

Soutenance de stage master SAD

Correction orthographique en contexte

Guillaume Pinot

Laboratoire d'Informatique de Nantes-Atlantique

5 septembre 2005

*Soutenance de stage
master SAD*

Guillaume Pinot

Introduction

État de l'art

*Méthode du contexte
non-ordonné*

*Approche intuitive de
l'algorithme*

Apprentissage

*Détection et correction
d'erreurs*

Expérimentations

Méthode expérimentale

Ajust des erreurs

Résultats et interprétation

Conclusion

1 Introduction

2 État de l'art

3 Méthode du contexte non-ordonné

- Approche intuitive de l'algorithme
- Apprentissage
- Détection et correction d'erreurs

4 Expérimentations

- Méthode expérimentale
- Ajout des erreurs
- Résultats et interprétation

5 Conclusion

*Soutenance de stage
master SAD*

Guillaume Pinot

Introduction

État de l'art

*Méthode du contexte
non-ordonné*

*Approche intuitive de
l'algorithme*

Apprentissage

*Détection et correction
d'erreurs*

Expérimentations

Méthode expérimentale

Ajout des erreurs

Résultats et interprétation

Conclusion

1 Introduction

2 État de l'art

- ## 3 Méthode du contexte non-ordonné
- Approche intuitive de l'algorithme
 - Apprentissage
 - Détection et correction d'erreurs

- ## 4 Expérimentations
- Méthode expérimentale
 - Ajout des erreurs
 - Résultats et interprétation

5 Conclusion

*Soutenance de stage
master SAD*

Guillaume Pinot

Introduction

État de l'art

*Méthode du contexte
non-ordonné*

*Approche intuitive de
l'algorithme*

Apprentissage

*Détection et correction
d'erreurs*

Expérimentations

Méthode expérimentale

Ajout des erreurs

Résultats et interprétation

Conclusion

*Soutenance de stage
master SAD*

Guillaume Pinot

Introduction

État de l'art

*Méthode du contexte
non-ordonné*

*Approche intuitive de
l'algorithme*

Apprentissage

*Détection et correction
d'erreurs*

Expérimentations

Méthode expérimentale

Ajout des erreurs

Résultats et interprétation

Conclusion

But : Corriger les erreurs cachées.

Les erreurs cachées surviennent lorsqu'une ou plusieurs modifications d'un mot le transforment en un autre mot du dictionnaire.

Exemple : pour → pur

Contrainte : Apprentissage sur corpus brut



But : Corriger les erreurs cachées.

Les erreurs cachées surviennent lorsqu'une ou plusieurs modifications d'un mot le transforment en un autre mot du dictionnaire.

Exemple : pour → pur

Contrainte : Apprentissage sur corpus brut



But : Corriger les erreurs cachées.

Les erreurs cachées surviennent lorsqu'une ou plusieurs modifications d'un mot le transforment en un autre mot du dictionnaire.

Exemple : pour → pur

Contrainte : Apprentissage sur corpus brut



But : Corriger les erreurs cachées.

Les erreurs cachées surviennent lorsqu'une ou plusieurs modifications d'un mot le transforment en un autre mot du dictionnaire.

Exemple : pour → pur

Contrainte : Apprentissage sur corpus brut



- 1 Introduction
- 2 **État de l'art**
- 3 Méthode du contexte non-ordonné
 - Approche intuitive de l'algorithme
 - Apprentissage
 - Détection et correction d'erreurs
- 4 Expérimentations
 - Méthode expérimentale
 - Ajout des erreurs
 - Résultats et interprétation
- 5 Conclusion

*Soutenance de stage
master SAD*

Guillaume Pinot

Introduction

État de l'art

*Méthode du contexte
non-ordonné*

*Approche intuitive de
l'algorithme*

Apprentissage

*Détection et correction
d'erreurs*

Expérimentations

Méthode expérimentale

Ajout des erreurs

Résultats et interprétation

Conclusion

Un ensemble de confusions est un ensemble de mots pouvant être confondus entre eux, en raison de leur orthographe proche (`{dessert, desert}`) ou parce qu'ils sont souvent confondus dans leur usage (`{between, among}`).

Exemples :

`{being, begin}`
`{past, passed}`
`{there, their, they're}`

*Soutenance de stage
master SAD*

Guillaume Pinot

Introduction

État de l'art

*Méthode du contexte
non-ordonné*

*Approche intuitive de
l'algorithme*

Apprentissage

*Détection et correction
d'erreurs*

Expérimentations

Méthode expérimentale

Ajout des erreurs

Résultats et interprétation

Conclusion



Un ensemble de confusions est un ensemble de mots pouvant être confondus entre eux, en raison de leur orthographe proche ({dessert, desert}) ou parce qu'ils sont souvent confondus dans leur usage ({between, among}).

Exemples :

{being, begin}
{past, passed}
{there, their, they're}

*Soutenance de stage
master SAD*

Guillaume Pinot

Introduction

État de l'art

*Méthode du contexte
non-ordonné*

*Approche intuitive de
l'algorithme*

Apprentissage

*Détection et correction
d'erreurs*

Expérimentations

Méthode expérimentale

Ajout des erreurs

Résultats et interprétation

Conclusion



Un ensemble de confusions est un ensemble de mots pouvant être confondus entre eux, en raison de leur orthographe proche ({dessert, desert}) ou parce qu'ils sont souvent confondus dans leur usage ({between, among}).

Exemples :

{being, begin}
{past, passed}
{there, their, they're}

*Soutenance de stage
master SAD*

Guillaume Pinot

Introduction

État de l'art

*Méthode du contexte
non-ordonné*

*Approche intuitive de
l'algorithme*

Apprentissage

*Détection et correction
d'erreurs*

Expérimentations

Méthode expérimentale

Ajout des erreurs

Résultats et interprétation

Conclusion



Un ensemble de confusions est un ensemble de mots pouvant être confondus entre eux, en raison de leur orthographe proche ({dessert, desert}) ou parce qu'ils sont souvent confondus dans leur usage ({between, among}).

Exemples :

{being, begin}
{past, passed}
{there, their, they're}

*Soutenance de stage
master SAD*

Guillaume Pinot

Introduction

État de l'art

*Méthode du contexte
non-ordonné*

*Approche intuitive de
l'algorithme*

Apprentissage

*Détection et correction
d'erreurs*

Expérimentations

Méthode expérimentale

Ajout des erreurs

Résultats et interprétation

Conclusion



Un ensemble de confusions est un ensemble de mots pouvant être confondus entre eux, en raison de leur orthographe proche ({dessert, desert}) ou parce qu'ils sont souvent confondus dans leur usage ({between, among}).

Exemples :

{being, begin}
{past, passed}
{there, their, they're}

*Soutenance de stage
master SAD*

Guillaume Pinot

Introduction

État de l'art

*Méthode du contexte
non-ordonné*

*Approche intuitive de
l'algorithme*

Apprentissage

*Détection et correction
d'erreurs*

Expérimentations

Méthode expérimentale

Ajout des erreurs

Résultats et interprétation

Conclusion



A Bayesian hybrid method for context-sensitive spelling correction [Golding, 1995]

Méthode *context words* :

- Apprentissage sur un corpus brut.
- Apprentissage des probabilités contextuelles (mot non-ordonnés).
- Correction : Calculer les probabilités de chaque mot de l'ensemble de confusions, puis proposer l'élément le plus probable.



