



# Exalead et Synapse Développement

## Module : Reprise et indexation de textes dégradés

Exalead travaille depuis plusieurs années avec Synapse Développement, un leader du marché des ressources linguistiques dédiées aux moteurs de recherche, aussi bien sur les problématiques de correction de documents que d'intégration de dictionnaires de synonymes.

Aujourd'hui, en intégrant les outils de correction de Synapse Développement dans sa plate-forme Exalead CloudView™, Exalead offre à ses clients de nouveaux modules pour nettoyer des bases documentaires dégradées. Qu'il s'agisse de corriger des fautes de frappe ou encore d'homogénéiser des textes « appauvris » (par exemple, disponible en majuscule uniquement ou sans accent), ces modules automatisent efficacement la correction orthographique et grammaticale à très grande échelle.

### ► Reprise de bases documentaires

Possibilité de reprise des documents dégradés d'une base documentaire indexée par la plate-forme Exalead CloudView. La qualité du correcteur autorise une reprise automatisée et donc peu coûteuse des documents.

### ► Adaptation à tous les formats

Le correcteur permet d'obtenir un texte correct et présentable, quel que soit le format du texte à l'entrée : intégralement en majuscule, sans accent, issu d'OCR ou d'un traitement de type "reconnaissance vocale."

### ► Amélioration de l'expérience utilisateur

Au-delà de l'amélioration de la présentation des textes, la reprise de bases documentaires dégradées permet au moteur de recherche d'Exalead d'exprimer toute sa puissance pour offrir aux utilisateurs une expérience riche et satisfaisante de la découverte de contenus.

### ► Champs applicatifs

L'utilisation des modules de correction de Synapse est particulièrement adaptée dans le cadre des problématiques de :

- « Numérisation et Optical Character Recognition
- « Saisie manuelle de texte
- « Intégration de dictionnaires de synonymes dans la plate-forme EXALEAD

#### BÉNÉFICES MÉTIERS

- Reprise simplifiée de bases documentaires obsolètes ou atypiques
- Homogénéisation de contenu de type « user generated » pour réintégration dans des applications métiers de type CRM ou BI
- Indexation et accès optimisés sur des bases dégradées
- Limitation des coûts liés à la mise à niveau
- Automatisation des processus de nettoyage des bases pour une intégration quasi-instantanée des bases dégradées dans le SI de l'entreprise
- Amélioration de la qualité des documents et donc de l'expérience utilisateur

#### EN UN COUP D'ŒIL

#### Correcteur orthographique

- Identification des mots absents des dictionnaires et proposition de suggestions qualifiées
- Disponible en français, anglais, allemand, espagnol
- Très large couverture des dictionnaires :
  - Français : 212 639 mots, soit 1 112 149 formes
  - Anglais : 169 124 mots, soit 417 634 formes
  - Espagnol : 70 689 mots, soit 713 255 formes
  - Allemand : 55 600 mots, soit 872 857 formes
- Dictionnaires techniques spécialisés disponibles : juridique, médical, commercial, bâtiment
- Possibilité d'intégrer des dictionnaires métiers et utilisateurs propriétaires
- Paramétrage du niveau de correction
- Identification de la bonne suggestion sur 98% des erreurs courantes en français
- Possibilité de correction automatique de requêtes ou de textes complets
- Réaccentuation automatique de texte en minuscules ou majuscules

#### Correcteur grammatical

- Correction des fautes d'accords entre les mots
- Disponible en français, espagnol, portugais
- 14 000 règles de grammaire pour le français
- Paramétrage du niveau de correction

#### Mise en minuscules et réaccentuation

- Chaque mot est analysé et défini comme un nom commun ou nom propre
- Désambiguïsation pour les noms communs possédant la même forme aux signes diacritiques près (ELEVE -> élève, élevé)

Disponibles sous Linux et Windows

## ► Comment fonctionne le correcteur de Synapse?

La correcteur distingue la correction orthographique et grammaticale. La correction orthographique consiste à identifier les mots ne figurant pas dans les dictionnaires et à proposer, pour ces mots, un nombre restreint et qualifié de suggestions. La correction grammaticale vérifie ensuite les fautes d'accords entre les mots. La précision des suggestions du correcteur orthographique de Synapse Développement autorise une correction automatique d'une base documentaire dégradée, contenant entre autre des fautes de français ou bien des encodages inappropriés, qui sont souvent le fruit de contraintes techniques (textes en majuscules ou non accentués)

### Exemples de corrections : avant et après

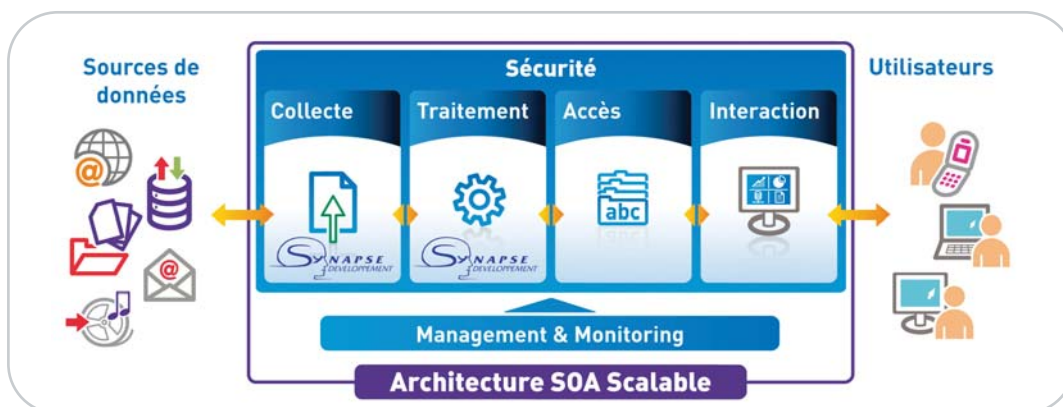
C'est génial la supression de la publicité sur les chaines de France <b>Y</b> élévision.	→	C'est génial la suppression de la publicité sur les chaînes de France <b>T</b> élévision.
Ces recettes pourraient <b>atteidnre</b> 5% du c.a. de France Télécom en quelques années.	→	Ces recettes pourraient <b>atteindre</b> 5% du c.a. de France Télécom en quelques années.
Vos remarques, il les a <b>suivi</b> à la lettre.	→	Vos remarques, il les a <b>suivies</b> à la lettre.
La fabrication d une piece ne peut etre <b>changee</b> sans une etude prealable.	→	La fabrication d une pièce ne peut être <b>changée</b> sans une étude préalable.
UNE ETUDE THEORIQUE A ETE ENGAGEE POUR DETERMINER LES CARACTERISTIQUES DE L'ECOULEMENT (CROISIERE) AU DROIT DU FUSELAGE.	→	Une étude théorique a été engagée pour déterminer les caractéristiques de l'écoulement (croisière) au droit du fuselage.

## ► Un module multilingue

Le module de correction de Synapse développement est disponible en plusieurs langues.

Module	Français	Anglais	Allemand	Espagnol
Correction orthographique	✓	✓	✓	✓
Correction grammaticale	✓			
Minuscule/Réaccentuation	✓			

## ► Totalement intégré à la plate-forme Exalead CloudView



## A propos de Synapse Développement

Pionniers dans le développement d'applications informatiques autour du texte, Synapse Développement propose un moteur de correction entièrement paramétrable et d'une qualité exceptionnelle. Synapse Développement est située à Toulouse et ses outils de correction ont été choisis par des les milliers d'utilisateurs et de grands industriels comme Microsoft Corporation.

