Fouille de textes

Thierry Hamon

Bureau H202 - Institut Galilée
Tél. : 33 1.49.40.35.53
LIMSI-CNRS
hamon@limsi.fr
https://perso.limsi.fr/hamon/Teaching/P13/FDT-2016-2017/

Ma recherche

- Domaine : Traitement Automatique des Langues
  - appliqué aux domaines de spécialité (électricité, médecine, risque chimique, etc.)
  - sur des textes en français, anglais, ukrainien, arabe
- Proposition d'approches pour
  - Constitution et enrichissement de ressources terminologiques
  - Articulation d'outils de TAL
  - Extraction et recherche d'information
- Type de données textuelles : bases bibliographiques, documents techniques, dossiers patients, articles scientifiques, forums de discussion, etc.
- Développement d'outils en Perl (modules CPAN)

Ma recherche

- Constitution et enrichissement de ressources terminologiques
  - Exploration de corpus textuels
  - Extraction de termes : YATEA

    [Aubin et Hamon 2006, Hamon et al. 2014, Neifar et al. 2016, Hamon et Grabar 2016]
  - Acquisition de relations sémantiques : SynoTerm

    [Hamon 2000, Grabar et al. 2008, Hamon et Grabar 2009]
  - Application : Identification automatique de mots-clés d'articles

    [Hamon 2012, Hamon 2016]
- Articulation d'outils de TAL [Hamon et al. 2007]

  Définition d'une plate-forme modulaire combinant des outils de TAL pour annoter des documents avec des informations linguistiques

Ma recherche

- Extraction et recherche d'information
  - Fouille de dossiers patients (en anglais) :
    Identification des maladies, des médicaments et des traitements mais aussi des relations entre eux et de leur niveau de certitude
    [Grabar et Hamon 2009, Hamon et al. 2010, Périnet et al. 2011]
  - Fouille de données bibliographiques :
    Collecte des facteurs de risque associés aux maladies
    [Hamon et al. 2010]
  - Collecte ciblée de documents sur le Web [Gollub et al. 2007]
  - Intégration d'informations linguistiques dans le moteur de recherche de l'intranet [Grabar et al. 2009]
  - Fouille de forums de discussion [Hamon et Gagnayre 2013]
    - Identifier des compétences requises par des programmes d'éducation thérapeutique
    - Profilage des internautes (identification des informations démographiques – en cours)

1 Organisation

2 Introduction à la Fouille de Textes
- Textes et Applications
- Objectif
- Fouille de données vs. Fouille de textes
- Caractéristiques de la fouille de textes

3 Difficultés de la fouille de texte
- Exemples de données textuelles
- Recherche vs. extraction d'information

4 Données textuelles
- Type de données textuelles
- Corpus
- Exemple de corpus spécialisés
- Problèmes associés et solutions

Organisation

- 8 séances (12h) de cours
  - principe de la fouille de données textuelles
  - Méthodologies d'extraction d'information à partir de textes
  - Outils pour la mise en place de systèmes de fouille de textes
- 8 séances (12h) de TP

Introduction à la Fouille de Textes

Constat : prépondérance des données non-structurées

- plus 90% des données produites en 2011 (sur 1.8 trillion de gigaoctets) [Oracle, 2011]
- 85% des données d'entreprise
- Données non-structurées : des textes
  - Importante source d'informations
  - Mais la lecture est coûteuse en temps, voir impossible
  - Besoin d'aller au-delà de la recherche de documents

Introduction à la Fouille de Textes

Constat : prépondérance des données non-structurées

- plus 90% des données produites en 2011 (sur 1.8 trillion de gigaoctets) [Oracle, 2011]
- 85% des données d'entreprise
- Données non-structurées : des textes
  - Importante source d'informations
  - Mais la lecture est coûteuse en temps, voir impossible
  - Besoin d'aller au-delà de la recherche de documents

Accèder à ces informations nécessite de fouiller les textes automatiquement

Introduction à la Fouille de Textes

Des textes de différentes natures :

- Pages Web, blogs, forums, emails
- Documents techniques, rapports produits dans les entreprises

- Livres, Bibliothèques numériques, articles scientifiques
- Enquête de consommateurs, Lettres de plaintes de consommateurs
- Dossiers patients (examens médicaux, lettres de médecin, compte-rendus, etc.)
- Rapports d'interaction entre médicaments, de décès
- Rapports d'incidents, constats d'accident
- Dépêches, articles de journaux

- ...

Introduction à la Fouille de Textes

Des textes de différentes natures :

- Pages Web, blogs, forums, emails
- Documents techniques, rapports produits dans les entreprises

- Livres, Bibliothèques numériques, articles scientifiques
- Enquête de consommateurs, Lettres de plaintes de consommateurs
- Dossiers patients (examens médicaux, lettres de médecin, compte-rendus, etc.)
- Rapports d'interaction entre médicaments, de décès
- Rapports d'incidents, constats d'accident
- Dépêches, articles de journaux

- ...

Tous contiennent des informations utiles dans de nombreux cadres applicatifs

Introduction à la Fouille de Textes

Des textes de différentes natures :

- Analyse d'opinion, de comportement, filtrage des spams
- Documents techniques, rapports produits dans les entreprises

- Livres, Bibliothèques numériques, articles scientifiques
- Enquête de consommateurs, Lettres de plaintes de consommateurs
- Dossiers patients (examens médicaux, lettres de médecin, compte-rendus, etc.)
- Rapports d'interaction entre médicaments, de décès
- Rapports d'incidents, constats d'accident
- Dépêches, articles de journaux

- ...

Tous contiennent des informations utiles dans de nombreux cadres applicatifs

Introduction à la Fouille de Textes

Des textes de différentes natures :

- Analyse d'opinion, de comportement, filtrage des spams
- Analyse d'incidents, Constitution d'une mémoire d'entreprise, Aide à la rédaction, Analyse du besoin, Formalisation de spécifications
- Livres, Bibliothèques numériques, articles scientifiques
- Enquête de consommateurs, Lettres de plaintes de consommateurs
- Dossiers patients (examens médicaux, lettres de médecin, compte-rendus, etc.)
- Rapports d'interaction entre médicaments, de décès
- Rapports d'incidents, constats d'accident
- Dépêches, articles de journaux

- ...

Tous contiennent des informations utiles dans de nombreux cadres applicatifs

Introduction à la Fouille de Textes

Des textes de différentes natures :

- Analyse d'opinion, de comportement, filtrage des spams
- Analyse d'incidents, Constitution d'une mémoire d'entreprise, Aide à la rédaction, Analyse du besoin, Formalisation de spécifications
- Agrégation de données, Structuration de connaissances, Aide à la décision
- Enquête de consommateurs, Lettres de plaintes de consommateurs
- Dossiers patients (examens médicaux, lettres de médecin, compte-rendus, etc.)
- Rapports d'interaction entre médicaments, de décès
- Rapports d'incidents, constats d'accident
- Dépêches, articles de journaux

- ...

Tous contiennent des informations utiles dans de nombreux cadres applicatifs

Introduction à la Fouille de Textes

Des textes de différentes natures :

- Analyse d'opinion, de comportement, filtrage des spams
- Analyse d'incidents, Constitution d'une mémoire d'entreprise, Aide à la rédaction, Analyse du besoin, Formalisation de spécifications
- Agrégation de données, Structuration de connaissances, Aide à la décision
- Etude des tendances, analyse d'opinion, rapports d'incidents
- Dossiers patients (examens médicaux, lettres de médecin, compte-rendus, etc.)
- Rapports d'interaction entre médicaments, de décès
- Rapports d'incidents, constats d'accident
- Dépêches, articles de journaux

- ...

Tous contiennent des informations utiles dans de nombreux cadres applicatifs

Introduction à la Fouille de Textes

Des textes de différentes natures :

- Analyse d'opinion, de comportement, filtrage des spams
- Analyse d'incidents, Constitution d'une mémoire d'entreprise, Aide à la rédaction, Analyse du besoin, Formalisation de spécifications
- Agrégation de données, Structuration de connaissances, Aide à la décision
- Etude des tendances, analyse d'opinion, rapports d'incidents
- Aide à la décision médicale, Aide à la lecture, Codage et facturation des interventions
- Rapports d'interaction entre médicaments, de décès
- Rapports d'incidents, constats d'accident
- Dépêches, articles de journaux

- ...

Tous contiennent des informations utiles dans de nombreux cadres applicatifs

Introduction à la Fouille de Textes

Des textes de différentes natures :

- Analyse d'opinion, de comportement, filtrage des spams
- Analyse d'incidents, Constitution d'une mémoire d'entreprise, Aide à la rédaction, Analyse du besoin, Formalisation de spécifications
- Agrégation de données, Structuration de connaissances, Aide à la décision
- Etude des tendances, analyse d'opinion, rapports d'incidents
- Aide à la décision médicale, Aide à la lecture, Codage et facturation des interventions
- Etude épidémiologique, rédaction de RCP, Pharmacovigilance
- Rapports d'incidents, constats d'accident
- Dépêches, articles de journaux
- ...

Tous contiennent des informations utiles dans de nombreux cadres applicatifs

Introduction à la Fouille de Textes

Des textes de différentes natures :

- Analyse d'opinion, de comportement, filtrage des spams
- Analyse d'incidents, Constitution d'une mémoire d'entreprise, Aide à la rédaction, Analyse du besoin, Formalisation de spécifications
- Agrégation de données, Structuration de connaissances, Aide à la décision
- Etude des tendances, analyse d'opinion, rapports d'incidents
- Aide à la décision médicale, Aide à la lecture, Codage et facturation des interventions
- Etude épidémiologique, rédaction de RCP, Pharmacovigilance
- Analyse du risque, Estimation de coûts, aide à la décision
- Dépêches, articles de journaux

- ...

Tous contiennent des informations utiles dans de nombreux cadres applicatifs

Introduction à la Fouille de Textes

Des textes de différentes natures :

- Analyse d'opinion, de comportement, filtrage des spams
- Analyse d'incidents, Constitution d'une mémoire d'entreprise, Aide à la rédaction, Analyse du besoin, Formalisation de spécifications
- Agrégation de données, Structuration de connaissances, Aide à la décision
- Etude des tendances, analyse d'opinion, rapports d'incidents
- Aide à la décision médicale, Aide à la lecture, Codage et facturation des interventions
- Etude épidémiologique, rédaction de RCP, Pharmacovigilance
- Analyse du risque, Estimation de coûts, aide à la décision
- Analyse d'opinion, Aide à la lecture ou à la décision, prédiction de mouvements financiers
- ...

Tous contiennent des informations utiles dans de nombreux cadres applicatifs

Introduction à la Fouille de Textes

Des textes de différentes natures :

- Analyse d'opinion, de comportement, filtrage des spams
- Analyse d'incidents, Constitution d'une mémoire d'entreprise, Aide à la rédaction, Analyse du besoin, Formalisation de spécifications
- Agrégation de données, Structuration de connaissances, Aide à la décision
- Etude des tendances, analyse d'opinion, rapports d'incidents
- Aide à la décision médicale, Aide à la lecture, Codage et facturation des interventions
- Etude épidémiologique, rédaction de RCP, Pharmacovigilance
- Analyse du risque, Estimation de coûts, aide à la décision
- Analyse d'opinion, Aide à la lecture ou à la décision, prédiction de mouvements financiers
- Mais aussi : détection de plagiat/attribution d'auteurs, analyse/comparaison de brevets
- ...

