

Université Mohammed V- Rabat  
Ecole Mohammadia d'Ingénieurs  
Département Génie Informatique  
Filière Génie Informatique et Digitalisation



# Text Analytics

Dr. H. SEBBAQ

[h.sebbaq@gmail.com](mailto:h.sebbaq@gmail.com)

Pr. N. EL FADDOULI

[nfaddouli@gmail.com](mailto:nfaddouli@gmail.com)

2023-2024

CC-BY NC SA

# Approches, tâches et applications de l'analyse de texte : Approches

---

## ❑ Approche manuelle

- Les analystes collectent, traitent et analysent manuellement les données sans l'aide d'outils ou d'algorithmes automatisés.
- Implique une intervention humaine et une prise de décision tout au long du processus d'analyse.
- La prise de décision repose souvent sur l'expertise, l'expérience et le jugement subjectif de l'homme.
- **Exemple concret** : Un **analyste financier** passe à travers les états financiers d'une entreprise, **examinant individuellement chaque ligne des rapports comptables** pour repérer des incohérences, des tendances ou des anomalies. Par exemple, identifier des chiffres inhabituels dans les colonnes de revenus ou de dépenses.

# Approches, tâches et applications de l'analyse de texte : Approches

---

## ❑ Approche basée sur des règles

- Implique des ensembles prédéfinis de règles ou de conditions pour analyser les données.
- Ces règles sont établies sur la base de connaissances d'experts, de normes industrielles ou d'exigences réglementaires.
- L'analyse et la prise de décision suivent des règles prédéterminées sans qu'une intervention humaine constante soit nécessaire.
- **Exemple concret** : Un système de détection de fraude bancaire utilise des **règles préétablies** pour repérer des **transactions inhabituelles**. Par exemple, il peut bloquer automatiquement les transactions d'un montant excessivement élevé réalisées à partir de lieux géographiquement éloignés dans un court laps de temps.

# Approches, tâches et applications de l'analyse de texte : Approches

---

## ❑ Approche automatisée

- Utilise des algorithmes, l'apprentissage automatique et l'intelligence artificielle pour effectuer l'analyse des données.
- Les algorithmes sont conçus pour traiter et analyser automatiquement les données et en tirer des enseignements sans intervention humaine directe.
- **Exemple concret** : Les plateformes de streaming vidéo telles que **Netflix** utilisent des algorithmes d'apprentissage automatique pour recommander des films ou des séries en fonction des habitudes de visionnage. Par exemple, en suggérant des films similaires à ceux que l'utilisateur a déjà regardé ou en se basant sur les évaluations qu'il a donné à d'autres contenus.

# Approches, tâches et applications de l'analyse de texte : Approches

## Scénario :

- Lancement d'une Nouvelle Saison très attendue sur Netflix



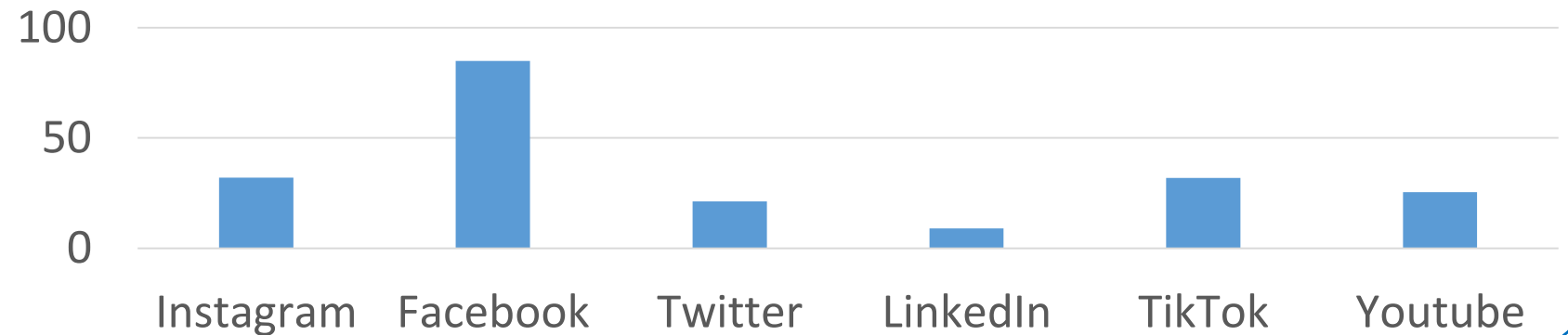
Netflix  @netflix · 17h

Ready for a first look at your next obsession

From the creators of DARK comes 1899, a mystery-horror series about a group of European migrants whose journey to New York turns into a nightmare when they come across another ship adrift on the ocean

[#GeekedWeek](#) 

## Abonnés en millions





# Approches, tâches et applications de l'analyse de texte : Approches

---

## ❑ Approche manuelle

### ➤ Points forts

- **Compréhension du contexte** : Les analystes humains fournissent des informations nuancées en comprenant la profondeur émotionnelle et les nuances subtiles des commentaires et des critiques des spectateurs.

