# Technologies sémantiques et accès à l'information dans le prescrit SNCF

Coralie Reutenauer, Luce Lefeuvre, Aurélie Fouqueray, Thibault Prouteau, Valentin Pelloin, Cédric Lopez, Camelin Nathalie, Frédérique Segond, Dugué Nicolas, Didier Bourigault

Institut pour la Maîtrise des Risques
Sûreté de Fonctionnement - Management - Cindyniques

# Technologies sémantiques et accès à l'information dans le prescrit SNCF
# Semantic technologies and information retrieval in SNCF prescriptive documentation

Reutenauer Coralie
*Direction Risques Sécurité Sûreté*
*SNCF*
La Plaine Saint-Denis, France
coralie.reutenauer@sncf.fr

Lefeuvre Luce
*Direction Innovation et Recherche*
*SNCF*
La Plaine Saint-Denis, France
luce.lefeuvre@sncf.fr

Fouqueray Aurélie
*Direction Risques Sécurité Sûreté*
*SNCF*
La Plaine Saint-Denis, France
aurelie.fouqueray@sncf.fr

Prouteau Thibault / Pelloin Valentin
*LIUM - Le Mans Université*
Le Mans
prenom.nom.etu@univ-lemans.fr

Camelin Nathalie
*LIUM - Le Mans Université*
Le Mans
nathalie.camelin@univ-lemans.fr

Dugué Nicolas
*LIUM - Le Mans Université*
Le Mans
nicolas.dugue@univ-lemans.fr

Lopez Cédric
*Emvista*
Montpellier, France
cedric.lopez@emvista.com

Segond Frédérique
*INRIA*
Grenoble, France
frederique.segond@inria.fr

Bourigault Didier
*Synomia*
Boulogne Billancourt, France
didier.bourigault@synomia.com

***Résumé*—Des expérimentations basées sur des technologies de traitement automatique du langage ont été menées au sein d'un programme de sécurité ferroviaire et de simplification documentaire afin d'améliorer la recherche d'information et la rédaction dans les textes de prescription SNCF.**

***Abstract*—In order to improve documentation production and information retrieval for a better rail safety, several prototypes based on Natural Language Processing technologies and applied to SNCF prescriptive documentation were developed and assessed.**

***Keywords—TALN, sécurité, transformation numérique, documentation, recherche d'information***

## I. INTRODUCTION

En 2015, suite à une série d'incidents, le programme PRISME Excellence sécurité est lancé à SNCF. Un de ses volets, la Simplification, vise à transformer le processus documentaire pour pallier les limites de la documentation sécurité. L'enjeu est de permettre, à travers une documentation orientée utilisateur, un juste accès à l'information métier dans une base de 90 000 textes prescriptifs.

La Simplification met en place de nouveaux outils numériques pour optimiser la production des contenus et leur consultation, en plusieurs étapes : numérisation des textes, réalisation d'une première application de consultation et de recherche (Digidoc), prototypage d'applications basées sur de la documentation modulaire et structurée avec réécriture des textes en XML et leur découpage en modules permettant de la production de documentation personnalisée à la volée. Au-delà se pose la question de nouveaux systèmes intelligents d'accès aux contenus, d'aide à l'interprétation et à la saisie. Les technologies sémantiques basées sur du Traitement Automatique du Langage se présentent comme une opportunité pour répondre aux besoins métiers et pour prévenir les risques langagiers, aux conséquences potentiellement graves en milieu professionnel lorsqu'ils se traduisent par une mauvaise transmission de l'information [1]. Elles s'inscrivent dans le sillage des tendances montantes de la recherche d'information aussi bien technologiques (montée en puissance du web sémantique et de la recherche sémantique [2] [3]) que d'usage (génération Google habituée à des réponses de plus en plus immédiates et guidées aux questions formulées). Dans le contexte des documents métiers SNCF, elles soulèvent des problématiques liées à la terminologie textuelle et à la recherche d'information dans un domaine de spécialité [4].

En collaboration avec plusieurs partenaires industriels et académiques (sociétés Synomia, Viseo, Lincoln, laboratoire LIUM), le plateau Simplification et la Direction Innovation et Recherche SNCF mènent entre 2017 et 2019 des travaux expérimentaux afin d'explorer et valider le potentiel de technologies sémantiques pour les usages métiers de la documentation numérique.

## II. OBJECTIFS

L'objectif est d'identifier, de maquetter et d'évaluer des solutions sémantiques pour l'aide à la rédaction, la recherche d'information et la navigation dans les contenus. La démarche vise à connaître les potentiels et limites des technologies testées, de valider leur adéquation aux besoins et d'identifier des outils et méthodes pertinents en vue d'une industrialisation. Pour l'utilisateur final, l'enjeu est de simplifier la production de contenus et l'accès à l'information pertinente, avec à la clé gain de temps et

efficacité accrue dans la production et la recherche d'information.

## III. METHODE

### A. Constitution du jeu de données

La base documentaire complète comporte 90 000 textes de prescription (directives, règles, procédures, …), destinés à tout le groupe SNCF. Le discours est normatif, les textes sont en français et la langue est spécialisée, avec l'usage de la terminologie métier SNCF.

Le corpus utilisé pour les expérimentations est limité à environ 10 000 textes pour des raisons de confidentialité, avec 95% de textes dits opposables, traitant majoritairement de sujets sécurité, du trafic et de l'exploitation ferroviaires, et 5% de textes produits par le service des ressources humaines. Les textes sont composés d'un titre, d'un thème attribué manuellement par les rédacteurs et administrateurs de la base, et du contenu textuel récupéré par un outil de conversion de PDF en texte.

### B. Définition des besoins

Etayée par une veille technologique et l'analyse de l'existant, une première étape d'analyse et d'expression de besoins structure la démarche exploratoire en plusieurs lots.

Le premier lot est la **création de ressources, l'annotation et la représentation sémantiques du corpus de textes**, prérequis aux autres lots. Il nécessite d'une part la construction de ressources sémantiques, avec identification de termes, définition de concepts et leurs relations sémantiques et création de thésaurus, lexiques et/ou ontologies ; d'autre part, l'annotation descriptive des documents et la génération de représentations des textes qui font le lien entre les documents et les ressources sémantiques (par exemple, repérage des termes et concepts mentionnés dans les documents).

Le deuxième lot, à destination des consultants, est le **prototypage d'outils de recherche sémantique** pour améliorer l'interface de recherche d'information (saisie de requête, consultation et filtrage des résultats de la requête), à travers une auto-complétion plus intelligente, la structuration et l'enrichissement des résultats de recherche (extraction de descripteurs facilitant l'interprétation des résultats, regroupements automatiques des résultats de recherche) et une recherche sémantique dynamique soit pour étendre une requête trop spécifique (recherche des idées liées, voisines ou un peu plus génériques) soit pour l'affiner (désambiguïsation, choix d'un sens particulier).