A Bayesian hybrid method for context-sensitive spelling correction [Golding, 1995]

Méthode *context words* :

- Apprentissage sur un corpus brut.
- Apprentissage des probabilités contextuelles (mot non-ordonnés).
- Correction : Calculer les probabilités de chaque mot de l'ensemble de confusions, puis proposer l'élément le plus probable.



A Bayesian hybrid method for context-sensitive spelling correction [Golding, 1995]

Méthode *context words* :

- Apprentissage sur un corpus brut.
- Apprentissage des probabilités contextuelles (mots non-ordonnés).
- Correction : Calculer les probabilités de chaque mot de l'ensemble de confusions, puis proposer l'élément le plus probable.



A Bayesian hybrid method for context-sensitive spelling correction [Golding, 1995]

Méthode *context words* :

- Apprentissage sur un corpus brut.
- Apprentissage des probabilités contextuelles (mots non-ordonnés).
- Correction : Calculer les probabilités de chaque mot de l'ensemble de confusions, puis proposer l'élément le plus probable.



*Soutenance de stage
master SAD*

Guillaume Pinot

Introduction

État de l'art

*Méthode du contexte
non-ordonné*

*Approche intuitive de
l'algorithme*

Apprentissage

*Détection et correction
d'erreurs*

Expérimentations

Méthode expérimentale

Ajout des erreurs

Résultats et interprétation

Conclusion

Méthode *collocation* :

- Utilisation d'un étiqueteur.
- Apprentissage de 3-grammes.
- Correction : Calculer les probabilités de chaque mot de l'ensemble de confusions, puis proposer l'élément le plus probable.

Méthode *collocation* :

- Utilisation d'un étiqueteur.
- Apprentissage de 3-grammes.
- Correction : Calculer les probabilités de chaque mot de l'ensemble de confusions, puis proposer l'élément le plus probable.

*Soutenance de stage
master SAD*

Guillaume Pinot

Introduction

État de l'art

*Méthode du contexte
non-ordonné*

*Approche intuitive de
l'algorithme*

Apprentissage

*Détection et correction
d'erreurs*

Expérimentations

Méthode expérimentale

Ajout des erreurs

Résultats et interprétation

Conclusion

*Soutenance de stage
master SAD*

Guillaume Pinot

Introduction

État de l'art

*Méthode du contexte
non-ordonné*

*Approche intuitive de
l'algorithme*

Apprentissage

*Détection et correction
d'erreurs*

Expérimentations

Méthode expérimentale

Ajout des erreurs

Résultats et interprétation

Conclusion

Méthode *collocation* :

- Utilisation d'un étiqueteur.
- Apprentissage de 3-grammes.
- Correction : Calculer les probabilités de chaque mot de l'ensemble de confusions, puis proposer l'élément le plus probable.



*Soutenance de stage
master SAD*

Guillaume Pinot

Introduction

État de l'art

*Méthode du contexte
non-ordonné*

*Approche intuitive de
l'algorithme*

Apprentissage

*Détection et correction
d'erreurs*

Expérimentations

Méthode expérimentale

Ajout des erreurs

Résultats et interprétation

Conclusion

Méthode *collocation* :

- Utilisation d'un étiqueteur.
- Apprentissage de 3-grammes.
- Correction : Calculer les probabilités de chaque mot de l'ensemble de confusions, puis proposer l'élément le plus probable.

Premier article de Golding : hybridations

Hybridation en liste de décision :

- Apprentissage : ordonner les données de la plus pertinente à la moins pertinente.
- Correction : utiliser la première donnée concordante pour calculer la probabilité.

Hybridation avec un classificateur Bayesiens :

- Apprentissage : ordonner les données de la plus pertinente à la moins pertinente.
- Correction : utiliser toutes les données concordantes pour calculer la probabilité.

*Soutenance de stage
master SAD*

Guillaume Pinot

Introduction

État de l'art

*Méthode du contexte
non-ordonné*

*Approche intuitive de
l'algorithme*

Apprentissage

*Détection et correction
d'erreurs*

Expérimentations

Méthode expérimentale

Ajust des erreurs

Résultats et interprétation

Conclusion

Premier article de Golding : hybridations

*Soutenance de stage
master SAD*

Guillaume Pinot

Introduction

État de l'art

*Méthode du contexte
non-ordonné*

*Approche intuitive de
l'algorithme*

Apprentissage

*Détection et correction
d'erreurs*

Expérimentations

Méthode expérimentale

Ajout des erreurs

Résultats et interprétation

Conclusion

Hybridation en liste de décision :

- Apprentissage : ordonner les données de la plus pertinente à la moins pertinente.
- Correction : utiliser la première donnée concordante pour calculer la probabilité.

Hybridation avec un classificateur Bayesiens :

- Apprentissage : ordonner les données de la plus pertinente à la moins pertinente.
- Correction : utiliser toutes les données concordantes pour calculer la probabilité.

Premier article de Golding : hybridations

*Soutenance de stage
master SAD*

Guillaume Pinot

Hybridation en liste de décision :

- Apprentissage : ordonner les données de la plus pertinente à la moins pertinente.
- Correction : utiliser la première donnée concordante pour calculer la probabilité.

Hybridation avec un classificateur Bayesiens :

- Apprentissage : ordonner les données de la plus pertinente à la moins pertinente.
- Correction : utiliser toutes les données concordantes pour calculer la probabilité.

Introduction

État de l'art

*Méthode du contexte
non-ordonné*

*Approche intuitive de
l'algorithme*

Apprentissage

*Détection et correction
d'erreurs*

Expérimentations

Méthode expérimentale

Ajust des erreurs

Résultats et interprétation

Conclusion

Premier article de Golding : hybridations

Hybridation en liste de décision :

- Apprentissage : ordonner les données de la plus pertinente à la moins pertinente.
- Correction : utiliser la première donnée concordante pour calculer la probabilité.

Hybridation avec un classificateur Bayesiens :

- Apprentissage : ordonner les données de la plus pertinente à la moins pertinente.
- Correction : utiliser toutes les données concordantes pour calculer la probabilité.

*Soutenance de stage
master SAD*

Guillaume Pinot

Introduction

État de l'art

*Méthode du contexte
non-ordonné*

*Approche intuitive de
l'algorithme*

Apprentissage

*Détection et correction
d'erreurs*

Expérimentations

Méthode expérimentale

Ajust des erreurs

Résultats et interprétation

Conclusion

Premier article de Golding : hybridations

*Soutenance de stage
master SAD*

Guillaume Pinot

Introduction

État de l'art

*Méthode du contexte
non-ordonné*

*Approche intuitive de
l'algorithme*

Apprentissage

*Détection et correction
d'erreurs*

Expérimentations

Méthode expérimentale

Ajust des erreurs

Résultats et interprétation

Conclusion

Hybridation en liste de décision :

- Apprentissage : ordonner les données de la plus pertinente à la moins pertinente.
- Correction : utiliser la première donnée concordante pour calculer la probabilité.

Hybridation avec un classificateur Bayesiens :

- Apprentissage : ordonner les données de la plus pertinente à la moins pertinente.
- Correction : utiliser toutes les données concordantes pour calculer la probabilité.



Premier article de Golding : hybridations

Hybridation en liste de décision :

- Apprentissage : ordonner les données de la plus pertinente à la moins pertinente.
- Correction : utiliser la première donnée concordante pour calculer la probabilité.