Tous contiennent des informations utiles dans de nombreux cadres applicatifs

Introduction à la Fouille de données textuelles

Objectif :
Extraire des textes, les connaissances non-triviales et intéressantes
C'est-à-dire :

- Extraire les informations

Introduction à la Fouille de données textuelles

Objectif :
Extraire des textes, les connaissances non-triviales et intéressantes
C'est-à-dire :

- Extraire les informations
  - utiles pour une tâche donnée

Introduction à la Fouille de données textuelles

Objectif :
Extraire des textes, les connaissances non-triviales et intéressantes
C'est-à-dire :

- Extraire les informations
  - utiles pour une tâche donnée
  - à partir de données textuelles, rédigées en langage naturel

Introduction à la Fouille de données textuelles

Objectif :
Extraire des textes, les connaissances non-triviales et intéressantes
C'est-à-dire :

- Extraire les informations
  - utiles pour une tâche donnée
  - à partir de données textuelles, rédigées en langage naturel
- Regrouper et mettre en relation ces informations

Introduction à la Fouille de données textuelles

Objectif :
Extraire des textes, les connaissances non-triviales et intéressantes
C'est-à-dire :

- Extraire les informations
  - utiles pour une tâche donnée
  - à partir de données textuelles, rédigées en langage naturel
- Regrouper et mettre en relation ces informations

Transformer du texte en connaissances utiles pour une analyse donnée
Alimentation de bases de données ou des bases de connaissances

Introduction à la Fouille de données textuelles

Objectif :
Extraire des textes, les connaissances non-triviales et intéressantes
C'est-à-dire :

- Extraire les informations
  - utiles pour une tâche donnée
  - à partir de données textuelles, rédigées en langage naturel
- Regrouper et mettre en relation ces informations

Transformer du texte en connaissances utiles pour une analyse donnée
Alimentation de bases de données ou des bases de connaissances

Mais aussi, permettre indirectement la fouille de données sur des données non-structurées

Fouille de données vs. Fouille de textes

Data mining vs. Text mining

- Fouille de données (*data mining*)
  - Méthodes et algorithmes d'exploration des données structurées, issues de bases de données, d'entrepôts de données
  - Objectifs : Mise en évidence de règles, identification de tendances, comportement invisible pour un humain

Fouille de données vs. Fouille de textes

Data mining vs. Text mining

- Fouille de données (*data mining*)
  - Méthodes et algorithmes d'exploration des données structurées, issues de bases de données, d'entrepôts de données
  - Objectifs : Mise en évidence de règles, identification de tendances, comportement invisible pour un humain
- Fouille de textes (*text mining*)
  - Méthodes d'exploitation de données non-structurées, présentes dans des textes rédigés en langage naturel
  - Objectifs : Extraction et catégorisation d'informations présentes dans les textes

Caractéristiques de la fouille de textes

- Matériau initial : Corpus/collection de documents textuels
- Connaissances sur la langue : Ressources linguistiques et terminologiques
- Accès aux informations linguistiques : Méthodes d'analyse linguistique (TAL) plus ou moins sophistiquées
- Regroupement des informations : Classification ou catégorisation automatique des informations
- Quantification des informations identifiées dans les documents : Mesures de similarité
- Prise en compte de la sémantique dans l'accès à l'information

Domaines associés

La fouille de textes emprunte des méthodes ou s'appuie sur plusieurs domaines :

- Recherche d'information
- Analyse lexicographique
- Apprentissage automatique (supervisé ou non)
- Traitement Automatique des Langues (TAL)
- Statistiques
- Visualisation de données, de l'information
- ...

Difficultés de la fouille de texte

Collection de données *en texte libre*

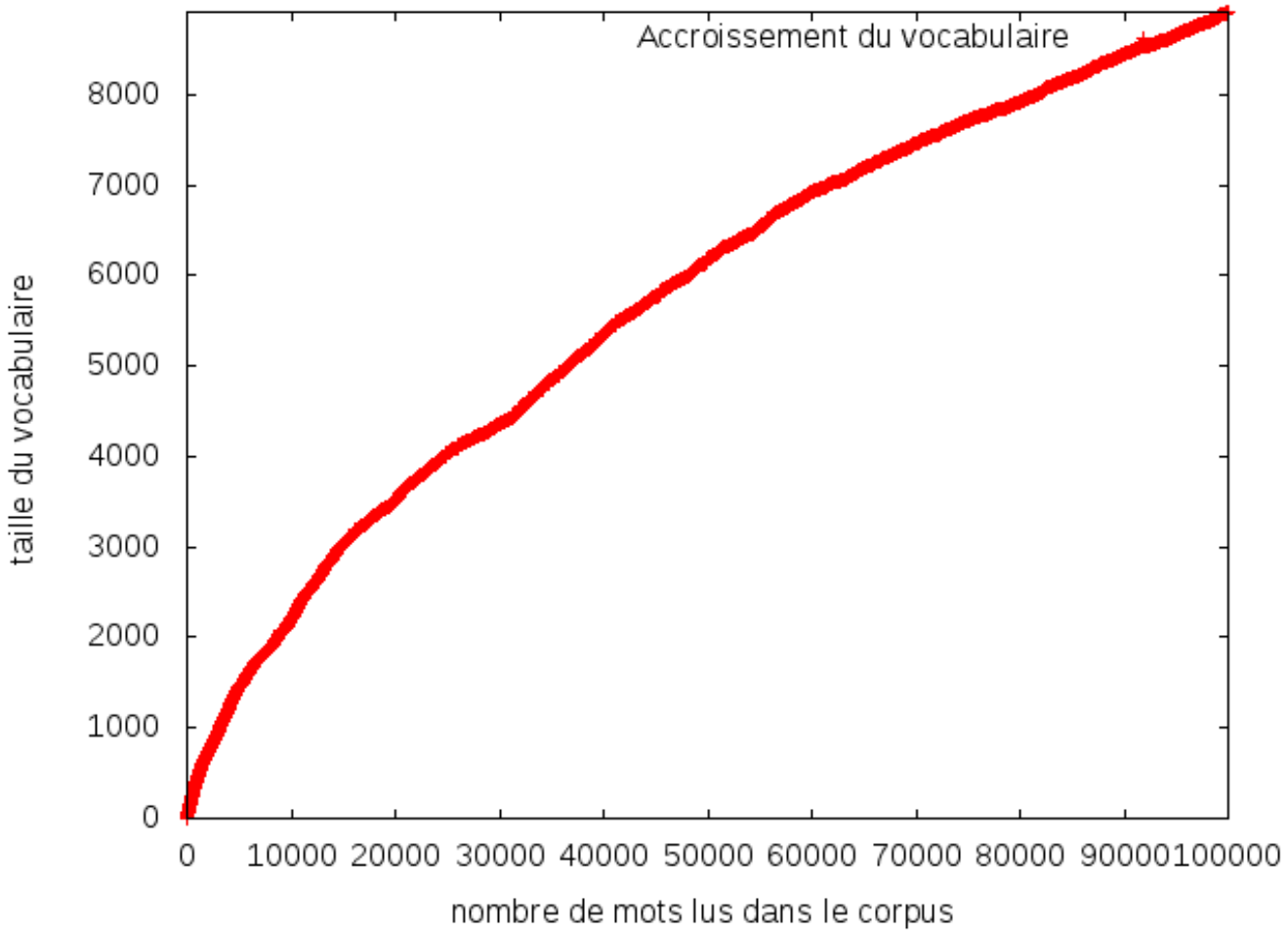
- Données éparpillées, mal organisées, bruitées
  - structure différente d'un texte à l'autre
  - pas statistiquement indépendant
- Données multilingues
- Données non-structurées, au mieux semi-structurées
- Volume de données textuelles important
- Temps de traitement de données textuelles élevés

Difficultés de la fouille de texte

- Importance du contexte dans l'accès à l'information
- Ambiguïtés de la langue à tous les niveaux
- Variation dans les énoncés similaires sémantiquement
- Nombre élévé et imprévisible de dimensions possibles :
  - Tous les mots et expressions de la langue peuvent être présents dans les textes (cf. accroissement du vocabulaire et Loi de Zipf)
  - On peut aussi avec de nouveaux mots
    - acronymes, néologismes
    - noms de marque, de personnes, de lieux, des dates, etc.)
    - mots empruntés à d'autres langues

Courbe d'accroissement du vocabulaire

Corpus Menelas

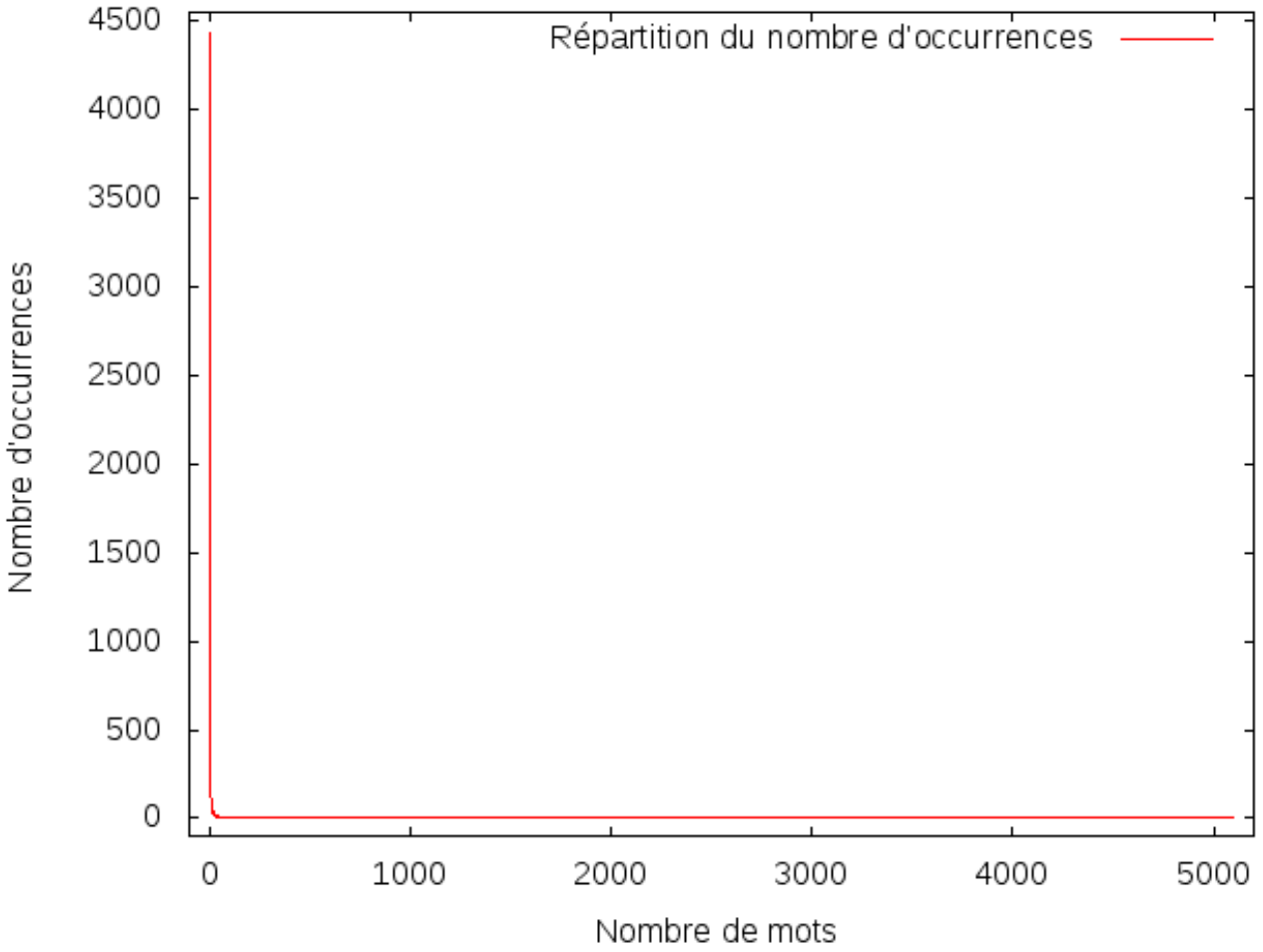


Loi de Zipf

Soient les mots d'un corpus donné, le rang d'un mot correspond à son nombre d'occurrence (dans l'ordre décroissant).
La loi de Zipf prévoit que le nombre d'occurrence d'un mot est inversement proportionnel à son rang.