Le troisième lot, également à destination des consultants, vise à explorer de nouvelles formes de navigation dans les contenus en utilisant des techniques de **data-visualisation**. Deux formats sont privilégiés : visualisation géographique des documents et visualisation de graphes de données.

Enfin le dernier lot, destiné aux rédacteurs, doit permettre d'améliorer la production de contenus à travers **l'attribution automatique de thèmes aux documents**.

### C. Choix de solutions technologiques

Les besoins requièrent des compétences en traitement automatique du langage, à la fois linguistiques, pour répondre aux problématiques sémantiques, et statistiques pour traiter la volumétrie. Ils nécessitent également la mise en place d'Interfaces Homme-Machine, indispensables pour aboutir à des maquettes testables par des utilisateurs finaux.

Plusieurs acteurs académiques et du marché spécialisés en traitement automatique du langage sont sélectionnés, avec différentes approches : 1) analyse syntaxique et distributionnelle de la société Synomia ; 2) approche ontologique de la société Viseo ; 3) approche par apprentissage automatique du laboratoire LIUM pour la partie algorithmique et de la société Lincoln pour l'intégration des traitements dans une interface.

### D. Prototypage et évaluation

Les réalisations ont été échelonnées dans le temps du fait des contraintes projet (montage des collaborations, contraintes budgétaires, mise en œuvre de traitements de complexité et maturité variables) :

- octobre 2017 à février 2018 : maquettes de recherche sémantique et de visualisation de données par approches syntaxique (société Synomia) et ontologique (société Viseo), suivies de tests utilisateurs communs ;
- septembre 2017 à juillet 2019 : maquettes de regroupements automatiques des résultats de recherche et de prédiction thématique par apprentissage automatique (laboratoire LIUM / société Lincoln).

#### 1) Approche syntaxique pour la recherche sémantique et la visualisation de graphes thèmes-termes

La première approche expérimentée repose sur une technologie propriétaire de Synomia, une société spécialisée en Traitement Automatique du Langage et en moteur de recherche sémantique. Basée sur un algorithme d'analyse syntaxique, elle extrait les termes, puis elle réalise une analyse complète des relations entre termes en couplant analyse syntaxique et analyse distributionnelle [5]. Elle génère ainsi un réseau terminologique qui reflète la structure des concepts présents dans le corpus. L'analyse est rapide et complètement automatisée, sans intervention manuelle experte pour prédéfinir les patrons recherchés.

Dans le corpus SNCF, 2,8 millions d'occurrences de termes simples ou complexes ont été extraits de 12330 textes de prescription. Les termes extraits sont précis et bien formés syntaxiquement, avec des syntagmes nominaux, verbaux, adjectivaux (*circulation des trains*, *circulation des trains de travaux*, *arrêt de la circulation des trains*, *interdire la circulation des trains*). Le corpus a été structuré en sous-corpus à partir des 13 principaux thèmes, avec identification automatique des termes spécifiques du corpus complet (*cf.* Fig. 1) et de chaque sous-corpus (*cf.* Fig.2).


Fig. 1. Syntagmes nominaux les plus fréquents extraits automatiquement par la solution d'analyse syntaxique dans tout le corpus, avec nombre total d'occurrences et nombre de documents les contenant


Fig. 2. Syntagmes nominaux les plus fréquents extraits automatiquement par la solution d'analyse syntaxique dans le sous-corpus des documents produits par les Ressources Humaines, avec nombre total d'occurrences et nombre de documents les contenant

Cette représentation des textes a été mise à disposition à travers deux interfaces : une interface d'exploration du réseau terminologique (*cf.* Fig. 3), appelée par la suite « SYNOMIA SETI », et une plateforme de consultation du corpus, « SYNOMIA TILE », qui permet l'exploration par index de termes ou directement dans les textes. Cette dernière plateforme met à disposition les outils utilisés habituellement par les analystes de la société Synomia, sans adaptation particulière de l'interface : elle comporte des outils avancés et n'a pas fait l'objet d'évolutions ergonomiques pour être plus adaptée au profil des utilisateurs finaux. L'enjeu de ces deux interfaces était de permettre à des utilisateurs d'explorer les contenus et d'évaluer la qualité de l'information, autrement dit la capacité de la technologie à extraire et à accéder au contenu sémantique ; l'ergonomie de l'interface pour les utilisateurs finaux n'était pas au cœur de la démarche. Les deux interfaces offrent de l'auto-complétion intelligente (constituée de suggestions cohérentes d'un point de vue syntaxique et sémantique), ainsi que des principes de navigation dynamique par suggestion de nouveaux termes sémantiquement liés et mise à jour de la recherche en fonction des termes choisis.


Fig. 3. Recherche du mot *procédure* dans la maquette d'exploration du réseau terminologique, avec les premiers termes spécifiques dans les corpus TR04 "Réalisation du trafic" et S08 "Manœuvre – évolutions"

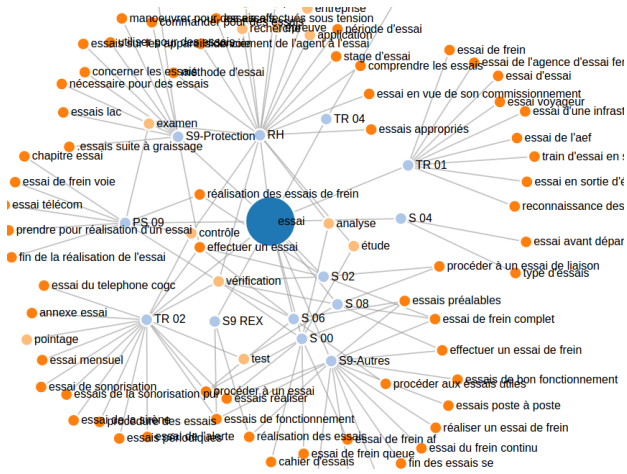


Fig. 4. Graphe thème-terme centré sur le mot *essai*, où les thèmes sont représentés par leurs codes d'une à deux lettres et deux chiffres

Le réseau terminologique a également été exploité pour constituer un graphe terme-thème (*cf.* Fig.4), afin de visualiser les interdépendances entre thèmes et termes caractéristiques de chaque thème.

### *2) Approche ontologique pour la recherche sémantique, la visualisation cartographique et par graphe de concepts*

La deuxième approche expérimentée est basée sur des méthodes issues du web sémantique et repose sur des ontologies. Elle a été réalisée par la société Viseo et son pôle Recherche et Innovation, détenteur d'expertise en Traitement Automatique du Langage (machine learning, web sémantique) et capable de répondre également aux besoins d'interfaces ergonomiques à travers des compétences de designers et de spécialistes de l'interaction.

