



# PACTE

## Apprendre aux machines à chercher plus vite

L'humanité est ensevelie sous les données, qui ne cessent de se multiplier, et les scientifiques doivent souvent creuser à travers des milliers de points de données avant de trouver ce qu'ils cherchent et d'assembler des statistiques. Bien qu'il excelle dans certaines tâches de traitement de texte comme la recherche de mots-clés, l'ordinateur se perd aisément dans les complexités du langage humain et bute quand il tombe sur des coquilles, des expressions idiomatiques, une orthographe vieillotte et, surtout, le sens caché des mots. Sans formation en langage naturel, il ne peut passer au crible pareille masses de données textuelles, sauf dans les cas les plus simples.

C'est ce qui a incité les chercheurs du Centre de recherche informatique de Montréal (CRIM) à créer PACTE, une plateforme collaborative d'annotation et d'analyse de texte. On ne peut apprendre à un ordinateur à saisir un texte sans annotations, petites étiquettes glissées ici et là pour lui expliquer ce que signifie telle ou telle expression. Les annotations expliquent comment le texte est structuré grammaticalement, les sens qu'on lui prête et les particularités syntaxiques auxquelles il faut porter attention. La plateforme PACTE a été conçue pour simplifier l'apprentissage automatique. Elle gère d'énormes bases de données textuelles, autorise l'annotation manuelle du texte, perfectionne les algorithmes d'apprentissage, puis passe à l'annotation automatique et à l'analyse des résultats.

## Annotation automatique et collaboration répartie

La plateforme PACTE est unique en raison de son approche collaborative. En répartissant la tâche grâce à un portail Web, une armée d'annotateurs se divisent le travail sur une masse de données. Une fois que leurs annotations ont été saisies en nombre suffisant, des équipes apprennent à la plateforme à s'en servir. PACTE applique alors des algorithmes qui annotent automatiquement le reste du

*PACTE est une plateforme collaborative d'annotation et d'analyse de texte. PACTE a été conçue pour simplifier l'apprentissage automatique.*

texte, réduisant de façon draconienne le temps nécessaire pour réaliser d'autres recherches et analyses.

Si l'apprentissage automatique figure parmi les applications utiles de PACTE, le texte annoté a lui aussi des usages typiquement humains. En effet certains textes sont très courts ou ultra spécialisés. Ils ne conviennent donc pas à l'apprentissage machine. Or, les spécialistes dans un domaine précis sont souvent disséminés, ici et là sur le globe. En laissant de nombreux collaborateurs annoter un texte, PACTE facilite les projets très pointus comme l'identification d'auteurs inconnus des pièces du théâtre parisien du dix-septième siècle ou la cartographie des voyages d'un ancien explorateur à partir de ses carnets de notes.

## Résoudre des problèmes très concrets

Tout récemment, on s'est servi de la plateforme PACTE pour traiter des rapports d'intervention dans des cas de délinquance juvénile. Les travailleurs sociaux rédigent des monceaux de rapports qui n'ont qu'une vie éphémère, les renseignements personnels sur les jeunes délinquants étant protégés. En raison de leur nombre, demander à quelqu'un d'analyser ces rapports avant leur destruction s'avèrerait difficile. Les chercheurs recourent donc à PACTE pour repérer et classer automatiquement les interventions en vue d'améliorer la qualité et l'efficacité du travail social.

## Réutilisation du logiciel pour de nouvelles découvertes

PACTE fait partie d'un riche ensemble de logiciels réutilisables, financés par CANARIE. La plateforme s'articule sur un projet antérieur, également subventionné par cette organisation, le système d'annotation de vidéos VESTA, qui offre des services de transcodage de fichiers audio et vidéo ainsi que d'extraction du langage. PACTE a également ajouté divers services logiciels au [Registre des plateformes logicielles de recherche](#) de CANARIE, notamment un logiciel d'analyse lexicale et linguistique, un dépôt d'annotations, un service de transcodage des fichiers texte et un autre d'analyse sémantique.

---

### Plateforme : PACTE

Description	Aide l'utilisateur à annoter de grands corpus de texte grâce à une interface Web. Services automatisés d'annotation linguistique, lexicale et sémantique dans les deux langues officielles. La plateforme comprend également un service d'apprentissage dynamique qui identifie les données les plus significatives à annoter, de manière à permettre une annotation semi-automatique.
Contributeur(s)	Centre de recherche informatique de Montréal (CRIM)
Champ de recherche	Traitement du langage naturel
Portail	<a href="http://patx-pacte.crim.ca">patx-pacte.crim.ca</a>
Accès au portail	Usage libre à des fins non lucratives sur demande
Autorise la poursuite de projets distincts	Oui
Science participative	Oui
Licence d'exploitation du logiciel	Propriété exclusive et source ouverte
En savoir plus	<a href="https://science.canarie.ca/res/115">https://science.canarie.ca/res/115</a>