Hybridation avec un classificateur Bayesiens :

- Apprentissage : ordonner les données de la plus pertinente à la moins pertinente.
- Correction : utiliser toutes les données concordantes pour calculer la probabilité.

*Soutenance de stage
master SAD*

Guillaume Pinot

Introduction

État de l'art

*Méthode du contexte
non-ordonné*

*Approche intuitive de
l'algorithme*

Apprentissage

*Détection et correction
d'erreurs*

Expérimentations

Méthode expérimentale

Ajust des erreurs

Résultats et interprétation

Conclusion



Publications fondées sur les ensembles de confusions :

- *Combining Trigram-based and Feature-based Methods for Context-Sensitive Spelling Correction* [Golding and Schabès, 1996]
- *A Winnow-Based Approach to Context-Sensitive Spelling Correction* [Golding and Roth, 1999]

Publications non fondées sur les ensembles de confusions :

- *Correcting Real-Word Spelling Errors by Restoring Lexical Cohesion* [Hirst and Budanitsky, 2005]
- *Un système multi-agent pour la détection et la correction des erreurs cachées en langue écrite* [Zribi et al., 2005]

*Soutenance de stage
master SAD*

Guillaume Pinot

Introduction

État de l'art

*Méthode du contexte
non-ordonné*

*Approche intuitive de
l'algorithme*

Apprentissage

*Détection et correction
d'erreurs*

Expérimentations

Méthode expérimentale

Ajout des erreurs

Résultats et interprétation

Conclusion

Publications fondées sur les ensembles de confusions :

- *Combining Trigram-based and Feature-based Methods for Context-Sensitive Spelling Correction* [Golding and Schabes, 1996]
- *A Winnow-Based Approach to Context-Sensitive Spelling Correction* [Golding and Roth, 1999]

Publications non fondées sur les ensembles de confusions :

- *Correcting Real-Word Spelling Errors by Restoring Lexical Cohesion* [Hirst and Budanitsky, 2005]
- *Un système multi-agent pour la détection et la correction des erreurs cachées en langue écrite* [Zribi et al., 2005]

*Soutenance de stage
master SAD*

Guillaume Pinot

Introduction

État de l'art

*Méthode du contexte
non-ordonné*

*Approche intuitive de
l'algorithme*

Apprentissage

*Détection et correction
d'erreurs*

Expérimentations

Méthode expérimentale

Ajout des erreurs

Résultats et interprétation

Conclusion

Publications fondées sur les ensembles de confusions :

- *Combining Trigram-based and Feature-based Methods for Context-Sensitive Spelling Correction* [Golding and Schabes, 1996]
- *A Winnow-Based Approach to Context-Sensitive Spelling Correction* [Golding and Roth, 1999]

Publications non fondées sur les ensembles de confusions :

- *Correcting Real-Word Spelling Errors by Restoring Lexical Cohesion* [Hirst and Budanitsky, 2005]
- *Un système multi-agent pour la détection et la correction des erreurs cachées en langage oral* [Zribi et al., 2005]

*Soutenance de stage
master SAD*

Guillaume Pinot

Introduction

État de l'art

*Méthode du contexte
non-ordonné*

*Approche intuitive de
l'algorithme*

Apprentissage

*Détection et correction
d'erreurs*

Expérimentations

Méthode expérimentale

Ajout des erreurs

Résultats et interprétation

Conclusion



Publications fondées sur les ensembles de confusions :

- *Combining Trigram-based and Feature-based Methods for Context-Sensitive Spelling Correction* [Golding and Schabes, 1996]
- *A Winnow-Based Approach to Context-Sensitive Spelling Correction* [Golding and Roth, 1999]

Publications non fondées sur les ensembles de confusions :

- *Correcting Real-Word Spelling Errors by Restoring Lexical Cohesion* [Hirst and Budanitsky, 2005]
- *Un système multi-agent pour la détection et la correction des erreurs cachées en langage oral* [Zribi et al., 2005]

*Soutenance de stage
master SAD*

Guillaume Pinot

Introduction

État de l'art

*Méthode du contexte
non-ordonné*

*Approche intuitive de
l'algorithme*

Apprentissage

*Détection et correction
d'erreurs*

Expérimentations

Méthode expérimentale

Ajout des erreurs

Résultats et interprétation

Conclusion



Publications fondées sur les ensembles de confusions :

- *Combining Trigram-based and Feature-based Methods for Context-Sensitive Spelling Correction* [Golding and Schabes, 1996]
- *A Winnow-Based Approach to Context-Sensitive Spelling Correction* [Golding and Roth, 1999]

Publications non fondées sur les ensembles de confusions :

- *Correcting Real-Word Spelling Errors by Restoring Lexical Cohesion* [Hirst and Budanitsky, 2005]
- *Un système multi-agent pour la détection et la correction des erreurs cachées en langage oral* [Zribi et al., 2005]

Soutenance de stage
master SAD

Guillaume Pinot

Introduction

État de l'art

Méthode du contexte
non-ordonné

Approche intuitive de
l'algorithme

Apprentissage

Détection et correction
d'erreurs

Expérimentations

Méthode expérimentale

Ajout des erreurs

Résultats et interprétation

Conclusion



Publications fondées sur les ensembles de confusions :

- *Combining Trigram-based and Feature-based Methods for Context-Sensitive Spelling Correction* [Golding and Schabes, 1996]
- *A Winnow-Based Approach to Context-Sensitive Spelling Correction* [Golding and Roth, 1999]

Publications non fondées sur les ensembles de confusions :

- *Correcting Real-Word Spelling Errors by Restoring Lexical Cohesion* [Hirst and Budanitsky, 2005]
- *Un système multi-agent pour la détection et la correction des erreurs cachées en langue arabe* [Zribi et al., 2005]

*Soutenance de stage
master SAD*

Guillaume Pinot

Introduction

État de l'art

*Méthode du contexte
non-ordonné*

*Approche intuitive de
l'algorithme*

Apprentissage

*Détection et correction
d'erreurs*

Expérimentations

Méthode expérimentale

Ajout des erreurs

Résultats et interprétation

Conclusion

- Beaucoup de publications utilisent des ressources linguistiques évoluées.
- Recherche encore active.
- Non-utilisation de tels correcteurs dans les applications grand public.

*Soutenance de stage
master SAD*

Guillaume Pinot

Introduction

État de l'art

*Méthode du contexte
non-ordonné*

*Approche intuitive de
l'algorithme*

Apprentissage

*Détection et correction
d'erreurs*

Expérimentations

Méthode expérimentale

Ajout des erreurs

Résultats et interprétation

Conclusion

- Beaucoup de publications utilisent des ressources linguistiques évoluées.
- Recherche encore active.
- Non-utilisation de tels correcteurs dans les applications grand public.

*Soutenance de stage
master SAD*

Guillaume Pinot

Introduction

État de l'art

*Méthode du contexte
non-ordonné*

*Approche intuitive de
l'algorithme*

Apprentissage

*Détection et correction
d'erreurs*

Expérimentations

Méthode expérimentale

Ajout des erreurs

Résultats et interprétation

Conclusion

- Beaucoup de publications utilisent des ressources linguistiques évoluées.
- Recherche encore active.
- Non-utilisation de tels correcteurs dans les applications grand public.