La première étape a été la création de ressources sémantiques à partir des termes extraits par Synomia (sans leurs relations), d'acronymes puis de liste de gares. Trois ontologies ont été créées :

- une ontologie globale qui représente les relations terme générique – termes spécifiques (par exemple, *consigne* a pour termes spécifiques *consigne neige* et *consigne de protection*), construite automatiquement à partir des termes issus de la solution Synomia et de scripts informatiques ;
- une ontologie spécifique à un sous-corpus concernant les procédés trains-travaux (corpus « S9 »), avec une description riche des principaux concepts des documents S9 et des différents types de relations (par exemple, la relation « réalise » lie un agent (conducteur, agent circulation, …) à une procédure) ;
- une ontologie d'association entre documents S9 et ontologie S9, pour faire le lien entre les concepts présents dans les textes et ceux décrits dans l'ontologie S9 ; elle a été constituée automatiquement à partir des individus de l'ontologie précédente (ontologie S9), des termes Synomia et de scripts informatiques pour faire l'appariement.

Ces ressources ont été exploitées dans une application web (*cf.* Fig. 5), appelée « MEAR ». L'application a été développée spécifiquement pour le projet, avec du design et une approche centrée utilisateur pour une interface plus intuitive et conçue pour les besoins métiers spécifiques du projet. Elle permet de répondre aux besoins de recherche sémantique avec :

- une **recherche sur un ou plusieurs mots-clés, avec suggestions d'autocomplétion sémantique** : la recherche n'est possible que sur les concepts de l'ontologie ;
- des **résultats de recherche ordonnés selon un score de pertinence**. Le score utilisé est le BM25, un score classique en recherche d'information [6]. Il est basé sur les fréquences des concepts dans les phrases et une modélisation « sac de mots » du texte.
- des **suggestions de recherche construites sur les relations entre concepts** si aucun résultat ne convient, avec proposition au maximum de 6 concepts plus spécifiques et un concept générique

tirés de l'ontologie générale. Par exemple, si le terme générique *alimentation* est saisi, l'outil suggère *alimentation en courant*, *alimentation en eau* ou *alimentation de caténaire*.

- des **suggestions de recherche extraites algorithmiquement des résultats de recherche** afin d'affiner celle-ci à l'aide des concepts suggérés. Par exemple, pour la recherche *alimentation*, l'outil propose *sectionneur*, *tension*, *caténaire* et *alimentation électrique*. Ces suggestions sont construites sur la base du score de pertinence BM25 : dans la liste de résultats, les concepts identifiés comme les plus pertinents pour la liste de résultats (c'est-à-dire qui ont le meilleur score) sont proposés à l'utilisateur qui peut les sélectionner et ainsi filtrer les résultats. Le processus est itératif.


Fig. 5. Maquette de recherche sémantique et fonctionnalités disponibles sur les résultats de recherche

L'application permet également deux formes de visualisation de données. La première est une **visualisation cartographique** (*cf.* Fig. 6) : celle-ci est construite sur la liste de lieux intégrés aux concepts de l'ontologie, sur un fichier de données SNCF disponible en open data avec les libellés de gare et leurs coordonnées géographiques. Les lieux et leurs désignations possibles sont intégrés à l'ontologie, détectés automatiquement dans les textes et reliés aux coordonnées géographiques. La maquette permet de : 1) voir sur une carte le nombre de documents répondant à une recherche, répartis par zone géographique ; 2) voir sur une zone géographique avec possibilité de zoom l'emplacement des gares associées aux documents issus d'une recherche, et un aperçu des documents correspondants. ; 3) voir tous les documents associés à une zone géographique.

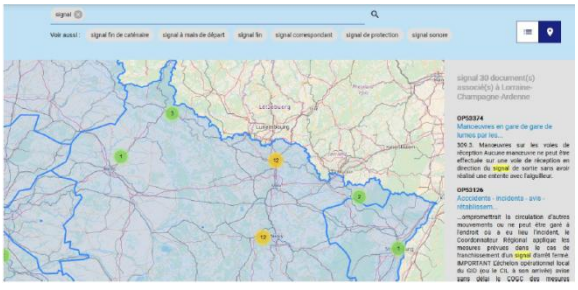

Fig. 6. Maquette de recherche cartographique (*copie d'écran*)

La deuxième forme de visualisation est une **visualisation par graphe de concepts** (*cf.* Fig.7) exploitant les concepts et relations répertoriés dans les ontologies, avec affichage de la liste de concepts, de leurs liens sémantiques, d'un aperçu des documents associés au concept de départ, avec possibilité de naviguer dans le graphe de concepts en cliquant successivement sur les concepts et avec accès à la liste complète des documents associés à un concept.

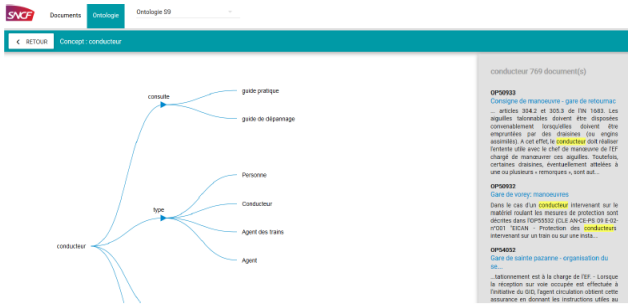


Fig. 7. Recherche par graphe de concept (*copie d'écran*)

### *3) Approche par apprentissage automatique pour la prédiction thématique et le clustering des résultats de recherche*

La troisième expérimentation vise à répondre à deux besoins : 1) structurer les résultats de recherche de façon automatique, en proposant des regroupements naturels des textes en fonction de leur contenu sémantique et ainsi guider l'utilisateur final pour qu'il converge plus facilement vers le résultat attendu ; 2) prédire automatiquement le ou les thèmes possibles d'un document afin d'aider les rédacteurs et administrateurs de textes à affecter le bon thème à un document, celui-ci étant recherché dans un plan de classement complexe à maîtriser car constitué de 628 éléments organisés selon une arborescence à 4 niveaux.

L'approche mise en œuvre repose sur des traitements TAL et d'apprentissage artificiel développés par le laboratoire LIUM (Laboratoire d'Informatique Université du Mans), spécialisé dans le traitement du langage. Ces traitements sont intégrés au sein d'une interface développée par la société Lincoln, spécialisée dans le traitement et l'analyse de données. Ils ont été réalisés sur un corpus de 7029 textes.

D'abord, des **prétraitements textuels et linguistiques** ont été appliqués au corpus de travail, avec une analyse et un nettoyage des données (gestion des problèmes de qualité dus au format initial (PDF)), puis un enrichissement linguistique comprenant de la lemmatisation, la détection de mots-outils, la détection de groupes nominaux et la normalisation de noms de lieux et dates.

Le **regroupement des résultats de recherche** est réalisé par apprentissage non supervisé (**clustering**). L'algorithme des K-moyennes est appliqué au corpus vectorisé des textes obtenus en réponse à une recherche de façon parallélisée pour un nombre de clusters variant de 2 à 15.

