

Connectôme : vers des services de données scientifiques ouverts grâce à l'IA

L'objectif du projet suisse Connectôme est d'organiser les métadonnées ouvertes nationales et internationales pertinentes de manière durable afin de les rendre facilement accessibles.

BIBLIOTHÈQUE
CANTONALE ET
UNIVERSITAIRE
BCU LAUSANNE

Personne ne doute plus du fait que les données de la recherche sont un atout essentiel pour les universités. L'accélération technologique de la recherche entraînant par ailleurs une augmentation de la masse de données ouvertes qui se retrouvent stockées dans diverses archives (inter)nationales décentralisées, et dépôts institutionnels, c'est donc un grand défi pour les universités et leurs bibliothèques de savoir quelles données sont stockées et où, quelle est leur qualité, comment elles peuvent être obtenues, récoltées collectivement, réutilisées, et comment elles peuvent être liées entre les disciplines afin de pouvoir engendrer de nouvelles recherches révolutionnaires. Afin de compléter l'infrastructure nationale, SWITCH¹, le prestataire du réseau national suisse de recherche et d'éducation, a lancé le projet Connectôme². La vision de Connectôme est d'interconnecter et d'organiser les métadonnées ouvertes nationales et internationales pertinentes de manière durable dans toutes les disciplines, afin de les rendre facilement trouvables, largement accessibles, interopérables et à valeur ajoutée.

HARMONISER, ENRICHIR, INTERCONNECTER LES DONNÉES

L'objectif de l'équipe à l'origine de cette initiative est de récolter, d'harmoniser, d'enrichir et d'interconnecter les métadonnées des fournisseurs de données décentralisés, et de les faire entrer dans le graphe de connaissances de Connectôme en utilisant des normes ouvertes et les meilleures pratiques internationales. Les données ouvertes liées qui en résultent sont alors utilisées pour activer de nouvelles fonctions de recherche et de découverte susceptibles de soutenir les cycles de vie des données de la recherche et de l'éducation. La mise en place de cette infrastructure représente un nouveau domaine

de développement stratégique de SWITCH. Dès son lancement, ce projet a intéressé les bibliothèques, étant donné que le Connectôme permettra d'offrir non seulement des services d'enrichissement automatique de métadonnées décrivant les publications et l'ORD, mais également la classification automatique des publications électroniques selon la systématique utilisée par une bibliothèque. Connectôme sera également en mesure de fournir des résumés des articles scientifiques aux lecteurs de niveau primaire ou secondaire, et d'ouvrir ainsi ces contenus à un nouveau périmètre de lecteurs.

UNE MISE EN ŒUVRE EN 2022

L'infrastructure a été créée en 2020 en collaboration avec 6 partenaires initiaux et étendue à 9 partenaires suisses en 2021. Certains partenaires ont été impliqués dans le développement du concept, la conception et les efforts de préparation, tandis que d'autres ont fourni un soutien par la livraison de logiciels *open source* et/ou ont offert des conseils et un retour d'information continu³. Des chercheurs de diverses disciplines ont été inclus dans le codéveloppement continu de l'infrastructure. Ils ont contribué au développement du graphe de connaissances en apportant des compétences sur les ontologies spécifiques.

Début 2022, le projet est entré dans sa phase de mise en œuvre. L'accent est actuellement mis sur l'extension des partenariats ainsi que sur le développement, l'évaluation stratégique et le déploiement plus large de l'infrastructure de base et des services de données tels que décrits ci-dessous.

L'INFRASTRUCTURE DE BASE DE CONNECTÔME

L'infrastructure de base permet d'assurer des services de données de recherche ouverts, autant du point de vue de la plate-

forme que des personnels nécessaires pour développer, exploiter et améliorer l'infrastructure de recherche, qui comprend le mappage des données pour convertir les métadonnées en une structure de données commune (RESCS.org), le graphe de données permettant d'importer et stocker les métadonnées liées, ainsi que l'API Connectôme pour rechercher, découvrir et extraire les données ouvertes liées. L'équipe fournit un savoir-faire en matière de données afin de collaborer avec la communauté suisse de l'éducation, de la recherche et de l'innovation (ERI) à la cocréation des solutions. Les éléments-clés de l'infrastructure de base sont alignés sur les engagements internationaux pertinents, par exemple à travers la participation à deux groupes de travail d'EOSC.

Les services de données de l'infrastructure de base permettent de répondre aux besoins d'amélioration de l'utilisation des données de recherche ouverte des parties prenantes du secteur ERI. Les services de données sont créés conjointement avec la communauté et gérés par SWITCH. Quatre services de données sont actuellement en développement et en évaluation continue :

• L'enrichissement des données

Ce service fournira différents outils pour augmenter la qualité et la réutilisation des données ouvertes par l'enrichissement des métadonnées. Les cas d'usage en cours de développement sont les suivants :

– Extraction et/ou génération d'informations de localisation à l'aide d'algorithmes de reconnaissance d'entités nommées afin d'enrichir les métadonnées de la collection Memobase dans le but de réutiliser les emplacements extraits dans les cartes de l'interface utilisateur et les chronologies des services.

– Désambiguïsation des noms d’auteurs pour les métadonnées de publications et de projets de recherche à l’aide d’un réseau neuronal profond entraîné sur les noms et prénoms suisses en utilisant des données ouvertes, manuellement étiquetées et vérifiées, provenant du Fonds national suisse de la recherche scientifique.