*Soutenance de stage
master SAD*

Guillaume Pinot

Introduction

État de l'art

*Méthode du contexte
non-ordonné*

*Approche intuitive de
l'algorithme*

Apprentissage

*Détection et correction
d'erreurs*

Expérimentations

Méthode expérimentale

Ajout des erreurs

Résultats et interprétation

Conclusion

1 Introduction

2 État de l'art

3 **Méthode du contexte non-ordonné**

- Approche intuitive de l'algorithme
- Apprentissage
- Détection et correction d'erreurs

4 Expérimentations

- Méthode expérimentale
- Ajout des erreurs
- Résultats et interprétation

5 Conclusion

*Soutenance de stage
master SAD*

Guillaume Pinot

Introduction

État de l'art

*Méthode du contexte
non-ordonné*

*Approche intuitive de
l'algorithme*

Apprentissage

*Détection et correction
d'erreurs*

Expérimentations

Méthode expérimentale

Ajout des erreurs

Résultats et interprétation

Conclusion



1 Introduction

2 État de l'art

3 **Méthode du contexte non-ordonné**

- Approche intuitive de l'algorithme
- Apprentissage
- Détection et correction d'erreurs

4 Expérimentations

- Méthode expérimentale
- Ajout des erreurs
- Résultats et interprétation

5 Conclusion

*Soutenance de stage
master SAD*

Guillaume Pinot

Introduction

État de l'art

*Méthode du contexte
non-ordonné*

*Approche intuitive de
l'algorithme*

Apprentissage

*Détection et correction
d'erreurs*

Expérimentations

Méthode expérimentale

Ajout des erreurs

Résultats et interprétation

Conclusion



Le but de l'apprentissage est de mémoriser :

- Les contextes des mots ;
- Leurs probabilités associées.

La correction :

- Trouver les mots apparaissant dans le contexte courant.
- Garder les mots proches du mot à corriger.
- Les ordonner.
- Déterminer si le mot est une erreur ou non.

Contexte non-ordonnés \Rightarrow sémantique.

*Soutenance de stage
master SAD*

Guillaume Pinot

Introduction

État de l'art

*Méthode du contexte
non-ordonné*

*Approche intuitive de
l'algorithme*

Apprentissage

*Détection et correction
d'erreurs*

Expérimentations

Méthode expérimentale

Ajout des erreurs

Résultats et interprétation

Conclusion



Le but de l'apprentissage est de mémoriser :

- Les contextes des mots ;
- Leurs probabilités associées.

La correction :

- Trouver les mots apparaissant dans le contexte courant.
- Garder les mots proches du mot à corriger.
- Les ordonner.
- Déterminer si le mot est une erreur ou non.

Contexte non-ordonnés \Rightarrow sémantique.

*Soutenance de stage
master SAD*

Guillaume Pinot

Introduction

État de l'art

*Méthode du contexte
non-ordonné*

*Approche intuitive de
l'algorithme*

Apprentissage

*Détection et correction
d'erreurs*

Expérimentations

Méthode expérimentale

Ajout des erreurs

Résultats et interprétation

Conclusion

Le but de l'apprentissage est de mémoriser :

- Les contextes des mots ;
- Leurs probabilités associées.

La correction :

- Trouver les mots apparaissant dans le contexte courant.
- Garder les mots proches du mot à corriger.
- Les ordonner.
- Déterminer si le mot est une erreur ou non.

Contexte non-ordonnés \Rightarrow sémantique.

*Soutenance de stage
master SAD*

Guillaume Pinot

Introduction

État de l'art

*Méthode du contexte
non-ordonné*

*Approche intuitive de
l'algorithme*

Apprentissage

*Détection et correction
d'erreurs*

Expérimentations

Méthode expérimentale

Ajout des erreurs

Résultats et interprétation

Conclusion



Le but de l'apprentissage est de mémoriser :

- Les contextes des mots ;
- Leurs probabilités associées.

La correction :

- Trouver les mots apparaissant dans le contexte courant.
- Garder les mots proches du mot à corriger.
- Les ordonner.
- Déterminer si le mot est une erreur ou non.

Contexte non-ordonnés \Rightarrow sémantique.

*Soutenance de stage
master SAD*

Guillaume Pinot

Introduction

État de l'art

*Méthode du contexte
non-ordonné*

*Approche intuitive de
l'algorithme*

Apprentissage

*Détection et correction
d'erreurs*

Expérimentations

Méthode expérimentale

Ajout des erreurs

Résultats et interprétation

Conclusion



Le but de l'apprentissage est de mémoriser :

- Les contextes des mots ;
- Leurs probabilités associées.

La correction :

- Trouver les mots apparaissant dans le contexte courant.
- Garder les mots proches du mot à corriger.
- Les ordonner.
- Déterminer si le mot est une erreur ou non.

Contexte non-ordonnés \Rightarrow sémantique.

*Soutenance de stage
master SAD*

Guillaume Pinot

Introduction

État de l'art

*Méthode du contexte
non-ordonné*

*Approche intuitive de
l'algorithme*

Apprentissage

*Détection et correction
d'erreurs*

Expérimentations

Méthode expérimentale

Ajout des erreurs

Résultats et interprétation

Conclusion



Le but de l'apprentissage est de mémoriser :

- Les contextes des mots ;
- Leurs probabilités associées.

La correction :

- Trouver les mots apparaissant dans le contexte courant.
- Garder les mots proches du mot à corriger.
- Les ordonner.
- Déterminer si le mot est une erreur ou non.

Contexte non-ordonnés \Rightarrow sémantique.

*Soutenance de stage
master SAD*

Guillaume Pinot

Introduction

État de l'art

*Méthode du contexte
non-ordonné*

*Approche intuitive de
l'algorithme*

Apprentissage

*Détection et correction
d'erreurs*

Expérimentations

Méthode expérimentale

Ajout des erreurs

Résultats et interprétation

Conclusion



Le but de l'apprentissage est de mémoriser :

- Les contextes des mots ;
- Leurs probabilités associées.

La correction :

- Trouver les mots apparaissant dans le contexte courant.
- Garder les mots proches du mot à corriger.
- Les ordonner.
- Déterminer si le mot est une erreur ou non.

Contexte non-ordonnés \Rightarrow sémantique.



Le but de l'apprentissage est de mémoriser :

- Les contextes des mots ;
- Leurs probabilités associées.

La correction :

- Trouver les mots apparaissant dans le contexte courant.
- Garder les mots proches du mot à corriger.
- Les ordonner.
- Déterminer si le mot est une erreur ou non.

Contexte non-ordonnés \Rightarrow sémantique.



Le but de l'apprentissage est de mémoriser :

- Les contextes des mots ;
- Leurs probabilités associées.

La correction :

- Trouver les mots apparaissant dans le contexte courant.
- Garder les mots proches du mot à corriger.
- Les ordonner.
- Déterminer si le mot est une erreur ou non.

Contexte non-ordonnés \Rightarrow sémantique.



- 1 Introduction
- 2 État de l'art
- 3 Méthode du contexte non-ordonné**
 - Approche intuitive de l'algorithme
 - Apprentissage
 - Détection et correction d'erreurs
- 4 Expérimentations
 - Méthode expérimentale
 - Ajout des erreurs
 - Résultats et interprétation
- 5 Conclusion

*Soutenance de stage
master SAD*

Guillaume Pinot

Introduction

État de l'art

*Méthode du contexte
non-ordonné*

*Approche intuitive de
l'algorithme*

Apprentissage

*Détection et correction
d'erreurs*

Expérimentations

Méthode expérimentale

Ajout des erreurs

Résultats et interprétation

Conclusion



Soit $k = 3$ et $c = 6$

On peut

s'	assurer	un	bon	revenu.
w_3	w_4	w_5	w_6	w_7

w_3 , w_4 , et w_5 sont dans le contexte de w_6 , le mot courant. Les cooccurrences $\{w_3, w_6\}$, $\{w_4, w_6\}$ et $\{w_5, w_6\}$ sont relevées.