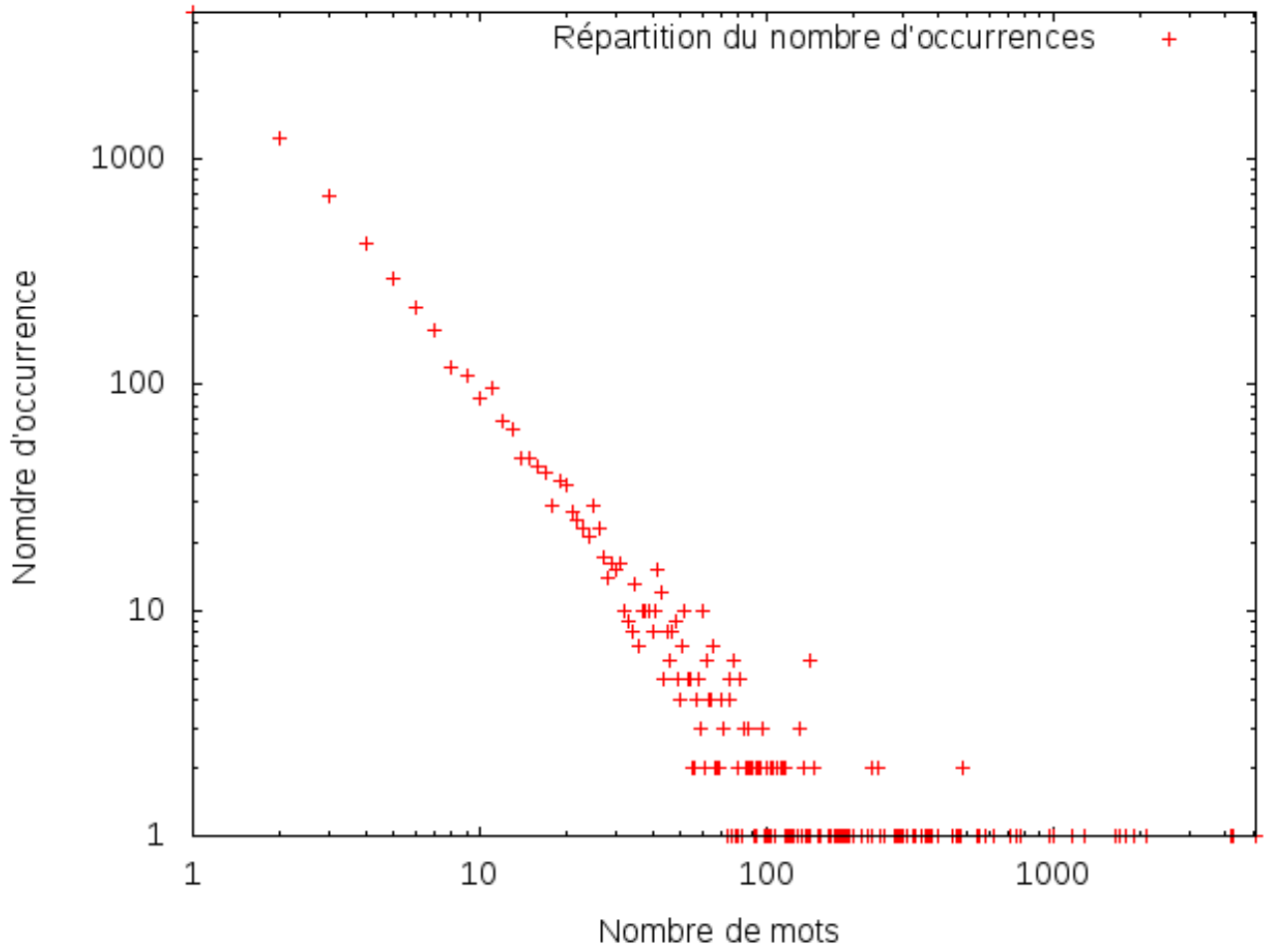
Répartition du nombre d'occurrences

Corpus Menelas



Répartition du nombre d'occurrences

Corpus Menelas



Difficultés de la fouille de texte

Ambiguïtés de la langue à tous les niveaux :

- lexical :
  - *avocat*[fruit] vs. *avocat*[métier]
  - *porte*[N] vs. *porte*[V]
  - *Apple*[société] vs. *apple*[fruit]
- syntaxique :
  - *la route est déviée par la gendarmerie.*
  - *Sylvain a vu un homme avec un télescope.*
- sémantique :
  - *une personne curieuse*
  - *L'homme descend du singe. Le singe descend de l'arbre*
- pragmatique :
  - *Le professeur a envoyé l'élève chez le proviseur parce qu'il le trouvait insupportable.*
  - *ce charmant patient*

Difficultés de la fouille de texte

- Variation dans les énoncés similaires sémantiquement :
  - *Bayer rachète Monsanto*
  - *Monsanto est racheté par Bayer*
  - *Bayer et Monsanto fusionnent*
  - *Monsanto fusionne avec Bayer*
  - *fusion Monsanto-Bayer*
- Métonymie (*le dernier Apple/Samsung*)
- Métaphore (*Géant du Web*, *or noir*)
- Erreurs d'orthographe : *Appel/Apple*
- ...

Exemples de données textuelles

Résumé d'article scientifique

```
Comparative Analyses of Hairpin Substrate Recognition by
Escherichia coli and Bacillus subtilis Ribonuclease P Ribozymes.

Ando T, Tanaka T, Kikuchi Y.

Division of Bioscience and Biotechnology, Department of Ecological
Engineering, Toyohashi University of Technology.

Previously, we reported that the substrate shape recognition of
the Escherichia coli ribonuclease (RNase) P ribozyme depends on
the concentration of magnesium ion in vitro. We additionally
examined the Bacillus subtilis RNase P ribozyme and found that the
B. subtilis enzyme also required high magnesium ion, above 10 mM,
for cleavage of a hairpin substrate. The results of kinetic
studies showed that the metal ion concentration affected both the
catalysis and the affinity of the ribozymes toward a hairpin RNA
substrate.
```

Exemples de données textuelles

Résumé d'article scientifique

```
Comparative Analyses of Hairpin Substrate Recognition by
Escherichia coli and Bacillus subtilis Ribonuclease P Ribozymes.

Ando T, Tanaka T, Kikuchi Y.

Division of Bioscience and Biotechnology, Department of Ecological
Engineering, Toyohashi University of Technology.

Previously, we reported that the substrate shape recognition of
the Escherichia coli ribonuclease (RNase) P ribozyme depends on
the concentration of magnesium ion in vitro. We additionally
examined the Bacillus subtilis RNase P ribozyme and found that the
B. subtilis enzyme also required high magnesium ion, above 10 mM,
for cleavage of a hairpin substrate. The results of kinetic
studies showed that the metal ion concentration affected both the
catalysis and the affinity of the ribozymes toward a hairpin RNA
substrate.
```

Commentaires :

Exemples de données textuelles

Résumé d'article scientifique

```
Comparative Analyses of Hairpin Substrate Recognition by
Escherichia coli and Bacillus subtilis Ribonuclease P Ribozymes.

Ando T, Tanaka T, Kikuchi Y.

Division of Bioscience and Biotechnology, Department of Ecological
Engineering, Toyohashi University of Technology.

Previously, we reported that the substrate shape recognition of
the Escherichia coli ribonuclease (RNase) P ribozyme depends on
the concentration of magnesium ion in vitro. We additionally
examined the Bacillus subtilis RNase P ribozyme and found that the
B. subtilis enzyme also required high magnesium ion, above 10 mM,
for cleavage of a hairpin substrate. The results of kinetic
studies showed that the metal ion concentration affected both the
catalysis and the affinity of the ribozymes toward a hairpin RNA
substrate.
```

Commentaires : Syntaxe correcte (ouf !)

Exemples de données textuelles

Résumé d'article scientifique

```
Comparative Analyses of Hairpin Substrate Recognition by
Escherichia coli and Bacillus subtilis Ribonuclease P Ribozymes.

Ando T, Tanaka T, Kikuchi Y.

Division of Bioscience and Biotechnology, Department of Ecological
Engineering, Toyohashi University of Technology.

Previously, we reported that the substrate shape recognition of
the Escherichia coli ribonuclease (RNase) P ribozyme depends on
the concentration of magnesium ion in vitro. We additionally
examined the Bacillus subtilis RNase P ribozyme and found that the
B. subtilis enzyme also required high magnesium ion, above 10 mM,
for cleavage of a hairpin substrate. The results of kinetic
studies showed that the metal ion concentration affected both the
catalysis and the affinity of the ribozymes toward a hairpin RNA
substrate.
```

Commentaires : Syntaxe correcte (ouf !), phrases longues et complexes

Exemples de données textuelles

Résumé d'article scientifique

```
Comparative Analyses of Hairpin Substrate Recognition by
Escherichia coli and Bacillus subtilis Ribonuclease P Ribozymes.

Ando T, Tanaka T, Kikuchi Y.

Division of Bioscience and Biotechnology, Department of Ecological
Engineering, Toyohashi University of Technology.

Previously, we reported that the substrate shape recognition of
the Escherichia coli ribonuclease (RNase) P ribozyme depends on
the concentration of magnesium ion in vitro. We additionally
examined the Bacillus subtilis RNase P ribozyme and found that the
B. subtilis enzyme also required high magnesium ion, above 10 mM,
for cleavage of a hairpin substrate. The results of kinetic
studies showed that the metal ion concentration affected both the
catalysis and the affinity of the ribozymes toward a hairpin RNA
substrate.
```

Commentaires : Abréviations

Exemples de données textuelles

Résumé d'article scientifique

```
Comparative Analyses of Hairpin Substrate Recognition by
Escherichia coli and Bacillus subtilis Ribonuclease P Ribozymes.

Ando T, Tanaka T, Kikuchi Y.

Division of Bioscience and Biotechnology, Department of Ecological
Engineering, Toyohashi University of Technology.

Previously, we reported that the substrate shape recognition of
the Escherichia coli ribonuclease (RNase) P ribozyme depends on
the concentration of magnesium ion in vitro. We additionally
examined the Bacillus subtilis RNase P ribozyme and found that the
B. subtilis enzyme also required high magnesium ion, above 10 mM,
for cleavage of a hairpin substrate. The results of kinetic
studies showed that the metal ion concentration affected both the
catalysis and the affinity of the ribozymes toward a hairpin RNA
substrate.
```

