



## *Association internationale francophone d'Extraction et de gestion des connaissances*<sup>1</sup>

**Objectifs.** — *L'association a pour objet de rassembler les chercheurs académiques et industriels des disciplines de l'informatique décisionnelle, de l'extraction de connaissances dans les bases de données, de la gestion des connaissances, de la fouille de données, de l'apprentissage automatique et de tout autre domaine connexe.*

**Les prix EGC2021 à Montpellier.** — Le prix de la catégorie *article académique* (1 500 euros) a été décerné à Arnaud Giacometti, Béatrice Markhoff et Arnaud Soulet pour leur article intitulé « VERSUS : générateur de tableaux comparatifs à partir de bases de connaissances<sup>2</sup> ».

Le prix de la catégorie *article applicatif* (1 500 euros) a été décerné à Agnès Braud, Pierre Gancarski, Corinne Grac, Agnès Herrmann, Florence Le Ber et Harrison Vernier pour « Classification de séries temporelles hétérogènes pour le suivi de l'état des cours d'eau<sup>3</sup> ».

Le prix de la catégorie *démonstration* (500 euros) a été décerné à Alexis Delaforge, Jérôme Azé, Arnaud Sallaberry, Maximilien Servajean, Sandra Bringay et

---

1. <http://www.egc.asso.fr>.

2. *Revue des nouvelles technologies de l'information*, vol.RNTI-E-37. pp.23-34, 2021, <https://editions-rnti.fr/?inprocid=1002634>.

3. *Revue des nouvelles technologies de l'information*, vol.RNTI-E-37, pp.71-82, 2021, <https://editions-rnti.fr/?inprocid=1002638>.

Caroline Mollevi pour « Expliquer les prédictions des réseaux de neurones par l'exploration de l'espace de représentation et de la frontière de décision à l'aide d'EBBE-Text<sup>4</sup> ».

Le prix de thèse (500 euros) a été décerné à Lamine Diop pour sa thèse « Échantillonnage sous contraintes de motifs structurés », thèse co-encadrée par Cheikh Talibouya Diop (université Gaston Berger de Saint-Louis, Sénégal), Arnaud Giacometti et Arnaud Soulet (université de Tours, France) et résumée maintenant.

La découverte de connaissances dans les bases de données est une activité de recherche qui permet de trouver de nouvelles tendances, corrélations et modèles, à partir d'une très grande masse de données. La littérature de la découverte de motifs a longtemps lutté avec deux problèmes majeurs. Premièrement, il n'est pas possible d'utiliser directement les motifs pertinents si le seuil d'intérêt minimal est petit car ils sont bien trop nombreux. A l'opposé, si le seuil d'intérêt minimal est trop grand, certaines instances seront peu ou pas décrites. Deuxièmement, l'ensemble complet des motifs ayant satisfait la contrainte de seuil d'intérêt minimal peut contenir de nombreuses redondances. L'échantillonnage en sortie est une méthode non exhaustive pour la découverte instantanée de motifs intéressants qui assure une bonne interactivité tout en offrant de solides garanties statistiques en raison de sa nature aléatoire. Curieusement, une telle approche étudiée pour différents types de motifs, y compris les *itemsets* et les sous-graphes, n'a pas encore été appliquée aux motifs séquentiels et aux bases de données distribuées. Cette thèse propose de nombreuses méthodes dédiées à l'échantillonnage en sortie de motifs séquentiels, l'échantillonnage en sortie de motifs dans des bases de données distribuées et l'échantillonnage en sortie de motifs fondé sur les tris. En plus de répondre à ces tâches complexes, l'originalité des approches proposées est d'introduire une classe de mesures d'intérêt reposant sur la norme des motifs, nommée « classe de mesures d'intérêt fondées sur la norme ». En particulier, cette classe permet d'ajouter des contraintes sur la norme des motifs échantillonnés pour contrôler leur longueur et éviter l'écueil de la « longue traîne » où les motifs les plus rares inondent l'utilisateur. Dans ce cadre, il est proposé, en premier lieu, deux algorithmes nommés *NUSSampling* pour les bases de données séquentielles et *DDSampling* pour les bases de données distribuées. Fondés sur des procédures aléatoires en deux étapes intégrant cette classe de mesures, ils tirent au hasard des motifs proportionnellement à la fréquence pondérée par une utilité fondée sur la norme. En second lieu, il est proposé *TPSampling*, un algorithme d'échantillonnage en sortie de motifs ensemblistes fondé sur la structure du tri. Moins consommateur en mémoire, il tire aussi aléatoirement des motifs en fonction de leur fréquence pondérée par une utilité fondée sur la norme. Il est montré que toutes les méthodes proposées effectuent un échantillonnage exact selon la mesure

---

4. Revue des nouvelles technologies de l'information, vol.RNTI-E-37, pp.485-492, 2021, <https://editions-rnti.fr/?inprocid=1002646>.

sous-jacente. Les applications se focalisent sur l'intérêt des contraintes de norme et de décroissance exponentielle qui aident à tirer des motifs généraux de la tête de la longue traîne. On illustre comment profiter de ces motifs échantillonnés pour construire des classificateurs dédiés aux séquences et aux *itemsets*. Cette approche de classification rivalise avec les propositions de l'état de l'art montrant l'intérêt de l'échantillonnage en sortie de motifs avec une mesure d'intérêt fondée sur la norme. On illustre également l'intérêt des motifs échantillonnés sur les données distribuées du Web sémantique pour détecter des entités aberrantes dans DBpedia et Wikidata.

Retrouvez toutes les informations sur notre site web<sup>1</sup> pour concourir aux prochains prix d'EGC2022 à Blois.



## Association Enseignement public et informatique (EPI)<sup>5</sup>

**Finalités.** — *L'EPI est une association loi 1901 fondée en 1971. Conformément à ses statuts, elle veut faire de l'informatique, et des technologies de l'information et de la communication en général, un facteur de progrès et un instrument de démocratisation.*

**Le 50<sup>e</sup> anniversaire de l'association.** — En cette période de pandémie de la Covid-19, l'association Enseignement public et informatique, à l'occasion de son 50<sup>e</sup> anniversaire, un bel âge pour une association, a publié un numéro spécial d'EpiNet<sup>6</sup>. Depuis sa création en 1971, l'EPI accompagne le développement de l'informatique pédagogique dans le système éducatif. À ce titre, elle est « mémoire collective » de ce développement. Son site en est l'illustration. Et l'association est lieu de réflexion, de proposition et d'action.

5. <https://www.epi.asso.fr>.

6. <https://www.epi.asso.fr/epinet/epinet232.htm>.