Formule des probabilités :

$$P(w_i|w_j) = \frac{c_{i,j}}{c_j}$$

Exemple des données générées durant l'apprentissage

w_i	une	sorte	un	lui	...
w_j	sorte	une	lui	un	...
$P(w_i w_j)$	0.6982	0.0365	0.0750	0.0267	...

Introduction

État de l'art

*Méthode du contexte
non-ordonné*

*Approche intuitive de
l'algorithme*

Apprentissage

*Détection et correction
d'erreurs*

Expérimentations

Méthode expérimentale

Ajout des erreurs

Résultats et interprétation

Conclusion

Formule des probabilités :

$$P(w_i|w_j) = \frac{c_{i,j}}{c_j}$$

Exemple des données générées durant l'apprentissage :

w_i	une	sorte	un	lui	...
w_j	sorte	une	lui	un	...
$P(w_i w_j)$	0.6982	0.0365	0.0750	0.0267	...

1 Introduction

2 État de l'art

3 **Méthode du contexte non-ordonné**

- Approche intuitive de l'algorithme
- Apprentissage
- Détection et correction d'erreurs

4 Expérimentations

- Méthode expérimentale
- Ajout des erreurs
- Résultats et interprétation

5 Conclusion

*Soutenance de stage
master SAD*

Guillaume Pinot

Introduction

État de l'art

*Méthode du contexte
non-ordonné*

*Approche intuitive de
l'algorithme*

Apprentissage

*Détection et correction
d'erreurs*

Expérimentations

Méthode expérimentale

Ajout des erreurs

Résultats et interprétation

Conclusion



Compter le nombre minimal d'insertion, de suppression, de substitution et d'inversion de lettres pour transformer un mot en un autre.

Exemple :

test $\xrightarrow{s \rightarrow x}$ text $\xrightarrow{+e}$ texte

ce qui donne

$$\text{edist}(\text{test}, \text{texte}) = 2$$

*Soutenance de stage
master SAD*

Guillaume Pinot

Introduction

État de l'art

*Méthode du contexte
non-ordonné*

*Approche intuitive de
l'algorithme*

Apprentissage

*Détection et correction
d'erreurs*

Expérimentations

Méthode expérimentale

Ajout des erreurs

Résultats et interprétation

Conclusion

Compter le nombre minimal d'insertion, de suppression, de substitution et d'inversion de lettres pour transformer un mot en un autre.

Exemple :

test $\xrightarrow{s \rightarrow x}$ text $\xrightarrow{+e}$ texte

ce qui donne

$\text{edist}(\text{test}, \text{texte}) = 2$



Compter le nombre minimal d'insertion, de suppression, de substitution et d'inversion de lettres pour transformer un mot en un autre.

Exemple :

test $\xrightarrow{s \rightarrow x}$ text $\xrightarrow{+e}$ texte

ce qui donne

$$\text{edist}(\text{test}, \text{texte}) = 2$$

*Soutenance de stage
master SAD*

Guillaume Pinot

Introduction

État de l'art

*Méthode du contexte
non-ordonné*

*Approche intuitive de
l'algorithme*

Apprentissage

*Détection et correction
d'erreurs*

Expérimentations

Méthode expérimentale

Ajout des erreurs

Résultats et interprétation

Conclusion



Soit ϵ la chaîne vide :

$$\text{sim}(w_i, w_j) = \begin{cases} \text{vrai si } \text{edist}(w_i, w_j) \leq \frac{\text{edist}(w_i, \epsilon) + \text{edist}(w_j, \epsilon)}{\gamma} + c \\ \text{faux sinon} \end{cases}$$

En pratique, $\gamma = 8$ et $c = 1$. Ces valeurs ont été déterminées après plusieurs expérimentations.

Exemple :

$$\frac{\text{edist}(\text{test}, \text{texte})}{\gamma} + c = \frac{4+5}{8} + 1 = 1.25$$

Donc $\text{sim}(\text{test}, \text{texte}) = \text{vrai}$

*Soutenance de stage
master SAD*

Guillaume Pinot

Introduction

État de l'art

*Méthode du contexte
non-ordonné*

*Approche intuitive de
l'algorithme*

Apprentissage

*Détection et correction
d'erreurs*

Expérimentations

Méthode expérimentale

Ajout des erreurs

Résultats et interprétation

Conclusion

Soit ϵ la chaîne vide :

$$\text{sim}(w_i, w_j) = \begin{cases} \text{vrai si } \text{edist}(w_i, w_j) \leq \frac{\text{edist}(w_i, \epsilon) + \text{edist}(w_j, \epsilon)}{\gamma} + c \\ \text{faux sinon} \end{cases}$$

En pratique, $\gamma = 8$ et $c = 1$. Ces valeurs ont été déterminées après plusieurs expérimentations.

Exemple :

$$\frac{\text{edist}(\text{test}, \text{texte})}{\gamma} + c = \frac{4+5}{8} + 1 = 2.125$$

Donc $\text{sim}(\text{test}, \text{texte}) = \text{vrai}$

*Soutenance de stage
master SAD*

Guillaume Pinot

Introduction

État de l'art

*Méthode du contexte
non-ordonné*

*Approche intuitive de
l'algorithme*

Apprentissage

*Détection et correction
d'erreurs*

Expérimentations

Méthode expérimentale

Ajout des erreurs

Résultats et interprétation

Conclusion

Soit $k = 3$ et $c = 5$

On

peut	s'	assurer	un	bon	revenu	en	travaillant.
w_2	w_3	w_4	w_5	w_6	w_7	w_8	w_9

Nous avons ici w_2, w_3, w_4, w_6, w_7 et w_8 dans le contexte du mot courant w_5 , donc

$$K_5 = \{w_2, w_3, w_4, w_6, w_7, w_8\}$$



Soient $k = 3$ et $w_c = \text{game}$

And	so	my	mind	game	round	to	the	business.
	0.030	0.012		came	0.073	0.016	0.0157	
				same			0.0052	
				gate		0.003	0.0030	
		0.009		gave	0.003		0.0017	
	0.006			name			0.0008	
				game			0.0004	

Pour ordonner les propositions, nous utilisons une heuristique :

$$\begin{aligned} H_{\text{came}} &= 5 + 0.030 \times 0.012 \times 0.073 \times 0.016 \times 0.015 \\ &= 5.000000006601 \end{aligned}$$

$$H_{\text{game}} = 1 + 0.0004 = 1.0004$$

Détection :

$$H1 : H_{\text{came}} > H_{\text{game}} \Rightarrow \text{erreur}$$

$$H2 : \lfloor H_{\text{came}} \rfloor > \lfloor H_{\text{game}} \rfloor \Rightarrow \text{erreur}$$

