles de langage de grande taille (LLMs) pour l'extraction d'information contextuelle à partir de documents non structurés.

B. Vérification de Conformité aux Standards :

Exemple Concret : S'assurer que les nouvelles spécifications d'équipements pour le "boosting" sont compatibles avec les standards (API, ISO) utilisés sur le CPF existant.

L'IA peut rapidement comparer les exigences de l'ancien CPF avec les propositions du nouveau projet et signaler les écarts ou les non-conformités potentielles.

Quelque logiciels Open source pour la Conformité aux standards:

- **SpaCy :** Pour la REN des noms de standards (5).
- **Faiss :** Pour la recherche de similarité sémantique entre les descriptions techniques et les paragraphes de normes.

Comment ça marche : Création d'une base de données sémantique des standards et normes extraits, puis comparaison des descriptions des nouveaux équipements avec cette base.

Type de Document	Problématique sans IA	Solution IA	Bénéfice Clé pour le "Boosting"	Outils IA/Logiciels Open Source
P&ID / PFD	Lecture manuelle, risque d'erreurs, lourd	Extraction auto. des tags, paramètres, flux	Cohérence de conception, interconnexion facilitée	OpenCV, Tesseract OCR, LangChain
Manuels Équipements	Recherche fastidieuse de spécifications	Extraction de fiches techniques structurées	Gain de temps pour ingénierie et O&M, sélection optimisée	PyPDF2, SpaCy, Hugging Face (pour extraction)
Rapports d'Ingénierie	Volume élevé, infos dispersées	Résumé, modélisation thématique, extraction de faits	Identification rapide des études critiques, optimisations	Gensim, SpaCy, LangChain
Certificats Matériaux	Vérification manuelle, perte traçabilité	Extraction auto. des propriétés, liage aux équipements	Assurance qualité, conformité réglementaire	Tesseract OCR, SpaCy

Tableau2 : Application de l'IA aux Documents Techniques pour le Projet de "Boosting"

3.2.3 Analyse des Documents Juridiques et Administratifs

Les aspects légaux et contractuels sont cruciaux pour la conformité et la gestion des risques.

A. Analyse Contractuelle :

Pour un projet de "boosting", il est vital de comprendre les clauses de garantie avec les anciens fournisseurs du CPF, les pénalités pour non-respect des délais, ou les clauses de confidentialité qui pourraient impacter la collaboration avec de nouveaux partenaires.

L'IA peut analyser des centaines de contrats (de construction, d'approvisionnement, de services) pour extraire ces clauses clés et les comparer avec les exigences du nouveau projet.

Quelques logiciels Open source pour l'analyse Contractuelle :

- **SpaCy ou NLTK** : Pour la REN de termes juridiques spécifiques (parties contractantes, dates, montants, types de clauses).
- **SetFit / Sentence-Transformers** : Pour la classification de clauses contractuelles par type (ex: force majeure, propriété intellectuelle, résiliation).
- **DeepPavlov** : pour des capacités de Question Answering sur des textes juridiques.

Comment ça marche : Des modèles de TLN sont entraînés sur un corpus de contrats pour reconnaître les structures et les termes spécifiques au langage juridique. Des ontologies juridiques peuvent être développées pour structurer cette information.

B. Conformité Réglementaire et Permis :

Assurer que toutes les exigences en matière d'environnement, de sécurité et de permis applicables au CPF existant sont prises en compte pour le nouveau projet de "boosting".

L'IA peut scanner les rapports d'audit, les permis d'exploitation et les échanges avec les autorités pour extraire les obligations et les contraintes spécifiques.

Quelques logiciels Open source pour la Conformité Réglementaire et Permis

Apache OpenNLP (Java) ou SpaCy : Pour l'extraction d'information.

Elasticsearch (avec intégration NLP) : Pour une recherche full-text sémantique sur l'ensemble des documents réglementaires.

Comment ça marche : Les documents sont indexés de manière sémantique, permettant des requêtes complexes comme "Quelles sont les exigences environnementales pour l'évacuation de H2S dans la zone du CPF existant ?".