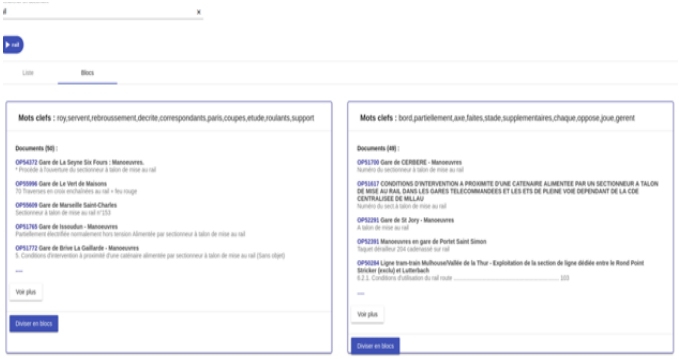

Fig. 8. Interface de clustering des résultats de recherche (*copie d'écran*)

Chaque cluster est étiqueté avec les mots ayant les valeurs de contraste les plus élevées. Les algorithmes de clustering sont implémentés dans une application web (*cf.* Fig. 8) qui propose à l'utilisateur une recherche par mot-clé, effectue la recherche dans la base documentaire à l'aide du moteur Elasticsearch, retourne une liste de résultats à laquelle les algorithmes de clustering sont appliqués. Elle permet ensuite de visualiser les regroupements de textes avec leurs étiquettes, puis de sélectionner l'un des groupes de textes pour lequel le processus est réitéré, permettant ainsi des affinements successifs.

La **prédiction de thèmes** est réalisée par apprentissage supervisé (classification). Plusieurs représentations textuelles du corpus ont été considérées et différents algorithmes classification automatique ont été appliqués. Le protocole expérimental est décrit plus en détail dans la partie IV. L'algorithme de classification est intégré à une application web (*cf.* Fig. 9) qui permet de charger un document depuis le poste de l'utilisateur, calculer le ou les thèmes possibles avec un score de probabilité, afficher les résultats à l'utilisateur avec possibilité de paramétrer les représentations ou l'algorithme utilisé.


Fig. 9. Interface de prédiction thématique. (*copie d'écran*)

## IV. EVALUATION ET RESULTATS

### A. *Evaluation des maquettes de recherche sémantique et de visualisation des données*

Afin d'évaluer les approches syntaxique et ontologique, des tests utilisateurs ont été mis en place.

#### *1) Protocole des tests utilisateurs*

##### *a. Objectifs et hypothèses des tests utilisateurs*

Les tests utilisateurs se sont inspirés de travaux s'intéressant à la compatibilité du système (ou logiciel) avec le mode de raisonnement de l'utilisateur, et interrogeant les notions d'utilisabilité ou d'utilité. Les tests réalisés avaient ainsi pour objectif d'évaluer la pertinence et l'intérêt des fonctionnalités mises en place dans les maquettes de recherche sémantique, l'application web basée sur une approche ontologique et les deux interfaces d'exploration et de consultation des verbatims basées sur l'approche terminologique, en évaluant leurs points forts et leurs points faibles. Les ressources utilisées dans les maquettes n'ont pas été testées indépendamment, bien que parfois la distinction ressource/fonctionnalité soit difficile à établir.

Trois hypothèses ont été testées :

- Les technologies sémantiques enrichissent les fonctionnalités des moteurs de recherche classiques
- Les technologies sémantiques apportent des filtres pertinents et utiles pour les utilisateurs
- La perception de la complexité des technologies sémantiques est fonction de l'expertise de l'utilisateur en recherche d'information

##### *b. Environnement et durée des tests*

Les passations se sont effectuées dans l'environnement de travail quotidien des testeurs. Cela signifie que dans la mesure du possible, les tests ont été réalisés à partir des postes de travail des agents, individuellement ou en groupe. En ce qui concerne le lieu, des salles de réunions ont également été utilisées, lors des passations en groupe notamment. Dans ce cas, les interactions sont davantage sollicitées, et mettent les testeurs dans un climat de confiance, d'échange entre pairs, et donc dans une situation naturelle [11].

Le temps de passation a été déterminé en fonction de deux critères : le temps de disponibilité des agents et le temps de concentration nécessaire. Ainsi, les tests ont été élaborés de manière à durer entre 40 minutes et 1 heure.

##### *c. Nombre de participants*

Plusieurs études en ergonomie démontrent qu'une dizaine d'utilisateurs suffisent à faire remonter la majeure partie des problèmes d'utilisabilité. Ce type de démarche, plutôt qualitatif, permet d'effectuer des évaluations en profondeur, et fournir de nombreux éléments de réflexion [12]. Par ailleurs, dans le domaine de l'ergonomie, Nielsen et son équipe [13, 14] montrent que le pourcentage d'erreurs identifiées augmente rapidement jusqu'à environ cinq personnes, pour se stabiliser autour de huit à dix personnes. Au-delà, l'ajout d'évaluateurs augmente très peu la proportion d'erreurs détectées. Pour ces raisons, 12 participants ont évalué les maquettes, avec des profils variés : hommes et femmes, experts sécurité ou non, experts du corpus « S9 ».

##### *d. Tâches et consignes*

La tâche réalisée consistait en une tâche de recherche d'information « classique ». L'objectif étant que les utilisateurs soient dans le contexte de recherche le plus naturel possible, ces derniers avaient pour première consigne de construire leurs propres scénarios de recherche, c'est-à-dire deux à trois cas d'usages liés à leur activité. Les testeurs effectuaient ensuite, au choix, un ou plusieurs scénarios de recherche sur chacune des maquettes. L'ordre de test des maquettes était contraint. De cette façon, l'évaluateur pouvait contrôler les tests : durée de passation, actions réalisées, *etc*.

Nous nous sommes appuyés sur une approche explicite pour recueillir le ressenti des utilisateurs. Lors de leur recherche, il leur a été demandé d'oraliser, c'est-à-dire de commenter leur démarche. Enfin, à la fin de chaque test sur une maquette, l'utilisateur devait répondre à un questionnaire de satisfaction.

##### *e. Métriques d'évaluation*

Trois sources de données ont été conservées pour l'évaluation :

1- Les observations de l'évaluateur

Lors de la passation des tests, l'évaluateur a un rôle d'observateur. Il relève tout ce qui a trait de l'interaction entre l'humain et le système, et relève notamment : le temps nécessaire à la réalisation des tâches, les réactions

immédiates et spontanées des utilisateurs, la manière d'interagir avec les outils (clics souris, ou plutôt interactions via le clavier), les comportements, actions qui ont apporté de la satisfaction, de la frustration, de la réussite, des échecs, *etc*. Ces éléments renseignent sur l'utilisation et l'utilisabilité de certaines fonctionnalités, et également sur la perception qu'ont les utilisateurs des maquettes.

