



# AuthorDetector

Un framework pour le développement rapide  
de méthodes d'identification d'auteur

Mises à jours

Par Stephen Larroque

Sous la direction de M. Jean-Gabriel Ganascia

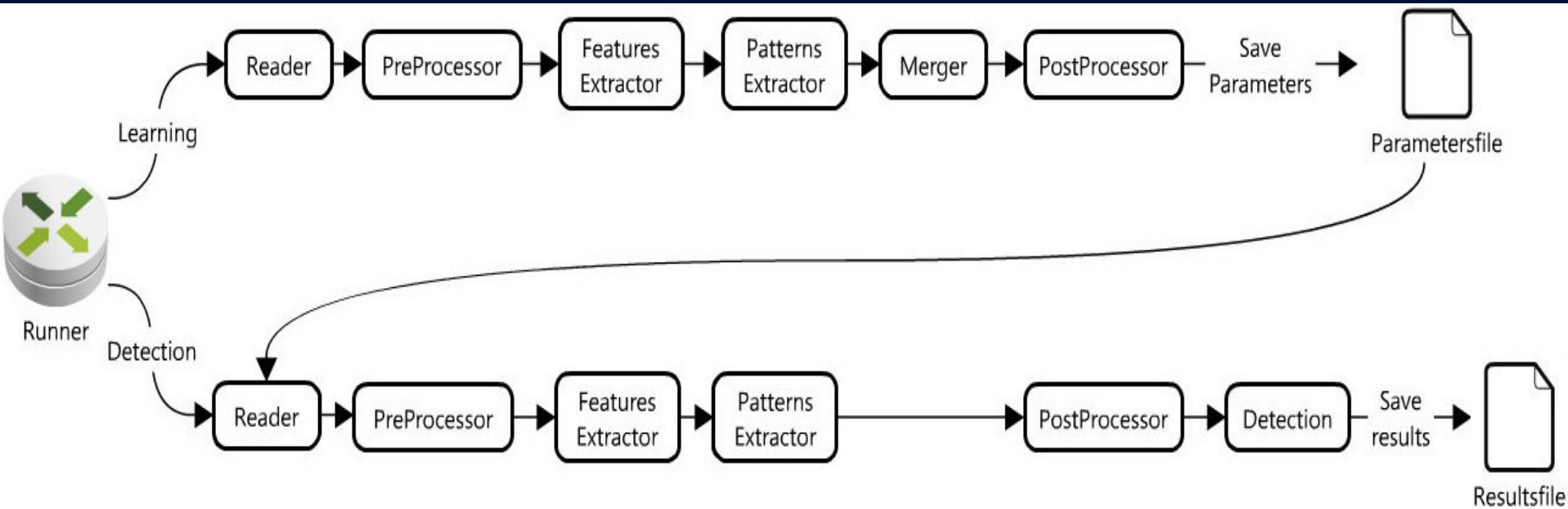
# Problématique

- Identifier l'auteur d'un texte non signé ou sous pseudonyme
- Principale hypothèse: à partir de caractéristiques du texte, on peut distinguer les auteurs