Commentaires : Abréviations, vocabulaire et sémantique spécialisés

Exemples de données textuelles

Résumé d'article scientifique

```
Comparative Analyses of Hairpin Substrate Recognition by
Escherichia coli and Bacillus subtilis Ribonuclease P Ribozymes.

Ando T, Tanaka T, Kikuchi Y.

Division of Bioscience and Biotechnology, Department of Ecological
Engineering, Toyohashi University of Technology.

Previously, we reported that the substrate shape recognition of
the Escherichia coli ribonuclease (RNase) P ribozyme depends on
the concentration of magnesium ion in vitro. We additionally
examined the Bacillus subtilis RNase P ribozyme and found that the
B. subtilis enzyme also required high magnesium ion, above 10 mM,
for cleavage of a hairpin substrate. The results of kinetic
studies showed that the metal ion concentration affected both the
catalysis and the affinity of the ribozymes toward a hairpin RNA
substrate.
```

Commentaires : Variation sémantique

Exemples de données textuelles

Message issues de forum de discussion

```
Sujet: galere pour initialiser un iphone 4
Auteur: tirando
Posté le 15-09-2016 à 11:52:13

Bonjour

On ma donne un iphone 4 desimlocker et a priori depareillé du compte
icloud
Je l ai reinitialisé via itunes (dabord sans carte sim et dans un 2eme
essai avec une carte sim dedans)
j ai cree un compte itunes (apple)
Quand il faut me connecter a mon nouveau compte apple au demarrage de
l iphone ca merde. Le serveur n a pas pu se connecte... Ca fait une
semaine que je suis dessus.
```

Exemples de données textuelles

Message issues de forum de discussion

```
Sujet: galere pour initialiser un iphone 4
Auteur: tirando
Posté le 15-09-2016 à 11:52:13

Bonjour

On ma donne un iphone 4 desimlocker et a priori depareillé du compte
icloud
Je l ai reinitialisé via itunes (dabord sans carte sim et dans un 2eme
essai avec une carte sim dedans)
j ai cree un compte itunes (apple)
Quand il faut me connecter a mon nouveau compte apple au demarrage de
l iphone ca merde. Le serveur n a pas pu se connecte... Ca fait une
semaine que je suis dessus.
```

Commentaires :

Exemples de données textuelles

Message issues de forum de discussion

```
Sujet: galere pour initialiser un iphone 4
Auteur: tirando
Posté le 15-09-2016 à 11:52:13

Bonjour

On ma donne un iphone 4 desimlocker et a priori depareillé du compte
icloud
Je l ai reinitialisé via itunes (dabord sans carte sim et dans un 2eme
essai avec une carte sim dedans)
j ai cree un compte itunes (apple)
Quand il faut me connecter a mon nouveau compte apple au demarrage de
l iphone ca merde. Le serveur n a pas pu se connecte... Ca fait une
semaine que je suis dessus.
```

Commentaires : accents sur caractères accentués quasiment toujours absents

Exemples de données textuelles

Message issues de forum de discussion

```
Sujet: galere pour initialiser un iphone 4
Auteur: tirando
Posté le 15-09-2016 à 11:52:13

Bonjour

On ma donne un iphone 4 desimlocker et a priori depareillé du compte
icloud
Je l ai reinitialisé via itunes (dabord sans carte sim et dans un 2eme
essai avec une carte sim dedans)
j ai cree un compte itunes (apple)
Quand il faut me connecter a mon nouveau compte apple au demarrage de
l iphone ca merde. Le serveur n a pas pu se connecte... Ca fait une
semaine que je suis dessus.
```

Commentaires : apostrophes manquants (espace, ou mots *regroupés*)

Exemples de données textuelles

Message issues de forum de discussion

```
Sujet: galere pour initialiser un iphone 4
Auteur: tirando
Posté le 15-09-2016 à 11:52:13

Bonjour

On ma donne un iphone 4 desimlocker et a priori depareillé du compte
icloud
Je l ai reinitialisé via itunes (dabord sans carte sim et dans un 2eme
essai avec une carte sim dedans)
j ai cree un compte itunes (apple)
Quand il faut me connecter a mon nouveau compte apple au demarrage de
l iphone ca merde. Le serveur n a pas pu se connecte... Ca fait une
semaine que je suis dessus.
```

Commentaires : peu de majuscule (iphone, itunes, apple)

Exemples de données textuelles

Message issues de forum de discussion

```
Sujet: galere pour initialiser un iphone 4
Auteur: tirando
Posté le 15-09-2016 à 11:52:13

Bonjour

On ma donne un iphone 4 desimlocker et a priori depareillé du compte
icloud
Je l ai reinitialisé via itunes (dabord sans carte sim et dans un 2eme
essai avec une carte sim dedans)
j ai cree un compte itunes (apple)
Quand il faut me connecter a mon nouveau compte apple au demarrage de
l iphone ca merde. Le serveur n a pas pu se connecte... Ca fait une
semaine que je suis dessus.
```

Commentaires : ponctuation approximative (points de fin de phrases absents)

Exemples de données textuelles

Message issues de forum de discussion

```
Sujet: galere pour initialiser un iphone 4
Auteur: tirando
Posté le 15-09-2016 à 11:52:13

Bonjour

On ma donne un iphone 4 desimlocker et a priori depareillé du compte
icloud
Je l ai reinitialisé via itunes (dabord sans carte sim et dans un 2eme
essai avec une carte sim dedans)
j ai cree un compte itunes (apple)
Quand il faut me connecter a mon nouveau compte apple au demarrage de
l iphone ca merde. Le serveur n a pas pu se connecte... Ca fait une
semaine que je suis dessus.
```

Commentaires : fautes d'orthographe, d'accord ou de conjugaison

Exemples de données textuelles

Message issues de forum de discussion

```
Sujet: galere pour initialiser un iphone 4
Auteur: tirando
Posté le 15-09-2016 à 11:52:13

Bonjour

On ma donne un iphone 4 desimlocker et a priori depareillé du compte
icloud
Je l ai reinitialisé via itunes (dabord sans carte sim et dans un 2eme
essai avec une carte sim dedans)
j ai cree un compte itunes (apple)
Quand il faut me connecter a mon nouveau compte apple au demarrage de
l iphone ca merde. Le serveur n a pas pu se connecte... Ca fait une
semaine que je suis dessus.
```

Commentaires : *franglais*

Exemples de données textuelles

Message issues de forum de discussion

```
Sujet: galere pour initialiser un iphone 4
Auteur: tirando
Posté le 15-09-2016 à 11:52:13

Bonjour

On ma donne un iphone 4 desimlocker et a priori depareillé du compte
icloud
Je l ai reinitialisé via itunes (dabord sans carte sim et dans un 2eme
essai avec une carte sim dedans)
j ai cree un compte itunes (apple)
Quand il faut me connecter a mon nouveau compte apple au demarrage de
l iphone ca merde. Le serveur n a pas pu se connecte... Ca fait une
semaine que je suis dessus.
```

Commentaires : *franglais*, semantique approximative

Exemples de données textuelles

Message issues de forum de discussion

```
Sujet: galere pour initialiser un iphone 4
Auteur: tirando
Posté le 15-09-2016 à 11:52:13

Bonjour

On ma donne un iphone 4 desimlocker et a priori depareillé du compte
icloud
Je l ai reinitialisé via itunes (dabord sans carte sim et dans un 2eme
essai avec une carte sim dedans)
j ai cree un compte itunes (apple)
Quand il faut me connecter a mon nouveau compte apple au demarrage de
l iphone ca merde. Le serveur n a pas pu se connecte... Ca fait une
semaine que je suis dessus.
```

Commentaires : niveau de langue variable...

Exemples de données textuelles

Message issues de forum de discussion

```
Sujet : fichiers corrompu supprimer sur Iphone 4S 64go
Auteur: Phoenix_spyme
Posté le 02-09-2016 à 23:16:12

Yo!
J'ai un soucis avec mon Iphone 4S.
Lors d'une synchronisation de mon téléphone pour y charger de la musique celui ci planté,
mettant la totalité de ma bibliothèque down, soit 5000 fichiers et des bananes (j'ai un 64 go)...
J'ai tenté une restauration des paramètre (depuis l'Iphone), c'est mieux , j'ai pu de nouveau
syncroniser mon iphone mais toujours pas de musique, enfin pas depuis mon Iphone, elle est par
contre visible depuis Itunes.
J'ai décidé de repartir de 0 et j'ai deleter toute la musique que j'avais manuellement (depuis
Itunes) mais ma mémoire est toujours encombré (Autre dans la barre de stockage en bas) de
fichiers corrompu (et visiblement ca serait ma musique...)
J'ai dll PhoneClean pour faire du ménage il voit bien les fichiers corrompu mais impossible de
les restaurer ou meme de les supprimer (il indique dans la cathégorie "temps" 0 secondes pour
la totalité des morceaux),je pense que c'est du a la suppression que j'ai effectué depuis Itunes.

Comment puis je faire pour virer ces fichiers corrompu? Dois je tout supprimer le contenue de
ma mémoire et restaurer mon iphone a une date antérieure depuis Itunes(qu'est ce que ca
implique de restaurer son Iphone?)

Merci! C'est mon dernier espoir avant de passer sous Android...
```

Exemples de données textuelles

Message issues de forum de discussion

```
Sujet : fichiers corrompu supprimer sur Iphone 4S 64go
Auteur: Phoenix_spyme
Posté le 02-09-2016 à 23:16:12

Yo!
J'ai un soucis avec mon Iphone 4S.
Lors d'une synchronisation de mon téléphone pour y charger de la musique celui ci planté,
mettant la totalité de ma bibliothèque down, soit 5000 fichiers et des bananes (j'ai un 64 go)...
J'ai tenté une restauration des paramètre (depuis l'Iphone), c'est mieux , j'ai pu de nouveau
syncroniser mon iphone mais toujours pas de musique, enfin pas depuis mon Iphone, elle est par
contre visible depuis Itunes.
J'ai décidé de repartir de 0 et j'ai deleter toute la musique que j'avais manuellement (depuis
Itunes) mais ma mémoire est toujours encombré (Autre dans la barre de stockage en bas) de
fichiers corrompu (et visiblement ca serait ma musique...)
J'ai dll PhoneClean pour faire du ménage il voit bien les fichiers corrompu mais impossible de
les restaurer ou meme de les supprimer (il indique dans la cathégorie "temps" 0 secondes pour
la totalité des morceaux),je pense que c'est du a la suppression que j'ai effectué depuis Itunes.

Comment puis je faire pour virer ces fichiers corrompu? Dois je tout supprimer le contenue de
ma mémoire et restaurer mon iphone a une date antérieure depuis Itunes(qu'est ce que ca
implique de restaurer son Iphone?)

Merci! C'est mon dernier espoir avant de passer sous Android...
```

Commentaires :

Exemples de données textuelles

Message issues de forum de discussion

```
Sujet : fichiers corrompu supprimer sur Iphone 4S 64go
Auteur: Phoenix_spyme
Posté le 02-09-2016 à 23:16:12

Yo!
J'ai un soucis avec mon Iphone 4S.
Lors d'une synchronisation de mon téléphone pour y charger de la musique celui ci planté,
mettant la totalité de ma bibliothèque down, soit 5000 fichiers et des bananes (j'ai un 64 go)...
J'ai tenté une restauration des paramètre (depuis l'Iphone), c'est mieux , j'ai pu de nouveau
syncroniser mon iphone mais toujours pas de musique, enfin pas depuis mon Iphone, elle est par
contre visible depuis Itunes.
J'ai décidé de repartir de 0 et j'ai deleter toute la musique que j'avais manuellement (depuis
Itunes) mais ma mémoire est toujours encombré (Autre dans la barre de stockage en bas) de
fichiers corrompu (et visiblement ca serait ma musique...)
J'ai dll PhoneClean pour faire du ménage il voit bien les fichiers corrompu mais impossible de
les restaurer ou meme de les supprimer (il indique dans la cathégorie "temps" 0 secondes pour
la totalité des morceaux),je pense que c'est du a la suppression que j'ai effectué depuis Itunes.

Comment puis je faire pour virer ces fichiers corrompu? Dois je tout supprimer le contenue de
ma mémoire et restaurer mon iphone a une date antérieure depuis Itunes(qu'est ce que ca
implique de restaurer son Iphone?)

Merci! C'est mon dernier espoir avant de passer sous Android...
```