2- Les réponses aux questionnaires

A la fin de chaque passation sur une maquette, un questionnaire de satisfaction était complété. Ce questionnaire était centré sur les fonctionnalités d'auto-complétion, sur les filtres de recherche mis en place, ainsi que sur l'interface graphique des maquettes. Des échelles de notation ont été proposées pour ces fonctionnalités, ce qui a permis d'évaluer la satisfaction ou l'insatisfaction des utilisateurs. Des questions ouvertes permettaient également aux utilisateurs de partager leur ressenti.

3- Les retours du débriefing final

À la fin de la passation des tests, un débriefing oral a permis de synthétiser les retours des utilisateurs, et également d'évoquer leurs attentes vis-à-vis d'un système de recherche d'informations. Les informations recueillies ont ensuite été rassemblées et croisées.

*2) Résultats*

Les retours des utilisateurs portaient sur la valeur des fonctionnalités dans la recherche d'information, et également sur les interfaces proposées.

*a. Retours sur les principales fonctionnalités*

**Autocomplétion** : la fonctionnalité guide et suggère d'autres idées de recherche, les suggestions sont bien formées, conformes aux attentes utilisateurs et pertinentes, dans l'approche syntaxique et ontologique. Elles améliorent l'existant, où les suggestions étaient jugées souvent peu pertinentes ou peu intéressantes.

**Recherche multi-mots** : dans l'approche ontologique, l'utilisateur est obligé de choisir un ou plusieurs des concepts proposés. Il peut saisir successivement plusieurs concepts dans la barre de recherche. Appréciée pour la même raison que pour l'autocomplétion, cette fonctionnalité a soulevé des réserves du fait d'une recherche contrainte, sans possibilité de saisie libre en langage naturel et avec un fonctionnement inhabituel.

**Suggestions de mots-clés** : pour les suggestions basées sur le réseau terminologique (SYNOMIA TILE et SETI), ou sur les relations de l'ontologie de MEAR, cette fonctionnalité est apparue comme utile pour guider et affiner la recherche, avec des mots-clés satisfaisants, mais manquait de visibilité dans l'interface. Pour les suggestions extraites algorithmiquement des résultats de recherche de MEAR, cette fonctionnalité a apporté peu de valeur car les mots-clés proposés n'étaient pas toujours pertinents, ou le résultat de la sélection d'un mot-clé n'était pas probant.

**Affichage du contexte du mot-clé** : absente de l'outil existant, cette fonctionnalité qui affiche le contexte d'apparition des mots recherchés est très appréciée car elle donne un premier niveau d'accès au mot dans son contexte sans avoir besoin d'ouvrir le document. L'affichage d'expressions même complexes et la pertinence des contextes sont ressortis comme un point fort de l'approche syntaxique.

**Ordre et tri des résultats** : les maquettes offraient différentes fonctionnalités permettant de retrier ou filtrer les résultats obtenus (tri par score, alphabétique ou titre dans « MEAR » ; filtres ou tris par sous-corpus, en fonction de la fréquence d'apparition, de la nature grammaticale des mots). D'une manière générale, les utilisateurs n'ont pas essayé de modifier l'ordre d'affichage des documents et l'utilisation des filtres n'étaient pas toujours intuitive. Le filtrage par sous-corpus de Synomia Tile et sa performance ont été appréciés.

Au niveau de la visualisation des résultats, la **recherche ou l'affichage des documents par cartographie** ont été bien accueillis, et leur intérêt pour les métiers opérationnels a été relevé. La **navigation par graphe** (au niveau de l'ontologie) a été perçue comme potentiellement utile, une fois le mode de navigation compris. Cette fonctionnalité est apparue comme réservée à des experts de la recherche d'information.

Au global, les fonctionnalités jugées prioritaires sont parfois héritées de l'utilisation des moteurs de recherche grand public type Google. Elles correspondent aux fonctionnalités de base, nécessaires, que les utilisateurs attendent. Elles concernent plusieurs phases dans la recherche d'information. Lors de la requête dans la barre de recherche, il apparaît primordial que les outils disposent de la recherche par mot-clé, l'auto-complétion, la possibilité de formuler en langage naturel, la combinaison mots-clés proposés et mots-clés non proposés par l'outil, la gestion des acronymes. Une fois la requête tapée, les filtres proposés pour trier les documents doivent permettre de sélectionner des textes selon leur niveau d'application et/ou selon leur type et de croiser plusieurs paramètres (filtres croisés) : par exemple le métier et le niveau d'application.

*b. Synthèse*

En majorité, les utilisateurs ont préféré la maquette de Viseo (MEAR), bien que ses fonctionnalités n'aient pas été jugées comme les plus pertinentes. Cela montre que l'intérêt pour un système plutôt qu'un autre n'est pas fonction seulement de la performance de ses fonctionnalités, mais fait appel aux notions d'utilisabilité et d'utilité mentionnées en première partie. En effet, les fonctionnalités d'autocomplétion, la présentation du contexte du mot-clé, la possibilité de filtrer les résultats selon un sous-ensemble de la documentation sont des fonctionnalités qui ont été évaluées comme utiles et plus pertinentes dans les maquettes Synomia que dans la maquette de Viseo. Cependant, d'une part, l'interface trop complexe de Synomia TILE et d'autre part, le mode de navigation non habituel de Synomia SETI, ont rendu ces maquettes plus difficiles à prendre en main. Finalement, la maquette ayant au niveau de l'interface le mode de fonctionnement le plus proche des moteurs de recherche grand public a rencontré le plus de suffrage.

Les tests utilisateurs ont par ailleurs permis de valider l'hypothèse selon laquelle plus un utilisateur est expert en recherche d'information, plus il utilisera des filtres de recherche avancés. En ce sens, les technologies sémantiques enrichissent les moteurs de recherche classiques. En revanche, la pertinence et l'utilité des différents filtres pour les utilisateurs n'a pas été démontrée : les limites de nos tests (nombre de participants, limites inhérentes aux maquettes) n'ont pas permis de valider cette hypothèse.

*B. Approches par apprentissage automatique*

Les secondes approches avaient pour objectif de tester des algorithmes de Machine Learning sur le corpus spécialisé SNCF.

*1) Évaluation des plongements lexicaux*

Avant l'application des algorithmes de classification et de clustering, le corpus a été prétraité et transformé en vecteurs de mots, via la méthode de plongements lexicaux. La modélisation du vocabulaire dans cet espace a fait l'objet d'une évaluation humaine originale [9].

*a. Une évaluation réalisée par le métier*

Afin d'évaluer la qualité de l'espace de représentation généré, un protocole d'évaluation humaine a été mis en place. Celui-ci consistait en l'évaluation empirique de la similarité lexicale par plusieurs experts SNCF, personnes ayant les compétences nécessaires pour juger de la proximité sémantique de deux mots dans le cadre dédié de SNCF. Au final, neuf experts ont été sollicités, (responsables ou chefs en poste depuis en moyenne 10 ans), exerçant à des postes variés : documentation métier, sécurité système, qualité et performance, organisation de travaux, *etc*.