Activité	Sans IA (Moyenne)	Avec IA (Cible)	Bénéfice Clé pour le "Boosting"	Outils IA/Logiciels Open Source
Recherche de clauses contractuelles	Lecture manuelle fastidieuse	Extraction automatique, recherche sémantique	Accélération des négociations, réduction des risques	SpaCy, SetFit, DeepPavlov
Vérification conformité réglementaire	Expertise manuelle, risques d'oubli	Analyse croisée automatisée des textes	Réduction des amendes, obtention rapide des permis	Apache OpenNLP, Elasticsearch
Gestion des litiges/échanges avec autorités	Processus long, difficile traçabilité	Structuration des correspondances, résumé	Meilleure défense, traçabilité facilitée	SpaCy, Hugging Face

Tableau 3 : Impact de l'IA sur la Gestion des Documents Juridiques et Administratifs

Ce chapitre a permis de mettre en lumière le potentiel transformateur de l'intelligence artificielle lorsqu'elle est appliquée à l'ingénierie documentaire des projets industriels complexes de Sonatrach. Trois catégories fondamentales de documents ont été explorées : les retours d'expérience, les documents techniques, et les contenus juridiques et administratifs révélant à chaque fois les apports concrets de l'IA en matière de rapidité d'accès, de fiabilité, et de valorisation de la connaissance.

À travers ces cas d'usage, il apparaît clairement que les corpus documentaires historiques, souvent perçus comme des passifs encombrants, peuvent devenir des leviers stratégiques puissants pour les projets futurs, notamment dans les contextes de "boosting". L'IA ne se contente pas d'automatiser les tâches ; **elle permet d'extraire du sens, d'anticiper les risques et d'uniformiser la conformité, réduisant ainsi les coûts, les délais et les incertitudes.**

Cette approche proactive et intelligente de la gestion documentaire préfigure une nouvelle manière de concevoir l'ingénierie projet, où la mémoire collective est structurée, accessible, et mise au service de l'innovation continue.

Défis et perspectives d'avenir

Malgré les avancées, plusieurs défis demeurent pour l'intégration de l'IA dans l'ingénierie documentaire :

Qualité des données : L'efficacité de l'IA dépend de la qualité des données d'entrée. Les systèmes IA doivent être alimentés par des données bien structurées et de haute qualité pour produire des résultats fiables.

Sécurité et confidentialité : Le traitement des documents sensibles par l'IA pose des enjeux de confidentialité, particulièrement dans des secteurs comme la santé, la finance et le juridique.

Adoption par les utilisateurs : L'acceptation de l'IA dans des processus critiques peut être freinée par des préoccupations liées à la confiance dans les systèmes automatisés.

Malgré ces défis, les perspectives sont prometteuses. À mesure que les technologies d'IA évoluent, de nouveaux outils encore plus intelligents et adaptatifs seront disponibles pour répondre aux besoins croissants des entreprises en matière de gestion documentaire.

Conclusion

L'ingénierie de l'information documentaire représente une avancée significative par rapport aux méthodes traditionnelles de gestion documentaire. En combinant des technologies avancées, des processus automatisés et une approche stratégique de la gestion de l'information, elle permet aux organisations de tirer le meilleur parti de leurs ressources informationnelles. Elle répond ainsi à des enjeux croissants de compétitivité, d'efficacité et de sécurité dans un environnement de plus en plus numérique et interconnecté.

L'évolution vers une ingénierie documentaire plus complexe et plus sophistiquée continue d'ouvrir de nouvelles perspectives pour les entreprises qui souhaitent exploiter pleinement la richesse de leurs données et optimiser leur gestion documentaire.

Dans un contexte de projets complexes et d'envergure, comme ceux menés par Sonatrach, la gestion efficace des documents devient un facteur déterminant pour garantir le succès et la rentabilité des projets. Les volumes massifs de données générés, ainsi que la diversité des types de documents (techniques, financiers, contractuels), représentent un défi majeur pour les équipes de gestion. C'est dans ce cadre que

l'ingénierie documentaire, soutenue par les avancées récentes en intelligence artificielle (IA), prend toute son importance.

L'ingénierie documentaire, en permettant l'organisation, l'archivage et l'accès aux informations de manière structurée, se voit désormais renforcée par l'IA qui, par ses capacités d'automatisation, d'analyse de données et de traitement intelligent de l'information, permet d'optimiser l'ensemble du cycle de gestion documentaire. Grâce à des technologies telles que le traitement du langage naturel (NLP), l'apprentissage automatique et l'analyse prédictive, l'IA offre des solutions novatrices pour classer, indexer et extraire les informations clés en temps réel, réduisant ainsi les risques d'erreurs humaines, les délais de recherche, et les coûts liés à la gestion des documents.