### ➤ Points faibles

- **Time-Consuming** : L'examen manuel d'un grand nombre de commentaires et d'évaluations peut prendre beaucoup de temps, ce qui peut retarder les temps de réponse aux problèmes émergents.
- **Extensibilité limitée** : L'approche pourrait avoir du mal à traiter efficacement une grande quantité de données, ce qui nuirait à sa capacité à recueillir tous les commentaires des utilisateurs.
- **Exemple concret** : Les analystes explorent manuellement les commentaires des médias sociaux, les forums et les blogs pour découvrir les réactions détaillées du public. Ils peuvent remarquer une critique récurrente sur le rythme lent de la série, mais ils peuvent avoir du mal à couvrir tous les commentaires des utilisateurs en raison de l'immense volume de réactions.

# Approches, tâches et applications de l'analyse de texte : Approches

---

## ❑ Approche basée sur des règles

### ➤ Points forts

- **Mesures automatisées** : Les systèmes automatisés peuvent évaluer rapidement des paramètres prédéfinis tels que les vues, les évaluations ou les partages dans les médias sociaux afin de déclencher des actions spécifiques.
- **Cohérence** : Les systèmes basés sur des règles fournissent des réponses et des actions cohérentes basées sur des critères prédéterminés.

### ➤ Faiblesses

- **Manque de flexibilité** : Les règles rigides peuvent négliger les sentiments nuancés ou changeants des utilisateurs qui ne correspondent pas aux critères prédéfinis.
- **Compréhension limitée du contexte** : Les règles peuvent ne pas saisir la profondeur émotionnelle ou les circonstances uniques entourant les commentaires des utilisateurs.
- **Exemple concret** : Netflix établit une règle déclenchant des activités promotionnelles supplémentaires si la série dépasse un certain nombre de vues dans les premières 24 heures → risque de ne pas tenir compte de la réaction émotionnelle positive massive des téléspectateurs qui ont suivi la série après la fenêtre initiale.

# Approches, tâches et applications de l'analyse de texte : Les approches

---

## ❑ Approche Automatisée

### ➤ Points forts

➤ **Analyse avancée** : Exploitation de l'apprentissage automatique pour traiter et analyser de grandes quantités de commentaires d'utilisateurs, en identifiant des modèles de sentiments et en prédisant des tendances.