Commentaires : accents sur caractères accentués parfois absents

Exemples de données textuelles

Message issues de forum de discussion

```
Sujet : fichiers corrompu supprimer sur Iphone 4S 64go
Auteur: Phoenix_spyme
Posté le 02-09-2016 à 23:16:12

Yo!
J'ai un soucis avec mon Iphone 4S.
Lors d'une synchronisation de mon téléphone pour y charger de la musique celui ci planté,
mettant la totalité de ma bibliothèque down, soit 5000 fichiers et des bananes (j'ai un 64 go)...
J'ai tenté une restauration des paramètre (depuis l'Iphone), c'est mieux , j'ai pu de nouveau
syncroniser mon iphone mais toujours pas de musique, enfin pas depuis mon Iphone, elle est par
contre visible depuis Itunes.
J'ai décidé de repartir de 0 et j'ai deleter toute la musique que j'avais manuellement (depuis
Itunes) mais ma mémoire est toujours encombré (Autre dans la barre de stockage en bas) de
fichiers corrompu (et visiblement ca serait ma musique...)
J'ai dll PhoneClean pour faire du ménage il voit bien les fichiers corrompu mais impossible de
les restaurer ou meme de les supprimer (il indique dans la cathégorie "temps" 0 secondes pour
la totalité des morceaux),je pense que c'est du a la suppression que j'ai effectué depuis Itunes.

Comment puis je faire pour virer ces fichiers corrompu? Dois je tout supprimer le contenue de
ma mémoire et restaurer mon iphone a une date antérieure depuis Itunes(qu'est ce que ca
implique de restaurer son Iphone?)

Merci! C'est mon dernier espoir avant de passer sous Android...
```

Commentaires : certains points de fin de phrases absents

Exemples de données textuelles

Message issues de forum de discussion

```
Sujet : fichiers corrompu supprimer sur Iphone 4S 64go
Auteur: Phoenix_spyme
Posté le 02-09-2016 à 23:16:12

Yo!
J'ai un soucis avec mon Iphone 4S.
Lors d'une synchronisation de mon téléphone pour y charger de la musique celui ci planté,
mettant la totalité de ma bibliothèque down, soit 5000 fichiers et des bananes (j'ai un 64 go)...
J'ai tenté une restauration des paramètre (depuis l'Iphone), c'est mieux , j'ai pu de nouveau
syncroniser mon iphone mais toujours pas de musique, enfin pas depuis mon Iphone, elle est par
contre visible depuis Itunes.
J'ai décidé de repartir de 0 et j'ai deleter toute la musique que j'avais manuellement (depuis
Itunes) mais ma mémoire est toujours encombré (Autre dans la barre de stockage en bas) de
fichiers corrompu (et visiblement ca serait ma musique...)
J'ai dll PhoneClean pour faire du ménage il voit bien les fichiers corrompu mais impossible de
les restaurer ou meme de les supprimer (il indique dans la cathégorie "temps" 0 secondes pour
la totalité des morceaux),je pense que c'est du a la suppression que j'ai effectué depuis Itunes.

Comment puis je faire pour virer ces fichiers corrompu? Dois je tout supprimer le contenue de
ma mémoire et restaurer mon iphone a une date antérieure depuis Itunes(qu'est ce que ca
implique de restaurer son Iphone?)

Merci! C'est mon dernier espoir avant de passer sous Android...
```

Commentaires : fautes d'orthographe

Exemples de données textuelles

Message issues de forum de discussion

```
Sujet : fichiers corrompu supprimer sur Iphone 4S 64go
Auteur: Phoenix_spyme
Posté le 02-09-2016 à 23:16:12

Yo!
J'ai un soucis avec mon Iphone 4S.
Lors d'une synchronisation de mon téléphone pour y charger de la musique celui ci planté,
mettant la totalité de ma bibliothèque down, soit 5000 fichiers et des bananes (j'ai un 64 go)...
J'ai tenté une restauration des paramètre (depuis l'Iphone), c'est mieux , j'ai pu de nouveau
syncroniser mon iphone mais toujours pas de musique, enfin pas depuis mon Iphone, elle est par
contre visible depuis Itunes.
J'ai décidé de repartir de 0 et j'ai deleter toute la musique que j'avais manuellement (depuis
Itunes) mais ma mémoire est toujours encombré (Autre dans la barre de stockage en bas) de
fichiers corrompu (et visiblement ca serait ma musique...)
J'ai dll PhoneClean pour faire du ménage il voit bien les fichiers corrompu mais impossible de
les restaurer ou meme de les supprimer (il indique dans la cathégorie "temps" 0 secondes pour
la totalité des morceaux),je pense que c'est du a la suppression que j'ai effectué depuis Itunes.

Comment puis je faire pour virer ces fichiers corrompu? Dois je tout supprimer le contenue de
ma mémoire et restaurer mon iphone a une date antérieure depuis Itunes(qu'est ce que ca
implique de restaurer son Iphone?)

Merci! C'est mon dernier espoir avant de passer sous Android...
```

Commentaires : fautes d'orthographe, d'accord ou de conjugaison

Exemples de données textuelles

Message issues de forum de discussion

```
Sujet : fichiers corrompu supprimer sur Iphone 4S 64go
Auteur: Phoenix_spyme
Posté le 02-09-2016 à 23:16:12

Yo!
J'ai un soucis avec mon Iphone 4S.
Lors d'une synchronisation de mon téléphone pour y charger de la musique celui ci planté,
mettant la totalité de ma bibliothèque down, soit 5000 fichiers et des bananes (j'ai un 64 go)...
J'ai tenté une restauration des paramètre (depuis l'Iphone), c'est mieux , j'ai pu de nouveau
syncroniser mon iphone mais toujours pas de musique, enfin pas depuis mon Iphone, elle est par
contre visible depuis Itunes.
J'ai décidé de repartir de 0 et j'ai deleter toute la musique que j'avais manuellement (depuis
Itunes) mais ma mémoire est toujours encombré (Autre dans la barre de stockage en bas) de
fichiers corrompu (et visiblement ca serait ma musique...)
J'ai dll PhoneClean pour faire du ménage il voit bien les fichiers corrompu mais impossible de
les restaurer ou meme de les supprimer (il indique dans la cathégorie "temps" 0 secondes pour
la totalité des morceaux),je pense que c'est du a la suppression que j'ai effectué depuis Itunes.

Comment puis je faire pour virer ces fichiers corrompu? Dois je tout supprimer le contenue de
ma mémoire et restaurer mon iphone a une date antérieure depuis Itunes(qu'est ce que ca
implique de restaurer son Iphone?)

Merci! C'est mon dernier espoir avant de passer sous Android...
```

Commentaires : *franglais*

Exemples de données textuelles

Message issues de forum de discussion

```
Sujet : fichiers corrompu supprimer sur Iphone 4S 64go
Auteur: Phoenix_spyme
Posté le 02-09-2016 à 23:16:12

Yo!
J'ai un soucis avec mon Iphone 4S.
Lors d'une synchronisation de mon téléphone pour y charger de la musique celui ci planté,
mettant la totalité de ma bibliothèque down, soit 5000 fichiers et des bananes (j'ai un 64 go)...
J'ai tenté une restauration des paramètre (depuis l'Iphone), c'est mieux , j'ai pu de nouveau
syncroniser mon iphone mais toujours pas de musique, enfin pas depuis mon Iphone, elle est par
contre visible depuis Itunes.
J'ai décidé de repartir de 0 et j'ai deleter toute la musique que j'avais manuellement (depuis
Itunes) mais ma mémoire est toujours encombré (Autre dans la barre de stockage en bas) de
fichiers corrompu (et visiblement ca serait ma musique...)
J'ai dll PhoneClean pour faire du ménage il voit bien les fichiers corrompu mais impossible de
les restaurer ou meme de les supprimer (il indique dans la cathégorie "temps" 0 secondes pour
la totalité des morceaux),je pense que c'est du a la suppression que j'ai effectué depuis Itunes.

Comment puis je faire pour virer ces fichiers corrompu? Dois je tout supprimer le contenue de
ma mémoire et restaurer mon iphone a une date antérieure depuis Itunes(qu'est ce que ca
implique de restaurer son Iphone?)

Merci! C'est mon dernier espoir avant de passer sous Android...
```

Commentaires : *franglais*, niveau de langue variable

Exemples de données textuelles

Message issues de forum de discussion

```
Sujet : fichiers corrompu supprimer sur Iphone 4S 64go
Auteur: Phoenix_spyme
Posté le 02-09-2016 à 23:16:12

Yo!
J'ai un soucis avec mon Iphone 4S.
Lors d'une synchronisation de mon téléphone pour y charger de la musique celui ci planté,
mettant la totalité de ma bibliothèque down, soit 5000 fichiers et des bananes (j'ai un 64 go)...
J'ai tenté une restauration des paramètre (depuis l'Iphone), c'est mieux , j'ai pu de nouveau
syncroniser mon iphone mais toujours pas de musique, enfin pas depuis mon Iphone, elle est par
contre visible depuis Itunes.
J'ai décidé de repartir de 0 et j'ai deleter toute la musique que j'avais manuellement (depuis
Itunes) mais ma mémoire est toujours encombré (Autre dans la barre de stockage en bas) de
fichiers corrompu (et visiblement ca serait ma musique...)
J'ai dll PhoneClean pour faire du ménage il voit bien les fichiers corrompu mais impossible de
les restaurer ou meme de les supprimer (il indique dans la cathégorie "temps" 0 secondes pour
la totalité des morceaux),je pense que c'est du a la suppression que j'ai effectué depuis Itunes.

Comment puis je faire pour virer ces fichiers corrompu? Dois je tout supprimer le contenue de
ma mémoire et restaurer mon iphone a une date antérieure depuis Itunes(qu'est ce que ca
implique de restaurer son Iphone?)

Merci! C'est mon dernier espoir avant de passer sous Android...
```

Commentaires : interprétation nécessitant une connaissance du monde

Exemples de données textuelles

Message issues de forum de discussion

```
Sujet : fichiers corrompu supprimer sur Iphone 4S 64go
Auteur: Phoenix_spyme
Posté le 02-09-2016 à 23:16:12

Yo!
J'ai un soucis avec mon Iphone 4S.
Lors d'une synchronisation de mon téléphone pour y charger de la musique celui ci planté,
mettant la totalité de ma bibliothèque down, soit 5000 fichiers et des bananes (j'ai un 64 go)...
J'ai tenté une restauration des paramètre (depuis l'Iphone), c'est mieux , j'ai pu de nouveau
syncroniser mon iphone mais toujours pas de musique, enfin pas depuis mon Iphone, elle est par
contre visible depuis Itunes.
J'ai décidé de repartir de 0 et j'ai deleter toute la musique que j'avais manuellement (depuis
Itunes) mais ma mémoire est toujours encombré (Autre dans la barre de stockage en bas) de
fichiers corrompu (et visiblement ca serait ma musique...)
J'ai dll PhoneClean pour faire du ménage il voit bien les fichiers corrompu mais impossible de
les restaurer ou meme de les supprimer (il indique dans la cathégorie "temps" 0 secondes pour
la totalité des morceaux),je pense que c'est du a la suppression que j'ai effectué depuis Itunes.

Comment puis je faire pour virer ces fichiers corrompu? Dois je tout supprimer le contenue de
ma mémoire et restaurer mon iphone a une date antérieure depuis Itunes(qu'est ce que ca
implique de restaurer son Iphone?)

Merci! C'est mon dernier espoir avant de passer sous Android...
```