Pour Sonatrach, leader dans le secteur de l'énergie, l'intégration de l'IA dans l'ingénierie documentaire ne se limite pas à une simple optimisation technique, mais représente un levier stratégique pour améliorer la performance des projets. Cette synergie entre l'IA et l'ingénierie documentaire permet non seulement d'améliorer la gestion de l'information mais aussi de renforcer la prise de décision, la conformité et la transparence dans les processus, en particulier dans des projets complexes où la précision et la rapidité sont cruciales.

Bibliographie

- [1] AMEUR, A. Intelligence artificielle et étude en ligne : acquisition des savoirs et performances en question. Revue algérienne des lettres, 2024, vol. 8, no 3, p. 89-102.
- [2] BENMOUSSA, S., et al. Intégration de l'IA dans les systèmes d'information documentaire des entreprises pétrolières. Revue Algérienne des Sciences de l'Information, 2023.
- [3] CAZALS, F., et CAZALS, C. Intelligence artificielle : L'intelligence amplifiée par la technologie. Bruxelles : De Boeck, 2019.

- [4] COLAVIZZA, G., BLANKE, T., JEURGENS, C., et NOORDEGRAAF, J. Archives and AI: An Overview of Current Debates and Future Perspectives. arXiv preprint arXiv:2105.01117, 2021.
- [5] DOUCET, G. La gestion de l'information dans les projets industriels. Paris : Éditions AFNOR, 2014.
- [6] EL KADDIOUI EL IDRISI, S., et JABRAOUI, S. Les effets de l'intelligence artificielle : une revue de littérature. Revue Marocaine de la Prospective en Sciences de Gestion, 2022, no 4.
- [7] GELIN, R. Dernières nouvelles de l'intelligence artificielle. Paris : Flammarion, 2022.
- [8] GORIA, S., et PINÈDE, N. Les systèmes d'information et les risques dans les grands projets. London : ISTE Editions, 2017.
- [9] HENDRYCKS, D. Introduction to AI Safety, Ethics, and Society. arXiv preprint arXiv:2411.01042, 2024.
- [10] METTLER, T. Digital Transformation of Document Management : Challenges and Opportunities. Journal of Enterprise Information Management, 2021.
- [11] NONAKA, I., et TAKEUCHI, H. The Knowledge-Creating Company. Oxford : Oxford University Press, 1995.
- [12] PEREIRA, V., BASILIO, M. P., et SANTOS, C. H. T. pyBibX -- A Python Library for Bibliometric and Scientometric Analysis Powered with Artificial Intelligence Tools. arXiv preprint arXiv:2304.14516, 2023.
- [13] RUSSELL, S., et NORVIG, P. Intelligence artificielle : une approche moderne. 4e éd. Paris : Pearson, 2021.
- [14] Revue Cognitive Computation. [En ligne]. Revue internationale sur les systèmes cognitifs naturels et artificiels. Consultée en 2025.
- [15] WELLER, T., et KANZA, S. Artificial Intelligence in Information Management. Cham : Springer, 2020.

Note

¹ La lemmatisation et la tokenisation sont des étapes fondamentales dans le traitement du langage naturel. Elles permettent de : Réduire la complexité du texte, facilitant l'identification des tendances, des motifs et des anomalies.

2 "Le modèle analyse les co-occurrences de mots", ça signifie que le modèle (souvent un programme d'ordinateur basé sur l'intelligence artificielle) ne regarde pas les mots un par un, mais plutôt quels mots apparaissent souvent ensemble dans les textes.

3 Mendenhall M. J., Peterson G. L., Graves A., Butler J. W. – Lessons Learned in Human-Artificial Intelligence Teaming in Business Processes. – États-Unis : Software Engineering Institute, 2022. – (Actes de symposium AAAI Spring Symposium, pp. 1-8

4 Le traitement du langage naturel

5 Reconnaissance d'Entités Nommées (REN) : C'est une technique d'intelligence artificielle (plus précisément du Traitement du Langage Naturel - NLP) qui permet à un ordinateur de détecter et d'identifier automatiquement des "entités" spécifiques dans un texte. Ces entités peuvent être des personnes, des lieux, des organisations, des dates, des montants, etc.