### ➤ Points faibles

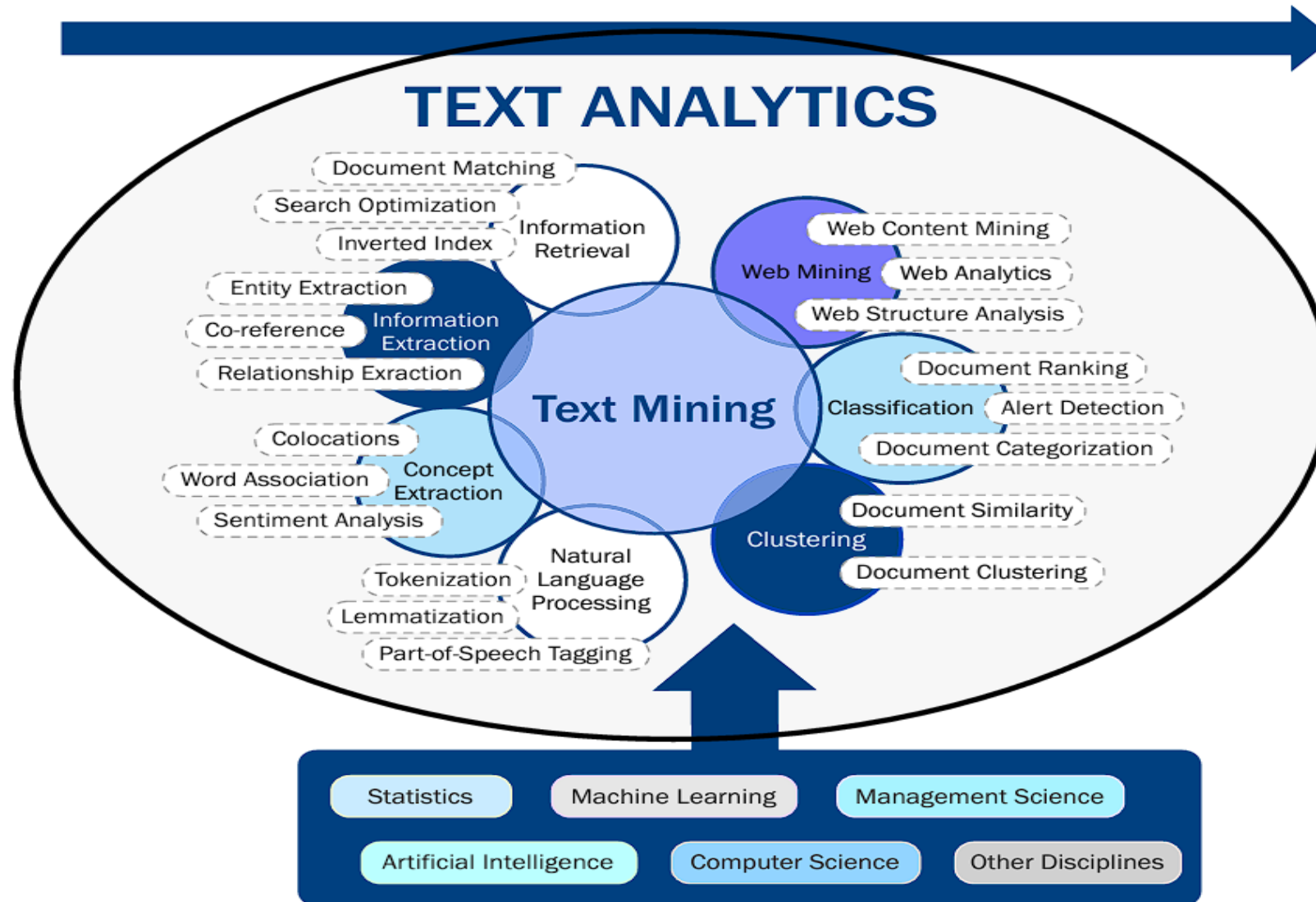
➤ **Dépendance à l'égard des données** : La précision dépend de la qualité et de la diversité des données de formation.

➤ **Interprétabilité** : Les modèles d'IA complexes peuvent manquer de transparence, ce qui rend difficile la compréhension du raisonnement qui sous-tend certaines prédictions.

➤ **Exemple concret** : Un modèle d'IA formé à partir d'un grand nombre de commentaires de téléspectateurs est utilisé pour prédire le succès potentiel d'une série sur la base des émotions exprimées. Cependant, il peut présenter des inexactitudes si les données d'entraînement sont biaisées ou limitées.



# Approches, tâches et applications de l'analyse de texte : Tâches



# Approches, tâches et applications de l'analyse de texte : Tâches

## ❑ Recherche et récupération d'informations (Information Retrieval)

- La recherche et l'extraction d'informations couvrent **l'indexation, la recherche et l'extraction de documents** à partir de grandes bases de données textuelles à l'aide de **requêtes par mots-clés**.

## ❑ Extraction de concepts

- **Définition** : L'extraction de concepts consiste à identifier des entités spécifiques telles que des noms de personnes, d'organisations ou de lieux à partir d'un texte non structuré.
- **Exemple** : Extraction d'entités comme **"Elon Musk" (personne)** et **"Tesla" (organisation)** à partir d'articles de presse traitant d'innovations technologiques.

## ❑ Extraction d'informations

- **Définition** : L'extraction d'informations vise à récupérer des données structurées, telles que des événements ou des relations, à partir de textes non structurés.
- **Exemple** : Analyse des critiques de produits pour en **extraire les caractéristiques** ("autonomie de la batterie", "qualité de l'appareil photo") et les sentiments correspondants (**positifs/négatifs**) afin de fournir des informations permettant d'améliorer le produit.

# Approches, tâches et applications de l'analyse de texte : Tâches

## ❑ Classification

- **Définition** : La classification des textes est une tâche qui consiste à **classer les données textuelles dans des classes ou catégories prédéfinies**.
- **Objectif** : Attribuer des étiquettes au texte en fonction de son contenu, ce qui permet de trier ou d'organiser automatiquement le texte en groupes spécifiques.
- **Exemple** : Analyse de sentiment des clients de l'opérateur telecom

## ❑ Clustering

- **Définition** : Le regroupement est une technique utilisée dans l'analyse des données et l'apprentissage automatique pour **regrouper des objets ou des points de données similaires sur la base de leurs caractéristiques inhérentes**.
- **Objectif** : Elle vise à diviser un ensemble de données en sous-ensembles, ou clusters, où les points de données d'un même cluster sont plus similaires les uns aux autres que ceux de clusters différents.
- **Exemple** : Le clustering peut être appliqué pour regrouper des articles d'actualité similaires. Par exemple, les articles traitant des tendances technologiques, de l'actualité politique et des mises à jour en matière de santé sont regroupés dans des groupes distincts.