Commentaires : variation sémantique

Exemples de données textuelles

Recette de cuisine - texte procédural

```
Tartiflette
Temps de préparation : 30 minutes
Temps de cuisson : 20 minutes

Ingrédients (pour 6 personnes) :
- 1 reblochon et demi (2 pour les gourmands)
- 1,2 kg de pommes de terre
- 1 tranche de jambon fumé coupée en dés (ou bien des lardons ou des dés de bacon)
- 500 g d'oignons
- 40 cl d'Apremont ou autre blanc de Savoie sec (facultatif, mais donne plus de goût)
- 1 pincée de sel
- 1 pincée de poivre

Préparation de la recette :
Faites cuire les pommes de terre avec leur peau. Épluchez-les et coupez-les en rondelles.
Émincez les oignons puis faites-les revenir avec un peu de beurre. Une fois dorés, y ajouter le
jambon (ou les lardons), ainsi que les pommes de terre. Laissez mijoter 15 min.
Si vous utilisez du vin blanc, c'est le moment de l'ajouter. Salez un peu, poivrez, laissez les
pommes de terre s'imprégner du vin blanc quelques minutes avant de transférer le tout dans un
plat à gratin.
Grattez au couteau les reblochons, coupez-les en 2 dans le sens de l'épaisseur et posez les sur
les pommes de terre.
Faites cuire à four chaud (220°C) pendant 20 à 30 min. Servez avec une salade verte, voire
quelques tomates, juste assaisonnées d'un peu de vinaigre d'échalote.

Boisson conseillée :
Apremont, blanc de savoie sec
```

Exemples de données textuelles

Recette de cuisine - texte procédural

```
Tiramisu
Temps de préparation : 20 minutes
Temps de cuisson : 0 minutes

Ingrédients (pour 6 personnes) :
- 3 gros oeufs
- 40 g de cassonade
- 125 g de mascarpone
- 120 g de petit-suisse à 3,8% de matière grasse
- 10 cl de café noir non sucré
- 10 g de poudre de cacao amer
- 1 cuillère à soupe de rhum vieux ou d'amaretto
- 12 biscuits à la cuillère

Préparation de la recette :
Séparer les blancs des jaunes.
Blanchir les jaunes avec la cassonade à l'aide d'un fouet électrique. Ajouter le mascarpone et
les petits-suisses puis bien fouetter. Ajouter l'alcool.
Monter les blancs en neige et les incorporer délicatement à la spatule au mélange précédent.
Réserver.

Préparer du café noir et y ajouter l'alcool.
Couper les biscuits en deux si besoin et les tremper légèrement dans le café.
Tapisser chaque verrine (24 cl) de deux moitiés de biscuits. Recouvrir de deux cuillères à
soupe de crème, d'une couche de deux moitiés de biscuits mouillés puis terminer avec deux
cuillères à soupe de crème. Mettre au réfrigérateur 6 heures minimum mais encore meilleur si
préparé la veille et laissé au frais toute la nuit.
Saupoudrer de cacao au moment de servir et déguster frais.
```

Exemples de données textuelles

Recette de cuisine - texte procédural

(suite)

```
Pour une 1 part :
- environ 240 calories

Équivalences dans un repas :
- 1 fruit moyen
- 1 portion de fromage (environ 30 g)
- 1 cuillère à café d'huile

Astuces : Si on décide de mettre plus de biscuits soit 4 par verrine au lieu de 2, on
rajoutera dans l'équivalence repas : 1 tranche de pain

Valeurs nutritionnelles :
- protéines : environ 5 g
- lipides : environ 15 g
- glucides : environ 20 g

Remarques :
Cette recette adaptée pour une personne diabétique a été élaborée en collaboration avec les
diététiciennes Jacqueline Delaunay, Céline Benoit et Patricia Fiquet
```

Exemples de données textuelles

Extraits de dossiers patients anonymisés

```
DISCHARGE MEDICATIONS:
HYDROCORTISONE 2.5% -RECTAL CREAM TP BID
Instructions: Apply to hemorrhoids
BEN-GAY TOPICAL TP BID
Instructions: Apply liberally to legs
Alert overridden: Override added on 9/8/03 by
FACK , PASQUALE DIEGO , M.D.
DEFINITE ALLERGY ( OR SENSITIVITY ) to SALICYLATES
Reason for override: aware
PREMARIN ( CONJUGATED ESTROGENS ) 1.25 MG PO QD
LASIX ( FUROSEMIDE ) 60 MG qam; 40 MG qpm PO BID 60 MG qam
40 MG qpm Starting Today ( 0/29 )
METAMUCIL SUGAR FREE ( PSYLLIUM ( METAMUCIL ) SU... )
1 PACKET PO TID Instructions: With meals
NORVASC ( AMLODIPINE ) 10 MG PO QD
Food/Drug Interaction Instruction
Avoid grapefruit unless MD instructs otherwise.
AMBIEN ( ZOLPIDEM TARTRATE ) 10 MG PO QHS PRN insomnia
```

Exemples de données textuelles

Extraits de dossiers patients anonymisés

```
RRR , lots of BS's , neuro nonfocal , ext with 1+ edema. On atenolol ,
zestril , norvasc , premarin , detrol , lasix 60 qd , nebs prn at home.
Labs sig for Cr 0.7 , CK 48 , TnI .05 , QBC 9.5 , Hct 41.3. From CV
point of view , thought to be CHF exac. ROMI'd without events on monitor
and diuresed 2L/day. IV Lasix 80 bid to start transitioned to 60 po
bid. BNP>assay. 6/17 dobut MIBI with mod sized ant septal wall defect
c/w diagonal lesion , 3/22 Echo with EF 55-60\% , mild LAE/RAE ,
no WMA , mod large RV. No further CV studies. Cont previously meds on
d/c. From FEN point of view , 2 L fluid restriction , 2 g Na
restriction. Nutrition consult , but pt very resistant to diet changes.
From GI point of view , GERD; nexium started. From pulm point of view ,
CXR c/w sl fluid overload , no focal findings , no pulm edema. Given NC
O2 and BiPAP at night.
```

Exemples de données textuelles

Extraits de dossiers patients anonymisés

```
The patient is currently off diuretics at this time. Daily
weights should be checked and if her weight increases by more
than 3 pounds Dr. Bockoven should be notified. The patient was also
started on calcitriol given elevation of parathyroid hormone.
Cardiovascular: Rate and rhythm: The patient has a history of
atrial fibrillation with a slow ventricular response. The
patient was started on metoprolol 12.5 mg p.o. q.6 h. for rate
control , however , this dose was decreased to 12.5 mg p.o. twice a
day, given some bradycardia on her telemetry. The patient was
also started on Flecainide 75 mg p.o. q.12 h. She will continue
on these two medications upon discharge.
```

Recherche vs. Découvrir

	Recherche	Découverte
données structurées	Extraction de données *BD*	Fouille de données *BD, entrepôts*
données non structurées	Recherche d'information *collection de textes (textes indexés)*	Fouille de textes *collection de textes (corpus)*

Méthodologie générale

1. Nettoyage et pré-traitement du corpus brut

Méthodologie générale

1. Nettoyage et pré-traitement du corpus brut
2. Reconnaissance/annotation des entités nommés (EN) et des termes

Méthodologie générale

1. Nettoyage et pré-traitement du corpus brut
2. Reconnaissance/annotation des entités nommés (EN) et des termes
3. Identification de relations entre EN et termes

Méthodologie générale

1. Nettoyage et pré-traitement du corpus brut
2. Reconnaissance/annotation des entités nommés (EN) et des termes
3. Identification de relations entre EN et termes
   Dans les deux cas : utilisation de
   - ressources (terminologies, liste d'entités nommées)

Méthodologie générale

1. Nettoyage et pré-traitement du corpus brut
2. Reconnaissance/annotation des entités nommés (EN) et des termes
3. Identification de relations entre EN et termes
   Dans les deux cas : utilisation de
   - ressources (terminologies, liste d'entités nommées)
   - régles d'extraction (expression régulières, etc.)

Méthodologie générale

1. Nettoyage et pré-traitement du corpus brut
2. Reconnaissance/annotation des entités nommés (EN) et des termes
3. Identification de relations entre EN et termes
   Dans les deux cas : utilisation de
   - ressources (terminologies, liste d'entités nommées)
   - régles d'extraction (expression régulières, etc.)
   - modèles d'apprentissage obtenus à partir d'un corpus d'exemples

Méthodologie générale

1. Nettoyage et pré-traitement du corpus brut
2. Reconnaissance/annotation des entités nommés (EN) et des termes
3. Identification de relations entre EN et termes
   Dans les deux cas : utilisation de
   - ressources (terminologies, liste d'entités nommées)
   - régles d'extraction (expression régulières, etc.)
   - modèles d'apprentissage obtenus à partir d'un corpus d'exemples
     - apprentissage supervisé (CRF, Arbre de décision, SVM, etc.)

Méthodologie générale

1. Nettoyage et pré-traitement du corpus brut
2. Reconnaissance/annotation des entités nommés (EN) et des termes
3. Identification de relations entre EN et termes
   Dans les deux cas : utilisation de
   - ressources (terminologies, liste d'entités nommées)
   - régles d'extraction (expression régulières, etc.)
   - modèles d'apprentissage obtenus à partir d'un corpus d'exemples
     - apprentissage supervisé (CRF, Arbre de décision, SVM, etc.)
     - apprentissage non-supervisé (règles d'association, k-means, etc.)

Méthodologie générale

1. Nettoyage et pré-traitement du corpus brut
2. Reconnaissance/annotation des entités nommés (EN) et des termes
3. Identification de relations entre EN et termes
   Dans les deux cas : utilisation de
   - ressources (terminologies, liste d'entités nommées)
   - régles d'extraction (expression régulières, etc.)
   - modèles d'apprentissage obtenus à partir d'un corpus d'exemples
     - apprentissage supervisé (CRF, Arbre de décision, SVM, etc.)
     - apprentissage non-supervisé (règles d'association, k-means, etc.)
4. Extraction d'information

Données textuelles

Données structurées vs. données non-structurés

- Données structurées
  données:
  - stockées dans une structure logique (BD, enregistrement, etc.)
  - consistantes
  - typées

  idéales pour la fouille de données
- Données non-structurées
  - textes libres, rédigés en langage naturel
  - pas de structure logique explicite, au sens informatique du terme
  - mélange d'information de différents types (dates, valeurs numériques, lieu, nom de personne, etc.)

  Le matériau principal de la fouille de textes

Données semi-structurées

- Combinaison de texte libre et de données structurés : XML, RDF
- Exemples
  1. Une partie contient des informations directement stockables dans une base de données
     Une autre partie contient des zones de texte libre

Exemple

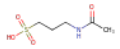
DrugBank

Identification | Taxonomy | Pharmacology | ADMET | Pharmacoeconomics | Properties | Spectra | References | Interactions | Targets (3) | Transporters (1) | Comments

Description

Acamprosate, also known by the brand name Campral™, is a drug used for treating alcohol dependence. Acamprosate is thought to stabilize the chemical balance in the brain that would otherwise be disrupted by alcoholism, possibly by blocking glutaminergic N-methyl-D-aspartate receptors, while gamma-aminobutyric acid type A receptors are activated. Reports indicate that acamprosate only works with a combination of attending support groups and abstinence from alcohol. Certain serious side effects include allergic reactions, irregular heartbeats, and low or high blood pressure, while less serious side effects include headaches, insomnia, and impotence. Acamprosate should not be taken by people with kidney problems or allergies to the drug.