*b. Tâche à réaliser*

L'évaluation consistait à valider la pertinence de l'association de deux mots donnés. Par exemple l'association de « train » et « wagon » est pertinente tandis que l'association de « billet » et « passage à niveau » ne l'est pas. Afin de simplifier le travail des experts, et pour minimiser le caractère subjectif de l'évaluation humaine, un système binaire a été mis en place : l'association des mots pouvait être pertinente ou non pertinente. Cela permettait de plus de couvrir suffisamment le corpus et de garantir la significativité statistique des résultats.

*c. Sélection du vocabulaire à évaluer*

Les mots évalués ont été regroupés et proposés aux experts selon 4 catégories : mots issus du lexique SNCF ; acronymes SNCF polysémiques ; acronymes SNCF non polysémiques ; N-grammes fréquents dans le corpus. Deux fois 20 mots de chacune des catégories ont été présentés aux experts : 20 parmi les mots les plus fréquents, et 20 parmi ceux dont la fréquence se situe au niveau de la médiane de la distribution. Ainsi chaque expert a annoté 160 mots auxquels étaient associés les 6 mots les plus proches dans l'espace de représentation généré, soit 960 paires de mots à juger comme étant pertinentes ou non.

*d. Interface d'annotation mise en œuvre*

Une plateforme regroupant plusieurs formulaires web a été développée par le LIUM. Pour chacune des 4 catégories, un tableau contenant le mot et ses 6 voisins était proposé. Les 6 voisins étaient associés à une case à cocher. Si l'expert estimait que le mot voisin n'était pas en relation avec le mot courant alors il cochait la case. S'il estimait que l'association des deux mots était correcte, il n'avait aucune action à faire. De plus, une fonctionnalité permettant d'indiquer que le mot n'est pas connu était proposée afin de distinguer une association qui ne serait pas pertinente d'une association qui ne peut être évaluée car au moins l'un des mots n'est pas connu.

*e. Résultats*

D'un point de vue quantitatif, les résultats montrent de fortes différences entre les deux catégories d'acronymes d'une part, et les catégories lexique et ngram d'autre part. Le pourcentage d'associations inconnues est très élevé dans les acronymes (40 à 50%), et il est faible dans les autres catégories (10%). Cette différence se retrouve au niveau de l'accord inter-annotateurs, plus faible sur les catégories d'acronymes que sur les autres catégories. Par ailleurs, les éléments lexicaux les plus fréquents sont mieux évalués (moins d'associations inconnues et un meilleur accord inter-annotateurs). Enfin, 30 à 40% des paires évaluées ont été jugées non pertinentes ce qui montre la difficulté de la tâche.

D'un point de vue qualitatif, les retours des participants ont permis de mettre au jour les limites des plongements lexicaux sur de la documentation métier. La catégorie des acronymes a été jugée comme étant la plus compliquée à évaluer. Étant donné le grand nombre d'acronymes polysémiques au sein du groupe, ce constat n'est pas surprenant. Souvent l'acronyme n'était pas connu de l'évaluateur, et surtout il était impossible pour ce dernier, hors contexte, de lui donner un sens.

*f. Synthèse*

L'évaluation mise en place a montré la difficulté à modéliser le vocabulaire spécialisé et notamment les acronymes. L'évaluation humaine reste difficile, les niveaux d'accord inter-annotateurs sont bas, et de nombreux mots sont inconnus, ou impossibles à évaluer sans contexte.

*2) Évaluation de la classification des documents SNCF*

L'objectif des algorithmes de classification était de prédire le thème d'un document donné. Le LIUM a également proposé un ensemble de mots-clés pour chaque document.

*a. Protocole expérimental*

Plusieurs protocoles ont été mis en place pour mieux évaluer les performances de classification. Les expériences menées ont fait varier : 1) les représentations des documents ; 2) les algorithmes de classification ; 3) les mesures d'évaluation. De nombreuses combinaisons ont été testées mais nous ne présentons dans la suite que les résultats les plus pertinents.

**Représentation d'un document**. La représentation finalement la plus pertinente est celle prenant en compte la forme de surface des mots et la suppression des mots outils. Chaque document est ensuite représenté par un vecteur creux du vocabulaire sélectionné (environ 10k mots) pondéré par ses scores *tf-idf*.

**Définition de mots-clés**. En plus des résultats de classification pour chaque nouveau document, l'ensemble des 50 mots clés ayant la mesure de *tf-idf* la plus élevée dans ce document a été proposé.

**Classifieurs**. Plusieurs classifieurs ont été testés via la libraire Scikit-Learn : les machines à vecteurs supports (SVM), les forêts aléatoires (random forest), le perceptron simple ou multicouches (MLP), le classifieur naïf bayésien (naive bayes) et la régression logistique. Un classifieur de boosting de gradient a aussi été testé à l'aide de la libraire XGBoost. Après analyse des performances issues de plusieurs expériences, seulement 3 classifieurs ont été conservés: XGBoost, SVM et la régression logistique.

**Corpus d'apprentissage et test**. Les résultats présentés ont tous été obtenus au moyen d'une évaluation par validation croisée stratifiée sur 5 plis. Ainsi, les performances obtenues sont révélatrices du comportement

qu'aura le système pour classer tout nouveau document dont la classe serait parmi celles du corpus utilisé pour la construction des modèles de classification.

**Ensemble de thèmes hiérarchiques**. L'association thème-document est issue de la classification hiérarchique SNCF sur 4 niveaux. Des statistiques générales montrent que la répartition des thèmes sur les 7k documents du corpus est très déséquilibrée sur tous les niveaux. Par exemple pour le niveau 1 comprenant 8 classes, 70% des documents appartiennent à une seule classe. Au niveau 2 comprenant 49 classes, c'est 36% des documents qui appartiennent à une seule classe. Ceci va forcément rendre l'apprentissage des classes peu représentées difficile. C'est pourquoi plusieurs mesures d'évaluation ont été proposées.

**Mesures d'évaluation**. Les mesures d'évaluation choisies sont celles classiques dans le domaine de la recherche d'information : la *précision*, le *rappel* et la *f-mesure (combinaison des deux précédentes)*. Afin d'analyser plus finement les résultats, deux variantes de ces mesures ont été utilisées : 1) les ***micro-mesures*** : elles permettent de comptabiliser chaque document avec le même poids et ainsi analyser sans distinction de classes, les performances du système ; 2) les ***macro-mesures*** : les mesures de *précision* et *rappel* sont calculées d'abord par classe puis une moyenne est faite. De manière générale, les résultats sont présentés en micro-mesure mais il peut être intéressant de regarder aussi le comportement du classifieur au niveau de chacune des classes via la macro-mesure.