*Soutenance de stage
master SAD*

Guillaume Pinot

Introduction

État de l'art

*Méthode du contexte
non-ordonné*

*Approche intuitive de
l'algorithme*

Apprentissage

*Détection et correction
d'erreurs*

Expérimentations

Méthode expérimentale

Ajout des erreurs

Résultats et interprétation

Conclusion

Soient $k = 3$ et $w_c = \text{game}$

And	so	my	mind	game	round	to	the	business.
	0.030	0.012		came	0.073	0.016	0.0157	
				same			0.0052	
				gate		0.003	0.0030	
		0.009		gave	0.003		0.0017	
	0.006			name			0.0008	
				game			0.0004	

Pour ordonner les propositions, nous utilisons une heuristique :

$$\begin{aligned}H_{\text{came}} &= 5 + 0.030 \times 0.012 \times 0.073 \times 0.016 \times 0.0157 \\ &= 5.000000006601\end{aligned}$$

$$H_{\text{game}} = 1 + 0.0004 = 1.0004$$

Détection :

$$H1 : H_{\text{came}} > H_{\text{game}} \Rightarrow \text{erreur}$$

$$H2 : \lfloor H_{\text{came}} \rfloor > \lfloor H_{\text{game}} \rfloor \Rightarrow \text{erreur}$$

*Soutenance de stage
master SAD*

Guillaume Pinot

Introduction

État de l'art

*Méthode du contexte
non-ordonné*

*Approche intuitive de
l'algorithme*

Apprentissage

*Détection et correction
d'erreurs*

Expérimentations

Méthode expérimentale

Ajout des erreurs

Résultats et interprétation

Conclusion

Soient $k = 3$ et $w_c = \text{game}$

And	so	my	mind	game	round	to	the	business.
	0.030	0.012		came	0.073	0.016	0.0157	
				same			0.0052	
				gate		0.003	0.0030	
		0.009		gave		0.003	0.0017	
	0.006			name			0.0008	
				game			0.0004	

Pour ordonner les propositions, nous utilisons une heuristique :

$$\begin{aligned}H_{\text{came}} &= 5 + 0.030 \times 0.012 \times 0.073 \times 0.016 \times 0.0157 \\ &= 5.000000006601\end{aligned}$$

$$H_{\text{game}} = 1 + 0.0004 = 1.0004$$

Détection :

$$H1 : H_{\text{came}} > H_{\text{game}} \Rightarrow \text{erreur}$$

$$H2 : \lfloor H_{\text{came}} \rfloor > \lfloor H_{\text{game}} \rfloor \Rightarrow \text{erreur}$$

*Soutenance de stage
master SAD*

Guillaume Pinot

Introduction

État de l'art

*Méthode du contexte
non-ordonné*

*Approche intuitive de
l'algorithme*

Apprentissage

*Détection et correction
d'erreurs*

Expérimentations

Méthode expérimentale

Ajout des erreurs

Résultats et interprétation

Conclusion

Soient $k = 3$ et $w_c = \text{game}$

And	so	my	mind	game	round	to	the	business.
	0.030	0.012		came	0.073	0.016	0.0157	
				same			0.0052	
				gate		0.003	0.0030	
		0.009		gave		0.003	0.0017	
	0.006			name			0.0008	
				game			0.0004	

Pour ordonner les propositions, nous utilisons une heuristique :

$$\begin{aligned}H_{\text{came}} &= 5 + 0.030 \times 0.012 \times 0.073 \times 0.016 \times 0.0157 \\ &= 5.000000006601\end{aligned}$$

$$H_{\text{game}} = 1 + 0.0004 = 1.0004$$

Détection :

$$H1 : H_{\text{came}} > H_{\text{game}} \Rightarrow \text{erreur}$$

$$H2 : \lfloor H_{\text{came}} \rfloor > \lfloor H_{\text{game}} \rfloor \Rightarrow \text{erreur}$$

*Soutenance de stage
master SAD*

Guillaume Pinot

Introduction

État de l'art

*Méthode du contexte
non-ordonné*

*Approche intuitive de
l'algorithme*

Apprentissage

*Détection et correction
d'erreurs*

Expérimentations

Méthode expérimentale

Ajout des erreurs

Résultats et interprétation

Conclusion

- 1 Introduction
- 2 État de l'art
- 3 Méthode du contexte non-ordonné
 - Approche intuitive de l'algorithme
 - Apprentissage
 - Détection et correction d'erreurs
- 4 Expérimentations**
 - Méthode expérimentale
 - Ajout des erreurs
 - Résultats et interprétation
- 5 Conclusion

*Soutenance de stage
master SAD*

Guillaume Pinot

Introduction

État de l'art

*Méthode du contexte
non-ordonné*

*Approche intuitive de
l'algorithme*

Apprentissage

*Détection et correction
d'erreurs*

Expérimentations

Méthode expérimentale

Ajout des erreurs

Résultats et interprétation

Conclusion



- 1 Introduction
- 2 État de l'art
- 3 Méthode du contexte non-ordonné
 - Approche intuitive de l'algorithme
 - Apprentissage
 - Détection et correction d'erreurs
- 4 Expérimentations**
 - Méthode expérimentale
 - Ajout des erreurs
 - Résultats et interprétation
- 5 Conclusion

*Soutenance de stage
master SAD*

Guillaume Pinot

Introduction

État de l'art

*Méthode du contexte
non-ordonné*

*Approche intuitive de
l'algorithme*

Apprentissage

*Détection et correction
d'erreurs*

Expérimentations

Méthode expérimentale

Ajout des erreurs

Résultats et interprétation

Conclusion



Corpus : *Les Misérables* de Victor Hugo.

Partie	Nb. de mots	Parties des <i>Misérables</i>
0	55451	Tome 1 livres 1 à 3
1	59127	Tome 1 livres 4 à 8
2	53332	Tome 2 livres 1 à 4
3	52915	Tome 2 livres 5 à 8, Tome 3 livre 1
4	47882	Tome 3 livres 2 à 7
5	53405	Tome 3 livre 8, Tome 4 livres 1 et 2
6	59836	Tome 4 livres 3 à 8
7	44698	Tome 4 livres 9 à 15
8	54349	Tome 5 livres 1 à 3
9	52998	Tome 5 livres 4 à 9
Total	533993	<i>Les Misérables</i>

*Soutenance de stage
master SAD*

Guillaume Pinot

Introduction

État de l'art

*Méthode du contexte
non-ordonné*

*Approche intuitive de
l'algorithme*

Apprentissage

*Détection et correction
d'erreurs*

Expérimentations

Méthode expérimentale

Ajout des erreurs

Résultats et interprétation

Conclusion



Corpus : *Les Misérables* de Victor Hugo.