Structure



MOL | SDF | 3D-SDF | PDB | SMILES | InChI | View 3D Structure

Synonyms

- 3-Acetamido-1-propanesulfonic acid
- Acamprosato
- Acamprosatum
- N-acetyl homotaurine
- N-Acetylhomotaurine

External Identifiers

Not Available

Approved Prescription Products

Show 10 entries — Search

Name	Dosage	Strength	Route	Labeller	Marketing Start	Marketing End	
Campral	tablet, delayed release	333 mg/1	oral	Forest Laboratories, Inc.	2005-01-11	2016-02-29	
Campral	tablet, delayed release	333 mg/1	oral	Carilion Materials Management	2005-01-11	Not applicable	
Campral	tablet (delayed-release)	333 mg	oral	Mylan Pharmaceuticals Ulc	2007-07-30	Not applicable	

Showing 1 to 3 of 3 entries — Previous 1 Next

Identification | Taxonomy | Pharmacology | ADMET | Pharmacoeconomics | Properties | Spectra | References | Interactions | Targets (3) | Transporters (1) | Comments

Pharmacology

Indication

For the maintenance of abstinence from alcohol in patients with alcohol dependence who are abstinent at treatment initiation

Pharmacodynamics

Pharmacodynamic studies have shown that acamprosate calcium reduces alcohol intake in alcohol-dependent animals in a dose-dependent manner and that this effect appears to be specific to alcohol and the mechanisms of alcohol dependence. Acamprosate calcium has negligible observable central nervous system (CNS) activity in animals outside of its effects on alcohol dependence, exhibiting no anticonvulsant, antidepressant, or anxiolytic activity.

Mechanism of action

The mechanism of action of acamprosate in maintenance of alcohol abstinence is not completely understood. Chronic alcohol exposure is hypothesized to alter the normal balance between neuronal excitation and inhibition. *in vitro* and *in vivo* studies in animals have provided evidence to suggest acamprosate may interact with glutamate and GABA neurotransmitter systems centrally, and has led to the hypothesis that acamprosate restores this balance. It seems to inhibit NMDA receptors while activating GABA receptors.

Related Articles

- New drug helps alcoholics kick the habit — Monica Shah, PharmD et al., Drug Topics, 2004
- Highlights of The American Society of Addiction Medicine 36th Annual Medical-Scientific Conference — Thomas L. Haynes, MD, FASAM, Medscape, 2005
- Acamprosate: a review of its use in alcohol dependence — MDLinx
- Topiramate in the Treatment of Alcohol Use Disorders — Sarah T. Mellon, PharmD, Medscape, 2015
- Highlights From the 13th Annual Meeting of the American Association for Addiction Psychiatry — John Talmadge, MD, Medscape, 2003

Powered by TrendMD

Absorption

The absolute bioavailability of acamprosate after oral administration is about 11%. The food effect on absorption is not clinically significant and no adjustment of dose is necessary.

Volume of distribution

- 72 to 109 L

Protein binding

Non detectable

Metabolism

Acamprosate does not undergo metabolism.

Route of elimination

Following oral administration of CAMPRAL®, the major route of excretion is via the kidneys as acamprosate.

Half life

20 - 33 hours

Clearance

Not Available

Toxicity

In all reported cases of acute overdosage with acamprosate (total reported doses of up to 56 grams of acamprosate calcium), the only symptom that could be reasonably associated with acamprosate was diarrhea.

Données semi-structurées

- Combinaison de texte libre et de données structurés : XML, RDF
- Exemples
  1. Une partie contient des informations directement stockables dans une base de données
     Une autre partie contient des zones de texte libre
  2. Questionnaire contenant une zone de commentaires

Type de données textuelles

Textes ou collection de textes: corpus texteuls
De grandes variations :

- Formats électronique (texte brut, HTML, XML, PDF, Word, etc.)
- Encodage de caractères (ASCII, ISO-LATIN-1, windows-1252, UTF-8, etc.)
- Type de documents (pages web, blogs, articles scientifiques, articles de journaux, livres, tables, messages de forum, emails, SMS, etc.)
- Taille: entre quelques Kilo-octets et plusieurs Giga-octets

→ En général : Manipulation de textes bruts

Texte brut/fichier plein texte

Résumés Medline

1: Biosci Biotechnol Biochem. 2003 Aug;67(8):1825-7. Related Articles, Links

Comparative Analyses of Hairpin Substrate Recognition by Escherichia coli and Bacillus subtilis Ribonuclease P Ribozymes.

Ando T, Tanaka T, Kikuchi Y.

Division of Bioscience and Biotechnology, Department of Ecological Engineering, Toyohashi University of Technology.

Previously, we reported that the substrate shape recognition of the Escherichia coli ribonuclease (RNase) P ribozyme depends on the concentration of magnesium ion in vitro. We additionally examined the Bacillus subtilis RNase P ribozyme and found that the B. subtilis enzyme also required high magnesium ion, above 10 mM, for cleavage of a hairpin substrate. The results of kinetic studies showed that the metal ion concentration affected both the catalysis and the affinity of the ribozymes toward a hairpin RNA substrate.

HTML

Page Web

```
<h1 id="sidrubrik">About me - Sergei Silvestrov</h1>
<div id="artikeldata">

<h4 align="justify"><img hspace="7" vspace="7" align="right"   src="/polopoly_f
<p align="left"><strong>I am Professor in Mathematics and Applied Mathematics</
<p><strong>I am also the Subject Chair for the subject of</strong><strong> </st
<p><strong>Mathematics and Applied Mathematics</strong></p>
<p><strong>in M&auml;lardalen University</strong></p>

<p><strong>(&Auml;mnesf&ouml;retr&auml;dare In Swedish) </strong></p>
<p> <a href="/polopoly_fs/1.49763!/CV%20for%20Sergei%20Silvestrov%201%20si
<p> <!--[if gte mso 9]>
```

→ Nécessite d'extraire le contenu utile

Texte brut issu de la page Web

About me - Sergei Silvestrov

I am Professor in Mathematics and Applied Mathematics

I am also the Subject Chair for the subject of

Mathematics and Applied Mathematics

in Mälardalen University

(Ämnesföreträdare In Swedish)

CV (short, 1 page, pdf)

Page Web Wikipedia

Article Discussion Lire Modifier Modifier le code Historique Rechercher

WIKIPÉDIA
L'encyclopédie libre

Accueil
Portails thématiques
Article au hasard
Contact

Contribuer
Débuter sur Wikipédia
Aide
Communauté
Modifications récentes
Faire un don

Outils
Pages liées
Suivi des pages liées
Importer un fichier
Pages spéciales
Adresse permanente
Information sur la page
Élément Wikidata
Citer cette page

Imprimer / exporter
Créer un livre
Télécharger comme PDF
Version imprimable

Diabète de type 1

*Cet article traite du **diabète de type 1**, une forme de diabète sucré. Il existe d'autres types de diabètes : voir la page d'homonymie Diabète.*

Le **diabète de type 1** ou **diabète insulino-dépendant** (DID), parfois **diabète inné** est une forme de diabète sucré qui apparaît le plus souvent de manière brutale chez l'enfant ou chez le jeune adulte (ou beaucoup plus rarement chez les personnes plus âgées) mais parfois aussi le diabète peut être présent depuis la naissance et ne se manifester qu'à l'adolescence. Il se manifeste par une émission d'urine excessive (polyurie), une soif intense (polydipsie) et un appétit anormalement augmenté (polyphagie). Il a aussi pour conséquence un amaigrissement malgré une prise de nourriture abondante, une hyperglycémie (c'est-à-dire un excès de glucose dans le sang) supérieure à 1,26 g/l de sucre dans le sang à jeun, ou supérieure à 2 g/l (11 mmol/l) à n'importe quel moment de la journée, avec parfois présence d'acétone dans les urines, accompagnée d'une haleine « de pomme reinette » caractéristique.

Les diabétiques de type 1 doivent régulièrement contrôler leur glycémie, s'injecter de l'insuline plusieurs fois par jour, manger de manière équilibrée et pratiquer une activité physique quotidienne.

Sommaire [masquer]

1 Causes
2 Physiopathologie
3 Épidémiologie
4 Diagnostic du DID
4.1 Circonstances d'apparition
4.2 Signes cliniques
4.2.1 Les signes fonctionnels et généraux
4.2.2 Les signes physiques
4.3 Signes biologiques
4.3.1 Auto-anticorps
4.3.2 L'hémoglobine glyquée
4.3.3 L'hyperglycémie
4.3.4 La glycosurie
4.3.5 Les corps cétoniques
5 Pronostic général et évolution

Diabète de type 1

Le cercle bleu, symbole universel du diabète[1]

CIM-10	E10
CIM-9	250.x1, 250.x3
OMIM	222100
DiseasesDB	3649
MedlinePlus	000305
eMedicine	med/546
MeSH	D003922

Mise en garde médicale

modifier

https://commons.wikimedia.org/w/index.php?title=File:Blue_circle_for_diabetes.svg&lang=fr&uselang=fr

Brochures médicales

Weight Bearing Using a Walker

Follow the directions as checked for your injured or weak leg.

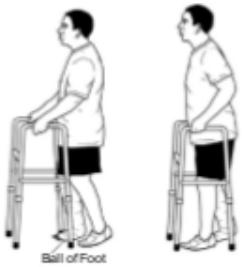
❑ **Non-Weight Bearing**

For this method, do not let your injured or weak leg touch the floor when standing or walking. When using the walker, hold your injured or weak leg up off the floor.

- Move your walker out in front of you. Be sure all 4 legs of your walker are flat on the floor.
- While pushing down on your walker with your arms, hop on your good foot to the center of your walker.

❑ **Touch Down Weight Bearing**

For this method, touch the ball of your foot to the floor to help your balance, but do not put your weight on it.



1

Taking Medicines Safely

Medicines are often part of treatment for illness or injury. Taking medicine is not without some risk for side effects. Follow these tips for taking medicines safely:

- Keep a list of your medicines with you. Include prescription and over the counter medicines, herbs, vitamins and any home remedies you use.
- Show this list to all doctors, pharmacists and dentists who are treating you.
- Take your medicine as ordered by your doctor. Do not change or stop taking any medicine until you talk to your doctor.
- Take your medicines at the same time each day.
- Tell your doctor if you are pregnant, breastfeeding or on a special diet.
- Get a refill at least one week before your medicine is gone.
- Throw away unused or expired medicines.
- Do not take any new herbal products or over the counter medicines without first checking with your doctor or pharmacist. These products may change the way your other medicines work.
- Do not give your medicines to anyone else.

1

Notion de corpus

Des textes sous forme électronique ne font pas un corpus

Un corpus

Notion de corpus

Des textes sous forme électronique ne font pas un corpus

Un corpus

est une collection de textes

Notion de corpus

Des textes sous forme électronique ne font pas un corpus

Un corpus

est une collection de textes
qui sont sélectionnés et organisés

Notion de corpus

Des textes sous forme électronique ne font pas un corpus

Un corpus

est une collection de textes
qui sont sélectionnés et organisés
selon des critères linguistiques et sociologiques

Notion de corpus

Des textes sous forme électronique ne font pas un corpus

Un corpus

- est une collection de textes
- qui sont sélectionnés et organisés
- selon des critères linguistiques et sociologiques
- explicites

Notion de corpus

Des textes sous forme électronique ne font pas un corpus

Un corpus

- est une collection de textes
- qui sont sélectionnés et organisés
- selon des critères linguistiques et sociologiques explicites
- pour servir d'échantillon du langage lié à l'application visée.