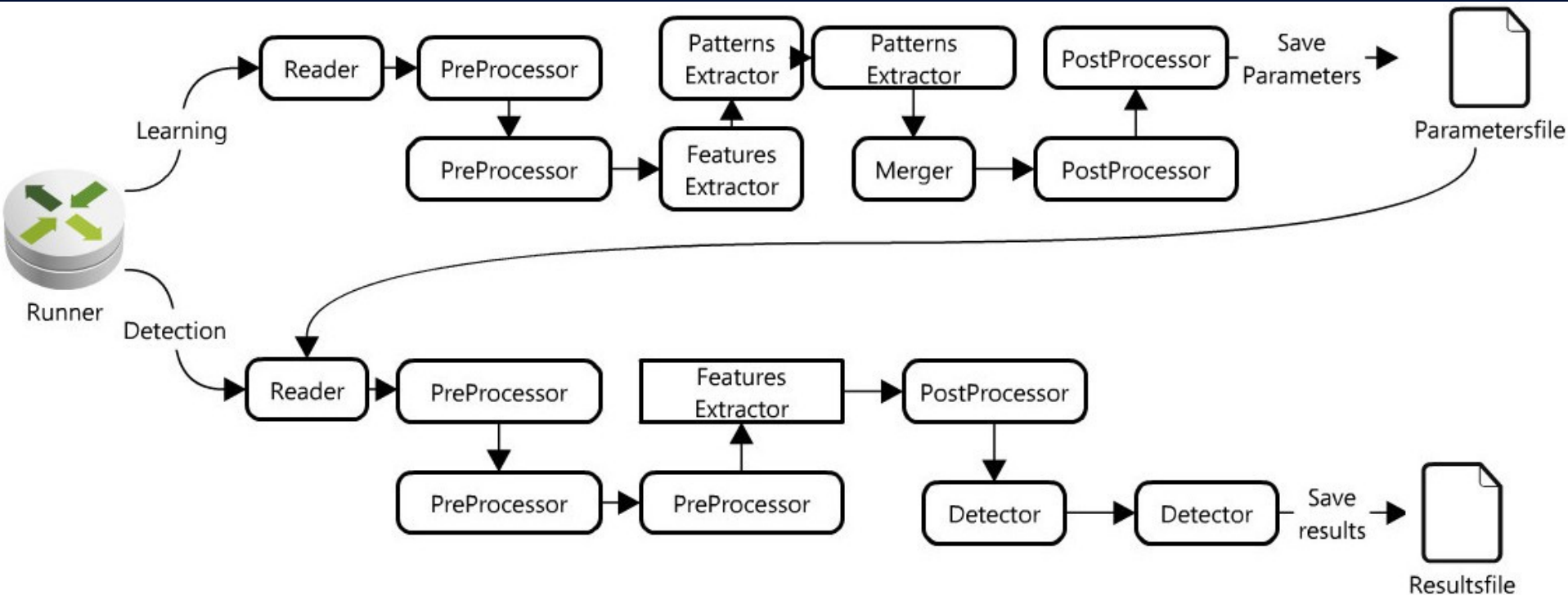
# Mécanisme global

- 2 phases/modes:
  - Apprentissage: apprend les paramètres (motifs récurrents) à partir d'un corpus de textes étiquetés.
  - Détection/Identification: à partir des paramètres appris, identifie l'auteur d'un texte non étiqueté.
- Workflow similaire (sauf exception config)

# Workflow

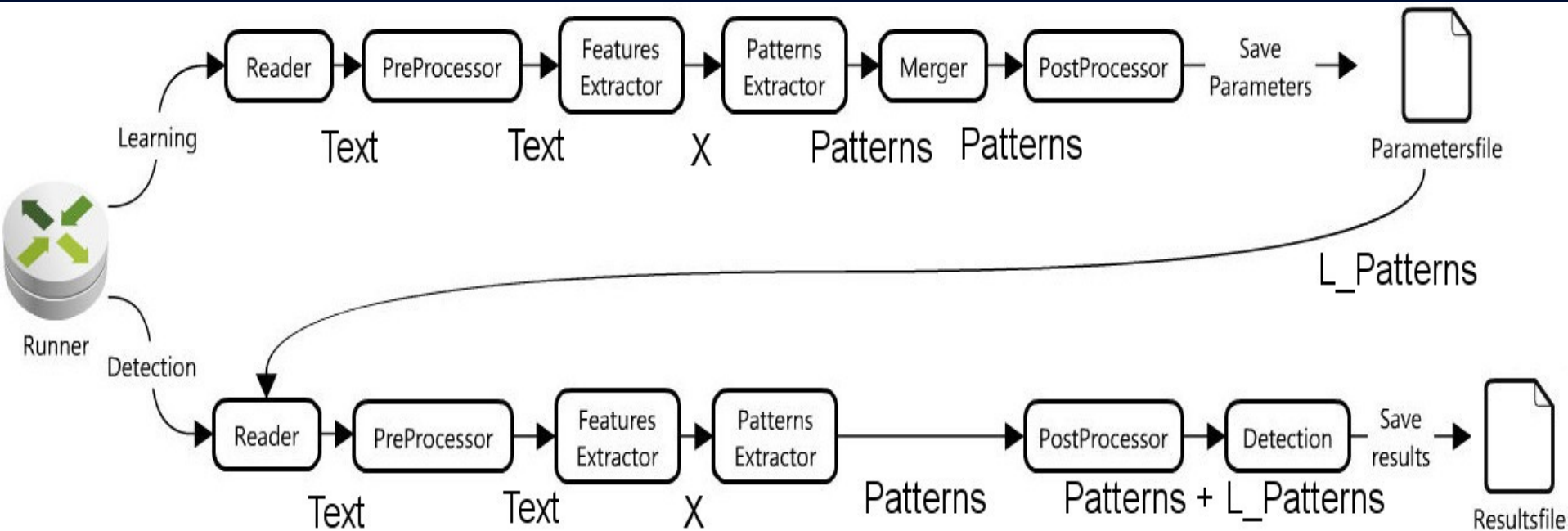


# Workflow 2



- Solution: class aliases

# Cohérence?



# Cohérence

Amélioration de la cohérence:

- Base classes plus qu'une interface: définit le comportement d'un type de module par métadonnées
- Déclaration méthode publique
- Constraints (executer après un autre module)
- Dataflags (traitement variables retournées)



# Nouveaux modules

Amélioration de la cohérence:

- TF-IDF
- Similarité cosinus
- N-grams à trous dynamiques

# Futurs modules

- Matrixify
- Wrapper pour d'autres libs?  
(déjà compatible avec celles supportant Numpy)

**Merci!**

THE END