Une mesure supplémentaire est introduite : la f-mesure à n. Elle permet de calculer la f-mesure des modèles lorsque les n classes de scores les plus élevés sont considérées et non plus seulement la classe qui obtient le score le plus élevé. Cette mesure est choisie car les classifieurs proposent pour chaque document nouvellement classé l'ensemble des scores associés à chacune des classes. Ainsi, il est possible de voir très rapidement les premières classes, ainsi que leur probabilité d'association au document et finalement choisir la classe 2 ou la classe 3 au lieu de celle obtenant le meilleur score.

*b. Résultats*

**Niveau 1 : Classification selon 8 classes.** Les résultats de classification selon ce niveau sont donnés dans le tableau Fig. 10.

| niveau 1 : 8 classes | | | | |
|---|---|---|---|---|
| | micro f.mes ± int. conf | macro fmes. | temps app. | temps test |
| xgb | 98.5 ± 0.3 | 48.9 | 5 min 1 sec | 46 sec |
| SVM | 98.1 ± 0.3 | 49.2 | 6 sec | 5 sec |
| log.reg. | 97.9 ± 0.3 | 43.2 | 26 sec | 5 sec |

Fig. 10. Performances de classification niveau 1.

Nous voyons que les résultats de classification sont très bons en micro-f-mesure. SVM se démarque avec une très bonne performance obtenue avec un temps d'apprentissage et de test les plus bas. Il est également intéressant de remarquer que si sa micro-f-mesure est de 98.1%, c'est à dire lorsqu'on considère que tous les documents ont le même poids, la valeur de macro-f-mesure, elle, s'effondre à 49.2%. Ce résultat n'est pas étonnant lorsque l'on garde à l'esprit que la répartition des classes sur le corpus est très déséquilibrée.

**Classification sur les niveaux suivants.** Au niveau 2, les résultats baissent d'environ 5 points par rapport au niveau 1. Cela s'explique évidemment par une augmentation du nombre de classes qui doivent maintenant être modélisées avec toujours le même nombre de documents. La baisse des performances continue sur les niveaux 3 et 4. Sur le dernier niveau, les meilleurs résultats sont obtenus avec SVM pour atteindre une micro-f-mesure de 84.8%.

Les résultats de f-mesure à n permettent alors d'évaluer si la bonne classe se trouve parmi les n classes ayant les plus hauts scores. Le tableau ci-dessous (*cf.* Fig.11) présente les résultats en considérant les 3 classes de plus hauts scores.

| F-mesure à n=3 | | | | |
|---|---|---|---|---|
| | #Classes | f-mesure (%) | | Tps app. (sec) |
| | | **micro** | **macro** | |
| Niveau 2 | 49 | 98.1 | 46.2 | 08 |
| Niveau 3 | 107 | 96.8 | 45.0 | 10 |
| Niveau 4 | 155 | 94.3 | 41.5 | 12 |

Fig. 11. Performances avec considération des 3 classes de plus hauts scores pour les niveaux 2 à 4.

On obtient alors des résultats qui sont très exploitables notamment toujours en prenant en compte le cadre de déséquilibre des classes. La macro-f-mesure, quant à elle, augmente sensiblement avec pour SVM un gain constaté de 15 points sur le niveau 2 et 10 points sur les niveaux 3 et 4.

*c. Synthèse des évaluations*

Le nombre de classe et le déséquilibre de leur répartition dans le corpus impacte les résultats. Néanmoins, de très bons résultats sont obtenus et les expériences nous indiquent que les classifieurs opèrent des choix très pertinents. Si leur premier choix n'est pas le bon, la classe attendue est néanmoins positionnée comme un prétendant de forte probabilité. La classification ayant pour but d'aider les rédacteurs à classer plus facilement un document, avoir le choix entre 3 classes semble tout à fait raisonnable et les résultats montrent que la classe de référence est positionnée à plus de 95% dans les 3 meilleures classes.

*3) Clustering thématique*

Le clustering a pour objectif d'organiser thématiquement les documents retournés pas une requête utilisateur sur le moteur de recherche documentaire. On associe l'algorithme d'apprentissage non supervisé des k-moyennes à des méthodes statistiques pour sélectionner le nombre optimal de groupes thématiques (cluster) et extraire des mots clés représentant chaque cluster [10].

*a. Clustering et extraction de mots-clés*

La méthode utilisée permet d'évaluer la qualité de la partition obtenue en maximisant l'indice EC. En appliquant successivement l'algorithme des k-moyennes et le calcul de l'indice EC pour un nombre de clusters variant de 2 à 15, on sélectionne la meilleure partition (clusters compacts et bien séparés entre eux). Ensuite, les mots-clés les plus prépondérants statiquement dans chaque groupe thématique sont utilisés pour représenter les documents contenus dans chacun des clusters.

*b. Résultats de la partition thématique*

Les expériences sont réalisées sur 46 requêtes ayant retournées chacune 99 documents. Le processus de clustering a été appliqué à chaque requête. Le point de variation des différentes expériences est la représentation des documents :

pour chaque expérience, seuls les 1 000 mots ayant obtenus la valeur de *tf-idf* la plus importante ont été retenus. Les documents ont été représentés de manières différentes:

- Par les lemmes, en conservant les mots ne pouvant être lemmatisés[1], en excluant les mots-outils ainsi que les dates, lieux et codes
- Par les lemmes en excluant les mots-outils ainsi que les dates, lieux et codes
- Par les lemmes en excluant les mots ne pouvant être lemmatisés et en ajoutant une liste de mots outils personnalisée à la liste de mots-outils classiques[2]

Les deux représentations de documents qui conservent les mots ne pouvant pas être lemmatisés donne des résultats avec beaucoup de termes non porteurs de sens pour les documents présents dans chaque cluster et notamment des prénoms tels que Jean et Charles correspondant à des noms de gares (Bordeaux Saint-Jean, Marseille Saint-Charles).

L'exclusion des termes non lemmatisés et l'ajout d'une liste de mots-outils permet de réduire la présence de codes, dates et noms de lieux dans les mots-clés utilisés pour représenter les clusters. Cependant, la cohérence sémantique des résultats reste difficile à évaluer. Ce manque de cohérence peut être la conséquence de prétraitements trop peu restrictifs (données bruitées) ou encore de la représentation des mots utilisée (*tf-idf*).

