Université de Montréal

# CLASSIFIEUR DE MÉTHODES DE RECHERCHE

## Faciliter les revues de littérature mixtes en santé

# CLASSIFIEUR DE MÉTHODES DE RECHERCHE

**Université de Montréal**

Département d'Informatique et de Recherche Opérationnelle

Laboratoire de recherche appliquée en linguistique informatique (RALI)

- Alexis Langlois, étudiant MS
- Jian-Yun Nie, PhD

**Université McGill**

Département de médecine de famille

- Quan Nha Hong, OT, MSc, PhDc
- Pierre Pluye, MD, PhD

**University College London Institute of Education**

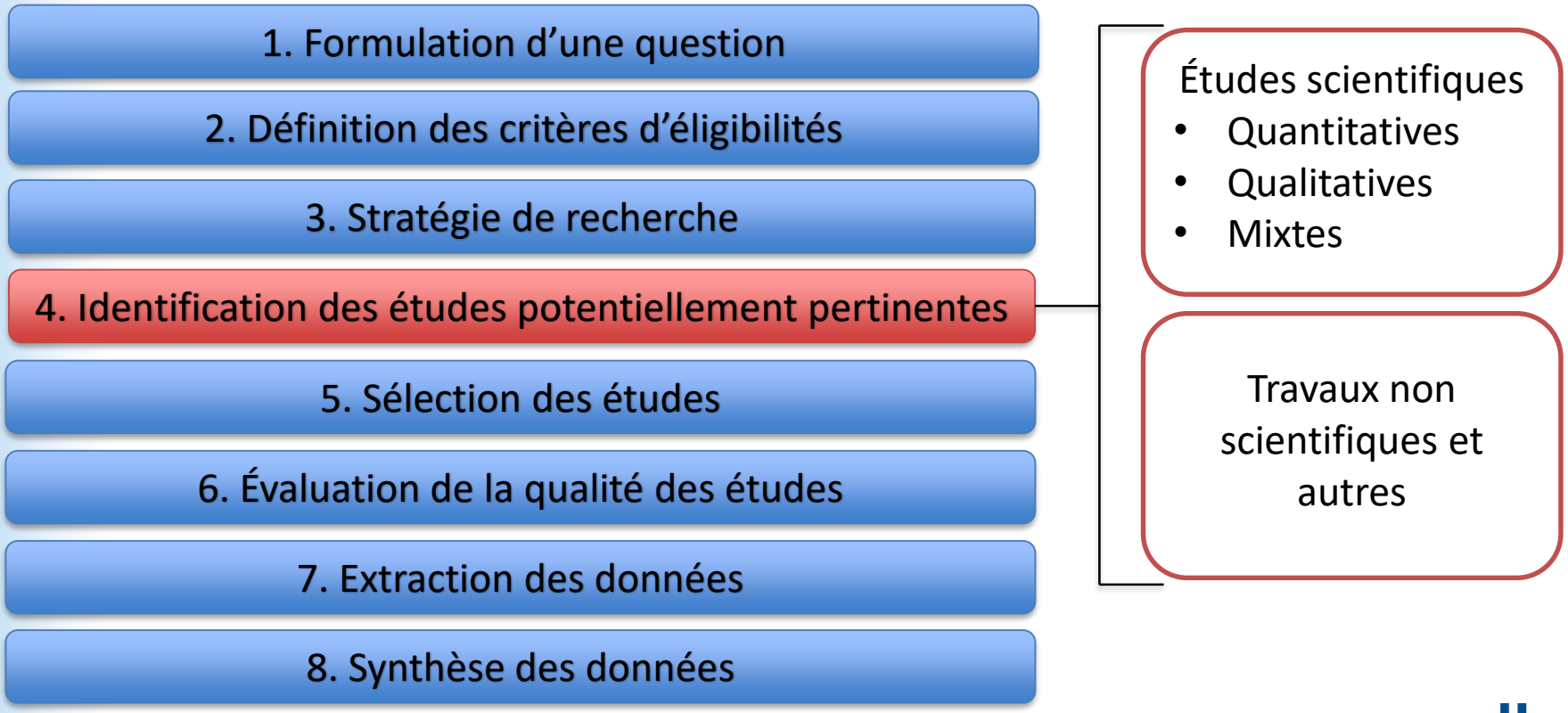
EPPI-Centre

- James Thomas, PhD



# REVUES DE LITTÉRATURE MIXTES

CONTEXTE




Université de Montréal

P. Pluye et Q.N. Hong. Annual Review of Public Health, 2014.

# REVUES DE LITTÉRATURE MIXTES

CONTEXTE

| Méthodes quantitatives | Méthodes qualitatives | Travaux non scientifiques |
|---|---|---|
| • Essai randomisé controllé<br>• Étude non-randomisée<br>  • Étude de cohort<br>  • Essai contrôlé<br>• Étude descriptive<br>  • Série de cas<br>  • Rapport de cas<br>  • Enquête transversal<br>• … | • Entretien<br>• Biographie<br>• Ethnographie<br>• Étude de cas qualitative<br>• Étude descriptive ou interpretative<br>• Phénoménographie<br>• … | • Éditorial<br>• Lettre<br>• Présentation<br>• Commentaire<br>• Débat<br>• Travail méthodologique<br>• … |