Partie	Nb. de mots	Parties des <i>Misérables</i>
0	55451	Tome 1 livres 1 à 3
1	59127	Tome 1 livres 4 à 8
2	53332	Tome 2 livres 1 à 4
3	52915	Tome 2 livres 5 à 8, Tome 3 livre 1
4	47882	Tome 3 livres 2 à 7
5	53405	Tome 3 livre 8, Tome 4 livres 1 et 2
6	59836	Tome 4 livres 3 à 8
7	44698	Tome 4 livres 9 à 15
8	54349	Tome 5 livres 1 à 3
9	52998	Tome 5 livres 4 à 9
Total	533993	<i>Les Misérables</i>

*Soutenance de stage
master SAD*

Guillaume Pinot

Introduction

État de l'art

*Méthode du contexte
non-ordonné*

*Approche intuitive de
l'algorithme*

Apprentissage

*Détection et correction
d'erreurs*

Expérimentations

Méthode expérimentale

Ajout des erreurs

Résultats et interprétation

Conclusion



- 1 Introduction
- 2 État de l'art
- 3 Méthode du contexte non-ordonné
 - Approche intuitive de l'algorithme
 - Apprentissage
 - Détection et correction d'erreurs
- 4 **Expérimentations**
 - Méthode expérimentale
 - Ajout des erreurs
 - Résultats et interprétation
- 5 Conclusion

*Soutenance de stage
master SAD*

Guillaume Pinot

Introduction

État de l'art

*Méthode du contexte
non-ordonné*

*Approche intuitive de
l'algorithme*

Apprentissage

*Détection et correction
d'erreurs*

Expérimentations

Méthode expérimentale

Ajout des erreurs

Résultats et interprétation

Conclusion

Exemple d'erreurs possibles :

	pie
	pirée
	paire
pire	poire
	pitre
	pires
	spire

Exemple d'insertion :

J'accuse Monsieur bien qu'il
<error correction="soit">soie</error>
innocent.

*Soutenance de stage
master SAD*

Guillaume Pinot

Introduction

État de l'art

*Méthode du contexte
non-ordonné*

*Approche intuitive de
l'algorithme*

Apprentissage

*Détection et correction
d'erreurs*

Expérimentations

Méthode expérimentale

Ajout des erreurs

Résultats et interprétation

Conclusion



Exemple d'erreurs possibles :

	pie
	pirée
	paire
pire	poire
	pitre
	pires
	spire

Exemple d'insertion :

J'accuse Monsieur bien qu'il
<error correction="soit">soie</error>
innocent.

*Soutenance de stage
master SAD*

Guillaume Pinot

Introduction

État de l'art

*Méthode du contexte
non-ordonné*

*Approche intuitive de
l'algorithme*

Apprentissage

*Détection et correction
d'erreurs*

Expérimentations

Méthode expérimentale

Ajout des erreurs

Résultats et interprétation

Conclusion

- 1 Introduction
- 2 État de l'art
- 3 Méthode du contexte non-ordonné
 - Approche intuitive de l'algorithme
 - Apprentissage
 - Détection et correction d'erreurs
- 4 Expérimentations**
 - Méthode expérimentale
 - Ajout des erreurs
 - Résultats et interprétation
- 5 Conclusion

*Soutenance de stage
master SAD*

Guillaume Pinot

Introduction

État de l'art

*Méthode du contexte
non-ordonné*

*Approche intuitive de
l'algorithme*

Apprentissage

*Détection et correction
d'erreurs*

Expérimentations

Méthode expérimentale

Ajout des erreurs

Résultats et interprétation

Conclusion



Err.	Dét.	Précision	Rappel	Prop. Cor.	PMC	MP
10%	H1	0.1081926	0.9363030	0.9622054	2.57	14.49
10%	H2	0.1561469	0.8800168	0.9635141	2.63	14.63
1%	H1	0.0206164	0.9615384	0.9500000	2.29	14.16
1%	H2	0.0301319	0.9120603	0.9696969	2.56	14.56

Légende :

- Err. : pourcentage d'erreurs ajoutées.
- Dét. : heuristique de détection utilisée.
- Précision : précision de la détection.
- Rappel : rappel de la détection.
- Prop. Cor. : proportion de correction.
- PMC : Position moyenne du mot correct au sein des propositions.
- MP : nombre moyen d'éléments dans la liste des propositions.

Introduction

État de l'art

*Méthode du contexte
non-ordonné*

*Approche intuitive de
l'algorithme*

Apprentissage

*Détection et correction
d'erreurs*

Expérimentations

Méthode expérimentale

Ajout des erreurs

Résultats et interprétation

Conclusion



Participation à la conférence RANLP 2005 (*Recent Advances in Natural Language Processing*) à Borovets, Bulgaria.

Présentation d'un *poster* le jeudi 22 septembre 2005.

*Soutenance de stage
master SAD*

Guillaume Pinot

Introduction

État de l'art

*Méthode du contexte
non-ordonné*

*Approche intuitive de
l'algorithme*

Apprentissage

*Détection et correction
d'erreurs*

Expérimentations

Méthode expérimentale

Ajout des erreurs

Résultats et interprétation

Conclusion

Participation à la conférence RANLP 2005 (*Recent Advances in Natural Language Processing*) à Borovets, Bulgaria.

Présentation d'un *poster* le jeudi 22 septembre 2005.

*Soutenance de stage
master SAD*

Guillaume Pinot

Introduction

État de l'art

*Méthode du contexte
non-ordonné*

*Approche intuitive de
l'algorithme*

Apprentissage

*Détection et correction
d'erreurs*

Expérimentations

Méthode expérimentale

Ajout des erreurs

Résultats et interprétation

Conclusion

- 1 Introduction
- 2 État de l'art
- 3 Méthode du contexte non-ordonné
 - Approche intuitive de l'algorithme
 - Apprentissage
 - Détection et correction d'erreurs
- 4 Expérimentations
 - Méthode expérimentale
 - Ajout des erreurs
 - Résultats et interprétation
- 5 **Conclusion**

*Soutenance de stage
master SAD*

Guillaume Pinot

Introduction

État de l'art

*Méthode du contexte
non-ordonné*

*Approche intuitive de
l'algorithme*

Apprentissage

*Détection et correction
d'erreurs*

Expérimentations

Méthode expérimentale

Ajout des erreurs

Résultats et interprétation

Conclusion



Les avantages de cet algorithme sont :

- la simplicité ;
- l'indépendance de toute information linguistique hormis la définition d'un mot ;
- l'utilisation d'un corpus brut pour l'apprentissage ;
- le peu de paramètres à régler (k la taille du contexte).

Les inconvénients sont :

- la taille des données générées par l'apprentissage ;
- la précision : l'algorithme propose des corrections pour beaucoup de mots corrects.

*Soutenance de stage
master SAD*

Guillaume Pinot

Introduction

État de l'art

*Méthode du contexte
non-ordonné*

*Approche intuitive de
l'algorithme*

Apprentissage

*Détection et correction
d'erreurs*

Expérimentations

Méthode expérimentale

Ajout des erreurs

Résultats et interprétation

Conclusion

Les avantages de cet algorithme sont :

- la simplicité ;
- l'indépendance de toute information linguistique hormis la définition d'un mot ;
- l'utilisation d'un corpus brut pour l'apprentissage ;
- le peu de paramètres à régler (k la taille du contexte).

Les inconvénients sont :

- la taille des données générées par l'apprentissage ;
- la précision : l'algorithme propose des corrections pour beaucoup de mots corrects.

*Soutenance de stage
master SAD*

Guillaume Pinot

Introduction

État de l'art

*Méthode du contexte
non-ordonné*

*Approche intuitive de
l'algorithme*

Apprentissage

*Détection et correction
d'erreurs*

Expérimentations

Méthode expérimentale

Ajout des erreurs

Résultats et interprétation

Conclusion



Les avantages de cet algorithme sont :

- la simplicité ;
- l'indépendance de toute information linguistique hormis la définition d'un mot ;
- l'utilisation d'un corpus brut pour l'apprentissage ;
- le peu de paramètres à régler (k la taille du contexte).

Les inconvénients sont :

- la taille des données générées par l'apprentissage ;
- la précision : l'algorithme propose des corrections pour beaucoup de mots corrects.