## V. CONCLUSION ET PERSPECTIVES

Les travaux, réalisés sur environ 10 000 textes prescrits, ont permis d'expérimenter plusieurs technologies, dont certaines à l'état de l'art, afin de :

- Créer des ressources sémantiques (terminologies, ontologies, représentations condensées des textes) à travers des prétraitements TALN classiques (lemmatisation, liste d'exclusion, …) et plus avancés (extraction terminologique, plongements lexicaux) ; ces ressources ont été exploitées dans les maquettes pour gagner en performance et en qualité informationnelle ;
- Maquetter des applications web à destination des rédacteurs et des utilisateurs finaux pour : 1) disposer d'une recherche sémantique à travers de l'auto-complétion sémantique, des suggestions de recherche issues d'ontologies, d'un réseau terminologique ou de méthodes algorithmiques, et une structuration automatique des résultats avec mise à jour dynamique obtenue à partir d'algorithmes de clustering, alternative aux classements métiers manuels ; 2) visualiser les données, à travers des graphes thèmes-termes, des graphes de concepts et de documents liés et une représentation cartographique obtenue par extraction des mentions de lieux ; 3) aider les rédacteurs à attribuer des thèmes aux documents par apprentissage des thèmes saisis et prédiction de thèmes sur de nouveaux documents

Les différentes solutions ont fait l'objet de tests utilisateurs qualitatifs et d'évaluation experte, pour aboutir à un bilan technologique et d'usage.

Pour les méthodes à base d'apprentissage artificiel, la prédiction est prometteuse, le clustering a un potentiel qui demande à être investigué davantage. La maturité est toutefois apparue insuffisante pour une industrialisation en l'état. Une partie des limites constatées est due à la qualité du jeu de données. Le plateau Simplification a pour ambition d'améliorer cette qualité de données, sur le fond par l'application de standards de rédaction et informatiquement, avec des données qui seront à terme en XML, modulaires et structurées, et donc plus propices au succès des méthodes d'intelligence artificielle. D'autres limites sont dues à des spécificités du corpus et permettent d'identifier des verrous scientifiques. Dans ce cadre, une collaboration scientifique entre le LIUM et la Direction Innovation et Recherche doit démarrer à l'automne 2020 pour mieux traiter les acronymes polysémiques, sources de bruit dans la qualité des résultats.

Pour les méthodes plus linguistiques (analyse syntaxique, ontologies), les tests utilisateurs ont démontré la capacité de ses approches à améliorer la pertinence des résultats, à mieux guider l'utilisateur et à offrir de nouvelles façons de naviguer dans les contenus qui lui permettent de mieux converger vers l'information qui le concerne.

Un passage à l'échelle a été enclenché, pour aller vers une industrialisation des sorties les plus matures et les plus concluantes d'un point de vue utilisateur. Cette phase, intégrée à la feuille de route du plateau Simplification, a démarré en mars 2019. Elle inclut la reprise de certains principes de recherche (auto-complétion de qualité, affichage du contexte d'occurrence par exemple), l'exploitation des ressources terminologiques (aide pour la construction d'un glossaire de l'ensemble des textes, enrichissement des données pour améliorer la recherche) et pour la réalisation de fonctionnalités dans le nouvel outil de consultation en cours de développement (recherche cartographique par exemple). Elle s'accompagne également d'une investigation plus poussée de la façon dont les utilisateurs conçoivent et expriment leurs besoins informationnels (acronymes, habitude de recherche par référence de document par exemple), afin d'approfondir des sujets révélés lors des tests utilisateurs et par analyse de corpus. L'industrialisation reste soumise à des contraintes : les compétences TAL restent encore rares en internes, notamment au niveau des DSI en charge de la reprise des travaux (mise en œuvre non triviale de certaines approches) ; un choix projet de développement spécifique a été fait au détriment de progiciels (par exemple, pas de reprise possible de la solution Synomia en l'état malgré ses performances) ; pour les approches ontologiques, un besoin d'expertise métier est indispensable en conception pour modéliser les connaissances, mais aussi pour maintenir durablement l'ontologie.

[1] Les mots ne pouvant être lemmatisés sont les mots pour lesquels il n'existe pas de lemme connu dans l'outil de lemmatisation, soit parce que ces termes sont très spécialisés, soit parce qu'ils correspondent à des entités nommées absentes du système.

[2] La liste de mots-outils classiques est issue de https://github.com/villmow/RAKE-tutorial/blob/master/FrenchStoplist.txt (adaptée de http://snowball.tartarus.org/algorithms/french/stop.txt)

## REFERENCES

[1] A. Condamines, "Peut-on prévenir le risque langagier dans la communication écrite ?", Langage et société, Vol. 125(3), 2008, pp. 77-97.

[2] F. Gandon, “Web de données liées et Web sémantique.”, 2017.

[3] J. Charlet, G. Kembellec, “Du web sémantique au web des données, quels enjeux professionnels?”, I2D Information, données documents, Vol. *53*(2), 2016, pp. 54-55.

[4] D. Bourigault, M. Slodzian, “Pour une terminologie textuelle”, Terminologies nouvelles, vol. 19, pp. 29-32, 1999.

[5] D. Bourigault, “Method and large syntactical analysis system of a corpus, a specialised corpus in particular”, U.S. Patent Application No 10/479,233, 2004.

[6] S. Robertson, H. Zaragoza, “The Probabilistic Relevance Framework: BM25 and Beyond”, Foundations and Trends in Information Retrieval., Vol. 3 Issue 4, 2009.

[7] K. Spärck Jones, *“*A statistical interpretation of term specificity and its application in retrieval*”*, Journal of Documentation, vol. 28-1, 1972, pp. 11–21.

[8] S. Rose, D. Engel, D., N. Cramer, N., W. Cowley, “Automatic keyword extraction from individual documents.”, Text mining: applications and theory, Vol. 1, 2010, pp. 1-20.

[9] N. Dugué., N. Camelin, L.Lefeuvre, X. Li, C. Reutenauer, et al.. “Apprentissage et évaluation de plongements lexicaux sur un corpus SNCF en langue spécialisée”, Proceedings of *Extraction et Gestion des Connaissances* Conference, Jan 2019, Metz, France, pp. 279-284.

[10] J.-C. Lamirel, N. Dugué, P. Cuxac, “New efficient clustering quality indexes”, International Joint Conference on Neural Networks (IJCNN), IEEE, 2016, pp. 3649-3657.

[11] J.C. Bastien, C. Leulier, D. L. Scapin, « L'ergonomie des sites web ». *Créer et maintenir un service Web*, 111-173J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon,1998, pp.68–73.

[12] A. Valentin, A. Lancry, C. Lemarchand, « La construction des échantillons dans la conception ergonomique de produits logiciels pour le grand public. Quel quantitatif pour les études qualitatives ? », *Le travail humain*, 2010/3 (Vol. 73), pp. 261-290

[13] J. Nielsen, R. Molich, Heuristic evaluation of user interfaces, *in* J. Carrasco & J. Whiteside (eds), *Proceedings of ACM Chi'90 Conference on Human Factors in Computing Systems*, 1-5 Avril 1990, Seatle, pp. 249-256.

[14] J. Nielsen, J. *Usability Engineering*, Boston, Academic Press, 1993.