Université de Montréal

# REVUES DE LITTÉRATURE MIXTES

## FILTRE MIXTE (BOOLÉEN)

- Combinaison de mots clés pour MEDLINE

  (focus group* OR ethno* …)

  OR (random*.mp. OR control*.mp. ...)

  NOT (letter OR comment …)

- Temps du triage réduit d'environ 50%
- Collection de 4 500 documents
  - 1 400 scientifiques
  - 3 100 non scientifiques
- Résultats
  - Taux de succès ≈ 57%
  - **Rappel ≈ 89%**
  - Précision ≈ 60%
  - Spécificité ≈ 55%

| Inconvénients |
| --- |
| Rigidité (forme fléchie du langage naturel?) |
| Taux de succès ≈ lancer d'une pièce |
| Dépendance aux bases de données bibliographiques |
| Maintenance et ajustements complexes |

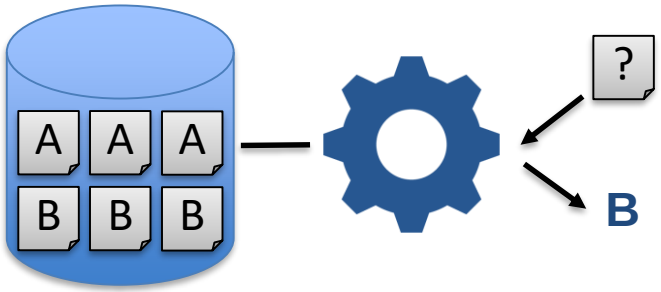
Université de Montréal

R. El Sherif et al. Performance of a mixed filter to identify relevant studies for mixed studies reviews, 2016.

# APPRENTISSAGE MACHINE

## CLASSIFICATION AUTOMATIQUE

- Apprentissage **statistique** guidé par un **algorithme**
- Apprentissage supervisé / **Classification binaire**
    - Scientifique ou non scientifique?
    - Nécessite une quantité raisonnable d'**entrées annotés**
    - **Généralisation** des connaissances à-priori
- Prédictions → Catégorie ou probabilité
- Définition explicite des **caractéristiques** d'une entrée




Université de Montréal

# APPRENTISSAGE MACHINE

## ALGORITHMES

On cherche une fonction mathématique x → y qui généralise bien l'ensemble d'entraînement (entrées déjà annotées)

Plusieurs algorithmes utilisés

- Machine à vecteurs de support
- Réseau de neurones
- Classifieur de Bayes
- Arbre de décision
- Méthode des *k* proches voisins
- Histogramme
- Perceptron
- …

Université de Montréal

# APPRENTISSAGE MACHINE

## CARACTÉRISTIQUES

1. Quelles sont les caractéristiques et les particularités d'un résumé de texte scientifique et non scientifique?
2. Comment les exploiter par un algorithme?

**Indexation** — TF-IDF

 → {mot 1, mot 2, mot 3, empirique} → {0.11, 0.25, 0.12, 1}

**Meta-caractéristiques** — Concepts

{0.11, 0.25, 0.12, 1} → {0.11, 0.12, **0.18, 0.8**, **0.3**, **0.5**, 1}

**Sélection** — Info Gain

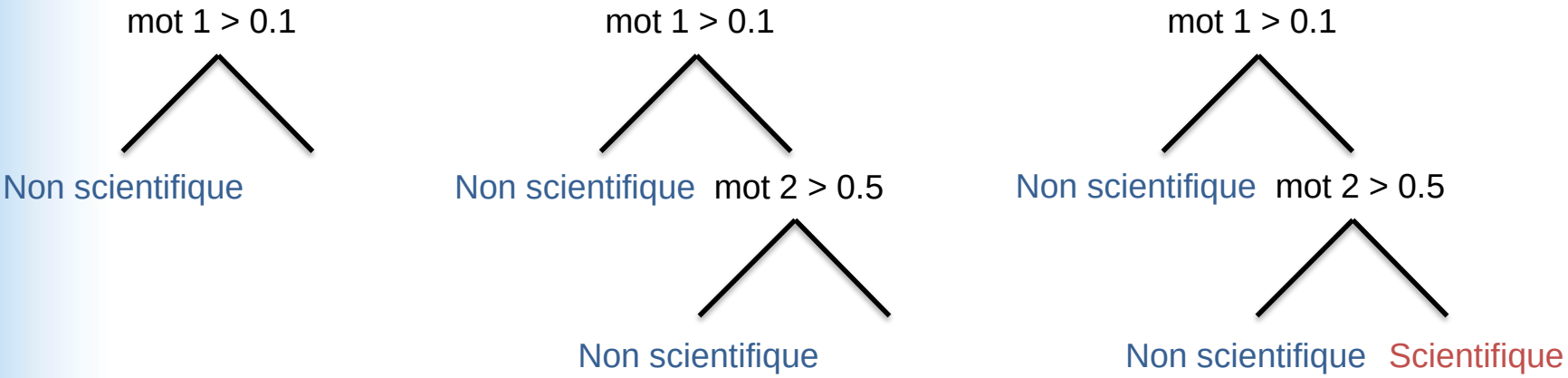
{0.11, 0.25, …, 1} → {0.11, 0.12, 0.8, 1}