Linguistique de corpus

Fondement principal des approches de fouille de texte

- Pas d'*a priori* sur la langue
- L'information nécessaire se trouve dans le corpus
- Corpus de spécialité caractérisés par :
  - un lexique limité (termes, synonymes)
  - schémas syntaxiques particuliers
- Disponibilité d'outils nécessaires
  dépouillement et traitements du corpus

Types de corpus

Corpus de référence

Corpus spécialisé

Mais aussi, Corpus de presse, de dépêches

Corpus de référence

- tend à fournir une vue complète d'une langue
  description des caractéristiques linguistiques
- sert de base pour la production de :
  - grammaires
  - dictionnaires
- sa composition :
  une sélection de textes selon un modèle des critères sociologiques, linguistiques, etc.
- structure interne :
  sous-corpus thématiques, origine, etc.

Corpus de spécialité

- textes issus d'un domaine de spécialité
  médecine, cogénération, télécommunications, etc.
- représente un langage de spécialité (sous-langage) pour un domaine donné
- a des particularités lexicales et grammaticales
- sert de base pour
  - beaucoup d'application de fouille de texte
  - la production de terminologies (thesaurus, etc.)
- structure interne :
  sous-corpus thématiques, origine, etc.

Démarches dans la constitution du corpus

- Regroupement de textes opportuniste
- Regroupement de textes raisonné et ciblé :
  - représentativité :
    - taille suffisante
    - données diversifiées
      genre, spécialisation
  - origine mémorisée

Exigences vis-à-vis d'un corpus

Le corpus doit satisfaire une triple exigence :

- pertinence par rapport au domaine
  textes représentatifs de ceux produits dans le domaine
- pertinence par rapport à la tâche
  textes représentatifs de ceux manipulés par l'application finale
- prise en compte des possibilités des outils de traitement automatique

Pertinence par rapport au domaine

- textes qui véhiculent la connaissance du domaine :
  - textes spécialisés
  - textes de vulgarisation

Quelques tactiques de constitution de corpus

- scannérisation
- collecte auprès des experts
- documentation d'une entreprise, d'un service
- exploration d'une base de données spécialisée
- exploration de l'Internet

Pertinence par rapport au domaine

- Caractérisation du domaine :
  mots clés, termes centraux du domaine, descripteurs
- Garantie de la centralité des documents :
  présence des mots clés dans ces documents
- Recensement de textes qui véhiculent la connaissance du domaine :
  - textes spécialisés
  - textes pour les non-spécialistes
  - textes de vulgarisation
    (à inclure suivant la tâche)

Pertinence par rapport à la tâche

Prendre en compte l'application finale :

- objectifs
  un phénomène linguistique, grammaire, style littéraire
- application
  recherche d'information, indexation, etc.
- portée
  interne, externe à une entreprise, nationale, etc.
- spécialisation
  corpus de langue générale, de langue de spécialité
- type de corpus
  contexte multilingue : corpus parallèle

Pertinence par rapport aux outils

- Prendre en compte les possibilités des outils disponibles
- Choisir des outils en fonction de textes à traiter :
  - robustesse
  - langue
  - format
- Outils monolingues ou bien multilingues

Exemple de corpus spécialisés

Corpus Menelas :

- genres :
  - manuels
  - comptes rendus d'examens et de traitements
  - lettres aux collègues
- domaine :
  maladies coronariennes

Exemple de corpus spécialisés

Corpus Clef médical :

- genres :
  - compte-rendus d'hospitalisation
  - RMO
  - portail médical (CISMeF)
- domaines :
  - stomatologie
  - néphrologie
  - neurologie
  - etc.

Exemple de corpus spécialisés

Corpus Safir (co-génération d'électricité et de chaleur) :

- documents de l'Internet :
  - textes reglementaires de la communauté européenne
  - documents de présentation des sociétés
  - documents techniques de vulgarisation
- documents des conférences sur la cogénération :
  - présentation des sociétés et de leurs produits
  - documents écologiques
- documents des bases de données internes :
  - fiches techniques
  - présentation des sous-domaines, des branches

Corpus I2B2

(I2B2 – Informatics for Integrating Biology and the Bedside)

2009 : Extraction d'information dans les dossiers patient (1 249 documents, 696 pour l'entrainement – 17 annotés manuellement, et 553 pour la phase de test)
Extraction des

- noms des médicaments administrés aux patients (dans le passé ou actuellement)
- informations liées (dosage, durée, fréquence, mode d'administration, raison(s) de la prescription)

Autres types de corpus :

corpus comparables constituent des sélections de textes similaires (langues ou variétés d'une langue)

corpus parallèles sont constitués de documents traduits dans une ou plusieurs langues

corpus alignés les passages correspondants sont réliés (Hansard)

corpus de suivi corpus en évolution, corpus glissant

corpus segmentés segmentation en mots, phrases, etc.

corpus étiquetés étiquetage morpho-syntaxique

corpus arborés analyse syntaxique

etc.

Problèmes juridiques

- Information confidentielle
  - Menelas (CRH : maladies coronariennes)
    secret médical
  - Safir (documents divers : cogénération)
    données confidentielles d'une entreprise (EDF)
- Propriété intellectuelle
  - droits d'auteur
  - droits d'annotateurs de corpus

Problèmes juridiques

Solutions :

- secret médical (Informatique et Liberté) anonymisation ou dé-identification :
  - nom du patient
  - service
  - date et lieu de naissance
  - coordonnées du patient
  - coordonnées du service
  - date de consultation, d'hospitalisation
  - nom du médecin

Problèmes juridiques

Solutions :

- propriété intellectuelle :
  - extraits
  - convention
    cession de droits, licence d'utilisation, etc.
- confidentialité vis-à-vis d'une entreprise :
  - convention, achat d'une licence
  - documents « non-sortables »
  - ???

Echantillonnage

- problèmes juridiques
- "équilibrer" en taille les textes
  - "représenter" une diversité maximale de situations de communication
  - ne pas sur-représenter des "lieux" de textes aux caractéristiques particulières
- problèmes :
  comportements hétérogènes des occurrences dans les documents

Documentation

Types d'informations

- contexte de production du texte
  auteur, date, taille, format, public visé, thème, objectif, etc.
- contexte de collecte de corpus
  date, responsables, taille, etc.

Enregistrement

- dans des tables externes
- encodage XML

Regroupement des documents dans un corpus

- Identification des sources de documents
- Source des documents :
  - base bibliographique
  - collection interne
  - Web
    - Utilisation d'un crawler, de requêtes Google
    - Aspiration d'un site Web

Aubin (Sophie) et Hamon (Thierry). –
Improving Term Extraction with Terminological Resources. *In : Advances in Natural Language Processing (5th International Conference on NLP, FinTAL 2006)*, éd. par Salakoski (Tapio), Ginter (Filip), Pyysalo (Sampo) et Pahikkala (Tapio). pp. 380–387. –
Springer.

Gollub (Koraljka), Hamon (Thierry) et Ardö (Anders). –
Automated classification of textual documents based on a controlled vocabulary in engineering. *Knowledge Organization*, vol. 34 (4), 2007, pp. 247–263.

Grabar (Natalia), Jaulent (Marie-Christine) et Hamon (Thierry). –
Combination of endogenous clues for profiling inferred semantic relations: experiments with Gene Ontology. *In : Proceedings of the AMIA 2008 Annual Symposium*, pp. 252–256. –
Washington, DC, November 2008. PMID 18999042.

Grabar (Natalia), Varoutas (Paul-Christophe), Rizand (Philippe), Livartowski (Alain) et Hamon (Thierry). –
Automatic acquisition of Synonym Ressources and Assessment of their Impact on the Enhanced Search in EHRs. *Methods of Information in Medicine*, vol. 48 (2), 2009, pp. 149–154. –
PMID 19283312, DOI 10.3414/ME9213.

Grabar (Natalia) et Hamon (Thierry). –
Exploitation of speculation markers to identify the structure of biomedical scientific writing. *In : Proceedings of AMIA 2009 Symposium*, pp. 203–207. –
San Francisco, USA, November 2009.

Hamon (Thierry), Nazarenko (Adeline), Poibeau (Thierry), Aubin (Sophie) et Derivière (Julien). –
A Robust Linguistic Platform for Efficient and Domain specific Web Content Analysis. *In : Proceedings of RIAO 2007*. –
Pittsburgh, USA, 2007. 15 pages.

Hamon (Thierry), Graña (Martin), Raggio (Víctor), Grabar (Natalia) et Naya (Hugo). –

Identification of relations between risk factors and their pathologies or health conditions by mining scientific literature. *In :* *Proceedings of MEDINFO 2010*, pp. 964–968. –
PMID: 20841827.

Hamon (Thierry), Engström (Christopher) et Silvestrov (Sergei). –
Term ranking adaptation to the domain: genetic algorithm based optimisation of the C-Value. *In :* *Proceedings of PolTAL 2014 – Advances in Natural Language Processing*, éd. par Springer , pp. 71–83.

Hamon (Thierry) et Gagnayre (RÃ©mi). –
Improving knowledge of patient skills thanks to automatic analysis of online discussions. *Patient Education and Counseling*, 2013. –
Special section on Health Communication and Artificial Intelligence (IF: 2.305, 5yIF: 2.929).

Hamon (Thierry) et Grabar (Natalia). –
Exploring Graph Structure for Detection of Reliability Zones within Synonym Resources: Experiment with the Gene Ontology. *In :* *Proceedings of BioNLP'2009 – Workshop of NAACL-HLT 2009*. –
Boulder, Colorado, June 2009.

Hamon (Thierry) et Grabar (Natalia). –
Adaptation of Cross-Lingual Transfer Methods for the Building of Medical Terminology in Ukrainian. *In :* *Proceedings of the 17th International Conference on Intelligent Text Processing and Computational Linguistics (CICLING2016)*. –
Springer.

Hamon (Thierry). –
Rapport de Stage à l'UQAM - 19 juin - 21 juillet 2000, octobre 2000.

Hamon (Thierry). –
Acquisition terminologique pour identifier les mots clés d'articles scientifiques. *In :* *Actes de l'atelier DEFT 2012*, pp. 25–31. –
Grenoble, France, Juin 2012.

Hamon (Thierry). –
Indexation automatique de notices bibliographiques à l'aide d'approches d'acquisition terminologique. *In : Actes de DEFT 2016*, pp. 20–26. –
Paris, France, Juillet 2016. Atelier de la conférence JEP-TALN-RECITAL 2016.

Neifar (Wafa), Hamon (Thierry), Zweigenbaum (Pierre), Khemakhem (Mariem Ellouze) et Belguith (Lamia Hadrich). –
Adaptation of a Term Extractor to Arabic Specialised Texts: First Experiments and Limits. *In : Proceedings of the 17th International Conference on Intelligent Text Processing and Computational Linguistics (CICLING2016)*, éd. par Springer .

Périnet (Amandine), Grabar (Natalia) et Hamon (Thierry). –
Identification des assertions dans les textes médicaux : application à la relation {patient, problème médical}. *Traitement Automatique des Langues (TAL)*, vol. 52 (1), 2011, pp. 97–132.