– Enrichissement des métadonnées d’organisation en utilisant des sources de données externes telles que les propriétés de Wikidata.

– Extraction de métadonnées contextuelles à partir d’images à l’aide d’algorithmes de vision formés manuellement sur des données d’images ouvertes provenant de Wikidata en collaboration avec la Haute école spécialisée bernoise et Wikimedia Suède.

– Transformation de métadonnées d’archives (Renouvaud Patrimum) en données ouvertes liées à des fins de réutilisation (par exemple, pour afficher des données ouvertes liées connexes).

• L’API Connectôme

L’API Connectôme vise à permettre aux fournisseurs de services tels que les archives, les dépôts ou les plateformes de découverte d’interroger un graphe de données ouvertes et d’extraire des informations à des fins de recherche et d’éducation.

Les cas d’utilisation actuellement en cours de développement sont les suivants :

– Accès aux données ouvertes liées via des fonctions API génériques (par exemple, récupération, recherche, exportation).

– Exportateurs spécifiques au client qui restructurent automatiquement les données ouvertes liées dans des structures de table spécifiques pour faciliter la réutilisation⁴.

• Insights en tant que service

Les *insights* sont des modèles significatifs dans les données qui permettent une prise de décision plus efficace.

Ce service vise à générer de nouvelles connaissances à partir de données ouvertes liées. Pour ce faire, des techniques d’intelligence artificielle sont utilisées pour analyser les données ouvertes liées dans le graphe de connaissances du Connectôme. Le public cible est constitué de fournisseurs de services et de données pour la recherche et l’éducation.

Les cas d’utilisation en cours de développement sont les suivants :

– Découverte d’informations de recherche similaires et/ou connexes et suggestions (recherche en cours de saisie) à l’aide d’algorithmes de recommandations.

– Analyse des réseaux (d’auteurs) pour identifier et/ou classer la pertinence des auteurs pour un groupe de sujets de recherche thématiques générés automatiquement et visualisation des graphes des réseaux résultants.

– Extraction automatique de mots-clés et leur catégorisation à partir de textes téléchargés et de documents scientifiques à l’aide d’algorithmes de modélisation thématique afin de découvrir des projets de recherche, des publications et des ensembles de données connexes.

– Résumé automatique et simplification des résumés des publications en *open access* à l’aide de grands modèles de langage pré-entraînés afin d’améliorer la compréhensibilité et la visibilité pour les différentes parties prenantes (par exemple, pour que les élèves et les citoyens puissent comprendre le contenu des articles de recherche).

• La plateforme de découverte

Une plateforme de découverte, actuellement utilisée à des fins de présentation, utilise l’API Connectôme et *Insights as a Service* pour améliorer l’expérience de recherche et de découverte des utilisateurs finaux. Les liens sémantiques visent à permettre aux utilisateurs finaux de rechercher des personnes, des organisations, des projets, des publications et des ensembles de données. Les fonctions d’aperçu permettent de retrouver des similitudes, des recommandations, des analyses de réseau et des visualisations. En outre, l’équipe travaille sur des simplifications automatisées et des résumés de publications, le calcul de scores d’ouverture et la recommandation de ressources sur la base de l’analyse de textes/documents.

UN OUTIL ADAPTÉ POUR LES BIBLIOTHÈQUES

Deux bibliothèques suisses sont actuellement impliquées dans la cocréation des cas d’utilisation sélectionnés. Le premier cas d’utilisation implique un processus de mappage, basé sur RML, des métadonnées d’archives de la base de données de patri-

moine numérique de la BCU Lausanne, Patrimum⁵ (MARCXML) aux structures de données ouvertes liées RiC-O et RESCS (RDF) avec le soutien de la Haute école spécialisée des Grisons dans le cadre d’un SWITCH Innovation Lab. Ces processus génèrent une base de qualité nécessaire pour fournir du *linked open data* (LOD) et du *data insight* à Patrimum et des enseignements pertinents pour les autres archives et bibliothèques. Un autre cas d’utilisation vise à récolter et filtrer les métadonnées bibliographiques de la *Swiss Library and Service Platform* - SLSP pour fournir du LOD aux autres fournisseurs de données et de services (par exemple la *Swiss AI Research Overview Platform*).

Les bibliothèques sont les championnes de la collecte, de la conservation, de la gestion durable et de la réutilisation des métadonnées archivistiques, bibliographiques et de recherche. Une infrastructure de recherche ouverte telle que le Connectôme vise à aider les bibliothèques et leurs utilisateurs à enrichir qualitativement les métadonnées existantes (par exemple par des extractions d’entités, des enrichissements avec des données provenant d’autres fournisseurs), à interconnecter les métadonnées bibliographiques avec les données (de recherche) ouvertes et à réutiliser les données ouvertes liées pour élaborer de nouvelles fonctionnalités de recherche, de découverte et de réutilisation pour les utilisateurs finaux des plateformes et des services des bibliothèques.

JEANNETTE FREY

Directrice de la Bibliothèque cantonale et universitaire de Lausanne
jeannette.frey@bcu.unil.ch

[1] www.switch.ch

[2] www.switch.ch/connectome

[3] Voir par exemple sur le Blue Brain Nexus : www.semantic-web-journal.net/content/blue-brain-nexus-open-secure-scalable-system-knowledge-graph-management-and-data-driven

[4] www.sairop.swiss par SATW

[5] <https://patrimum.ch/?ln=de>