Université de Montréal

# CLASSIFIEUR DE MÉTHODES DE RECHERCHE

## ARBRES DE DÉCISION

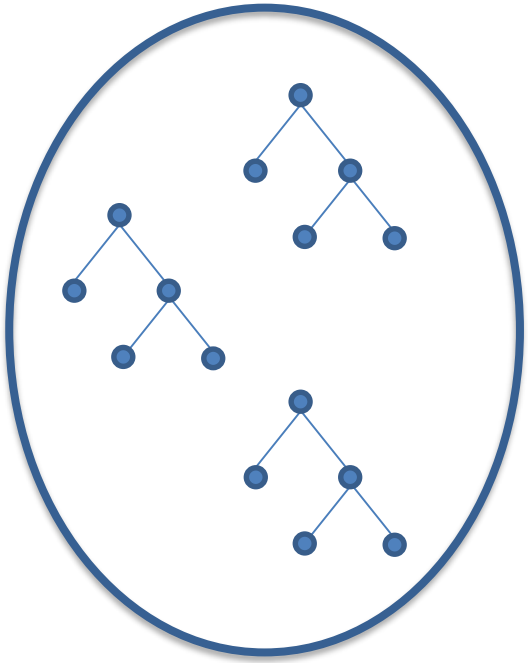
- Suite de règles définissant la classe d'un document:



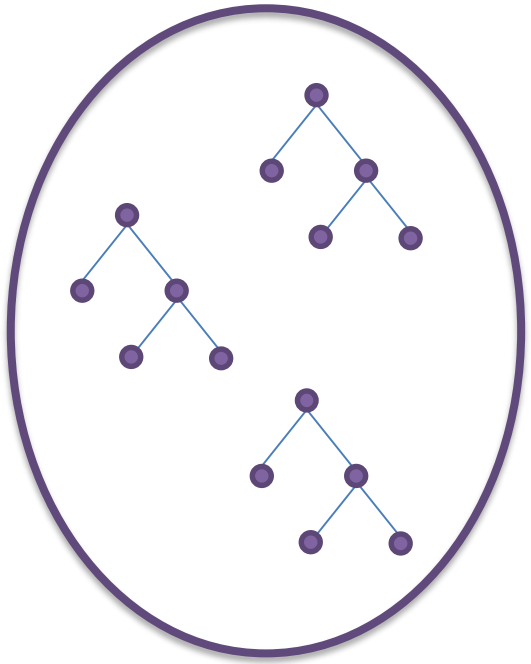

Université de Montréal

# CLASSIFIEUR DE MÉTHODES DE RECHERCHE

MODÈLE FINAL



+



Caractéristiques
formées par les termes

Caractéristiques
formées par les concepts

Université de Montréal

# CLASSIFIEUR DE MÉTHODES DE RECHERCHE

RÉSULTATS

| Catégorie | Précision | Sensibilité | Spécificité | Taux de succès |
|---|---|---|---|---|
| **Titres et résumés de texte** | | | | |
| Scientifique | 81.4 % | 85.2 % | 90.4 % | 88.7 % (+32.1) |
| Non scientifique | 92.5 % | 90.4 % | 85.2 % | |
| **Textes complets** | | | | |
| Scientifique | 86.3 % | 85.4 % | 93.3 % | 90.7 % (+34.1) |
| Non scientifique | 92.8 % | 93.3 % | 85.4 % | |

Université de Montréal

# CLASSIFIEUR DE MÉTHODES DE RECHERCHE

LA SUITE

- **Accessibilité**
  - Outil exécutable et exploitable par les libraires et les chercheurs
  - Interface conviviale sur le web

- **Recherche accentuée**
  - Apprentissage profond
  - Collection de données étendue
  - Accès centralisé et automatisé aux textes intégraux
  - …

Université de Montréal

Université de Montréal

# Merci!