*Soutenance de stage
master SAD*

Guillaume Pinot

Introduction

État de l'art

*Méthode du contexte
non-ordonné*

*Approche intuitive de
l'algorithme*

Apprentissage

*Détection et correction
d'erreurs*

Expérimentations

Méthode expérimentale

Ajust des erreurs

Résultats et interprétation

Conclusion



Les avantages de cet algorithme sont :

- la simplicité ;
- l'indépendance de toute information linguistique hormis la définition d'un mot ;
- l'utilisation d'un corpus brut pour l'apprentissage ;
- le peu de paramètres à régler (k la taille du contexte).

Les inconvénients sont :

- la taille des données générées par l'apprentissage ;
- la précision : l'algorithme propose des corrections pour beaucoup de mots corrects.

*Soutenance de stage
master SAD*

Guillaume Pinot

Introduction

État de l'art

*Méthode du contexte
non-ordonné*

*Approche intuitive de
l'algorithme*

Apprentissage

*Détection et correction
d'erreurs*

Expérimentations

Méthode expérimentale

Ajust des erreurs

Résultats et interprétation

Conclusion



Les avantages de cet algorithme sont :

- la simplicité ;
- l'indépendance de toute information linguistique hormis la définition d'un mot ;
- l'utilisation d'un corpus brut pour l'apprentissage ;
- le peu de paramètres à régler (k la taille du contexte).

Les inconvénients sont :

- la taille des données générées par l'apprentissage ;
- la précision : l'algorithme propose des corrections pour beaucoup de mots corrects.

*Soutenance de stage
master SAD*

Guillaume Pinot

Introduction

État de l'art

*Méthode du contexte
non-ordonné*

*Approche intuitive de
l'algorithme*

Apprentissage

*Détection et correction
d'erreurs*

Expérimentations

Méthode expérimentale

Ajust des erreurs

Résultats et interprétation

Conclusion



Les avantages de cet algorithme sont :

- la simplicité ;
- l'indépendance de toute information linguistique hormis la définition d'un mot ;
- l'utilisation d'un corpus brut pour l'apprentissage ;
- le peu de paramètres à régler (k la taille du contexte).

Les inconvénients sont :

- la taille des données générées par l'apprentissage ;
- la précision : l'algorithme propose des corrections pour beaucoup de mots corrects.

*Soutenance de stage
master SAD*

Guillaume Pinot

Introduction

État de l'art

*Méthode du contexte
non-ordonné*

*Approche intuitive de
l'algorithme*

Apprentissage

*Détection et correction
d'erreurs*

Expérimentations

Méthode expérimentale

Ajust des erreurs

Résultats et interprétation

Conclusion



Les avantages de cet algorithme sont :

- la simplicité ;
- l'indépendance de toute information linguistique hormis la définition d'un mot ;
- l'utilisation d'un corpus brut pour l'apprentissage ;
- le peu de paramètres à régler (k la taille du contexte).

Les inconvénients sont :

- la taille des données générées par l'apprentissage.
- la précision : l'algorithme propose des corrections pour beaucoup de mots corrects.

*Soutenance de stage
master SAD*

Guillaume Pinot

Introduction

État de l'art

*Méthode du contexte
non-ordonné*

*Approche intuitive de
l'algorithme*

Apprentissage

*Détection et correction
d'erreurs*

Expérimentations

Méthode expérimentale

Ajout des erreurs

Résultats et interprétation

Conclusion



Les avantages de cet algorithme sont :

- la simplicité ;
- l'indépendance de toute information linguistique hormis la définition d'un mot ;
- l'utilisation d'un corpus brut pour l'apprentissage ;
- le peu de paramètres à régler (k la taille du contexte).

Les inconvénients sont :

- la taille des données générées par l'apprentissage.
- la précision : l'algorithme propose des corrections pour beaucoup de mots corrects.

*Soutenance de stage
master SAD*

Guillaume Pinot

Introduction

État de l'art

*Méthode du contexte
non-ordonné*

*Approche intuitive de
l'algorithme*

Apprentissage

*Détection et correction
d'erreurs*

Expérimentations

Méthode expérimentale

Ajust des erreurs

Résultats et interprétation

Conclusion

Pour compléter cette recherche, il est possible de :

- définir une nouvelle heuristique ;
- créer un algorithme se basant plus sur la syntaxe que sur la sémantique.

*Soutenance de stage
master SAD*

Guillaume Pinot

Introduction

État de l'art

*Méthode du contexte
non-ordonné*

*Approche intuitive de
l'algorithme*

Apprentissage

*Détection et correction
d'erreurs*

Expérimentations

Méthode expérimentale

Ajout des erreurs

Résultats et interprétation

Conclusion

Pour compléter cette recherche, il est possible de :

- définir une nouvelle heuristique ;
- créer un algorithme se basant plus sur la syntaxe que sur la sémantique.

*Soutenance de stage
master SAD*

Guillaume Pinot

Introduction

État de l'art

*Méthode du contexte
non-ordonné*

*Approche intuitive de
l'algorithme*

Apprentissage

*Détection et correction
d'erreurs*

Expérimentations

Méthode expérimentale

Ajout des erreurs

Résultats et interprétation

Conclusion

Pour compléter cette recherche, il est possible de :

- définir une nouvelle heuristique ;
- créer un algorithme se basant plus sur la syntaxe que sur la sémantique.

*Soutenance de stage
master SAD*

Guillaume Pinot

Introduction

État de l'art

*Méthode du contexte
non-ordonné*

*Approche intuitive de
l'algorithme*

Apprentissage

*Détection et correction
d'erreurs*

Expérimentations

Méthode expérimentale

Ajout des erreurs

Résultats et interprétation

Conclusion



- 

Golding, A. R. (1995).
A bayesian hybrid method for context-sensitive
spelling correction.
CoRR, cmp-lg/9606001.
- 

Golding, A. R. and Roth, D. (1999).
A winnow-based approach to context-sensitive
spelling correction.
Mach. Learn., 34(1-3) :107–130.
- 

Golding, A. R. and Schabes, Y. (1996).
Combining trigram-based and feature-based
methods for context-sensitive spelling correction.
In *Proceedings of the 34th conference on
Association for Computational Linguistics*, pages
71–78. Association for Computational Linguistics.
- 

Hirst, G. and Budanitsky, A. (2005).

*Soutenance de stage
master SAD*

Guillaume Pinot

Introduction

État de l'art

*Méthode du contexte
non-ordonné*

*Approche intuitive de
l'algorithme*

Apprentissage

*Détection et correction
d'erreurs*

Expérimentations

Méthode expérimentale

Ajout des erreurs

Résultats et interprétation

Conclusion

Correcting real-word spelling errors by restoring lexical cohesion.

Natural Language Engineering, 11 :87–111.

 Zribi, C. B. O., Fraj, F. B., and Ahmed, M. B. (2005).

Un système multi-agent pour la détection et la correction des erreurs cachées en langue arabe.

In *TALN Récital 2005*, pages 143–152. ATALA.

*Soutenance de stage
master SAD*

Guillaume Pinot

Introduction

État de l'art

*Méthode du contexte
non-ordonné*

*Approche intuitive de
l'algorithme*

Apprentissage

*Détection et correction
d'erreurs*

Expérimentations

Méthode expérimentale

Ajout des erreurs

Résultats et interprétation

Conclusion