# Approches, tâches et applications de l'analyse de texte : Tâches

---

## ❑ Exploration du web (Web mining)

- ❑ **Définition** : L'exploration du Web permet **d'extraire des informations ou des modèles** utiles des données du Web, afin de **mieux comprendre le comportement des utilisateurs** ou d'analyser le contenu.
- ❑ **Exemple** : Analyse des modèles de recherche des utilisateurs sur un site web d'achat afin d'identifier les catégories de produits les plus populaires et d'optimiser les stratégies de marketing.

## ❑ Traitement du langage naturel (NLP)

- ❑ **Définition** : Le NLP implique l'interaction entre les ordinateurs et le langage humain, permettant aux machines de comprendre, d'interpréter et de générer du langage humain.
- ❑ **Exemple** : Les chatbots ou les assistants virtuels tels que Siri, Bixby ou Google Assistant utilisent le NLP pour comprendre et répondre aux requêtes vocales ou écrites des humains.

# Applications : Marketing

---

- **Analyse des concurrents** : Suivre la situation des concurrents pour comprendre leurs forces, leurs faiblesses et les opportunités du marché.
- **Gestion des campagnes** : Évaluer les réponses du public aux publicités afin d'affiner le ciblage et les budgets.
- **Prévision des tendances du marché** : Utilisation de données textuelles pour prédire les tendances futures du marché et adapter les stratégies.
- **Analyse des sentiments et retour d'information** : Comprendre les opinions des clients à partir des commentaires et des médias sociaux afin d'affiner les stratégies.
- **Segmentation et profilage des clients** : Regroupement des clients pour des messages et des offres personnalisés.

## Applications : Finance

---

- **Prise de décision en matière d'investissement** : Utilisation de l'analyse de texte pour éclairer les décisions d'investissement sur la base d'actualités et d'analyses d'experts.
- **Évaluation des risques** : Identification des risques potentiels dans les investissements par l'analyse des rapports financiers.
- **Gestion de portefeuille** : Optimisation des portefeuilles d'investissement et gestion des risques grâce à l'analyse des données.
- **Prévisions économiques** : Prévision des tendances économiques et des facteurs ayant un impact sur les marchés financiers par l'analyse de données textuelles.
- **Détection des fraudes** : Détecter les anomalies dans les transactions et les communications afin de prévenir la fraude.
- **Trading algorithmique** : Développement d'algorithmes pour des décisions commerciales automatisées à l'aide du traitement du langage naturel.



# Applications : Education

---

- **Formation des enseignants** : Améliorer le développement professionnel des éducateurs.
- **Amélioration de l'apprentissage en ligne** : Amélioration de l'expérience des utilisateurs sur les plateformes d'apprentissage.
- **Personnalisation du contenu** : Adaptation du contenu éducatif en fonction des préférences des étudiants.
- **Analyse du retour d'information** : Comprendre les commentaires des étudiants pour améliorer les méthodes d'enseignement.
- **Identification des étudiants à risque** : Détecter les étudiants ayant besoin d'un soutien supplémentaire.
- **Recherche pédagogique** : Examen d'articles et de recherches universitaires pour en tirer des enseignements.
- **Détection du plagiat** : Prévention du plagiat dans les travaux des étudiants.

# Applications : Santé

---

- **Surveillance des épidémies** : Prévoir les épidémies en analysant les données médicales et les médias sociaux.
- **Suivi des réactions aux vaccins** : Analyse des données pour surveiller les réactions et les opinions du public concernant les vaccins.
- **Diagnostic assisté par ordinateur** : Fournir des données supplémentaires pour le diagnostic médical.
- **Amélioration des diagnostics** : Extraction d'informations pour des diagnostics et des traitements précis.
- **Amélioration des soins** : Utilisation du retour d'information des patients pour améliorer la qualité des soins.
- **Recherche de nouveaux traitements** : Exploration d'articles et de données cliniques pour découvrir de nouvelles approches médicales.
- **Surveillance de la santé mentale** : Détection des tendances et des problèmes de santé mentale sur les plateformes en ligne.