

ÉTUDE DE CAS



DZD
Deutsches Zentrum
für Diabetesforschung

Le Centre allemand de recherche sur le diabète (DZD)

Le Centre allemand de recherche sur le diabète améliore considérablement ses capacités de recherche avec la technologie des graphes

SECTEUR

Bioinformatique /
Recherche médicale

CAS D'USAGE

Gestion des données
de référence /
Gestion des connaissances

DÉFI

Fusionner les données de différents sites et sources et en différents formats en une base de données centrale pour doter les chercheurs allemands de moyens de recherche plus efficaces pour prévenir et soigner le diabète.

SOLUTION

- La base de données de graphes DZDconnect pour des requêtes centralisées des métadonnées
- Langage de requête intuitif et visualisation puissante pour aider à une recherche efficace

RÉSULTATS

- Intégration rapide des données d'essais cliniques du DZD
- Identification de relations entre données destinées à être approfondies dans de nouveaux projets de recherche

Pour que les chercheurs sur le diabète bénéficient d'un meilleur accès aux données, le Centre allemand de recherche sur le diabète (DZD) élabore une nouvelle ressource de gestion des connaissances à base de graphes. Celle-ci s'appuie sur Neo4j, et combine les métadonnées issues d'une grande variété de recherches fondamentales et d'essais cliniques.

L'organisation

Le Centre allemand de recherche sur le diabète (DZD), fondé en 2009, est financé par le Ministère fédéral allemand de l'éducation et de la recherche (BMBF) et les états fédéraux associés. Il rassemble des experts de toute l'Allemagne pour développer des mesures efficaces de prévention et de traitement du diabète dans diverses disciplines et bénéficier des technologies biomédicales les plus récentes.

Le défi

Environ 7 millions d'Allemands souffrent de diabète, c'est l'une des maladies les plus répandues dans le pays. Afin de mieux comprendre ses causes, les chercheurs du DZD examinent la maladie sous un maximum d'angles possibles. Ils accumulent des volumes massifs de données disséminées et travaillent à établir une base de données de référence pour consolider toutes ces informations. Cette base de données va apporter à l'équipe de 400 chercheurs une vue holistique des informations disponibles et leur permettre d'en tirer des informations instructives sur les causes et la propagation du diabète.

La stratégie

«Notre objectif consiste à mettre en place un système centralisé de gestion des données et des connaissances,» explique le professeur et Dr. Martin Hrabě de Angelis, membre du conseil d'administration du DZD. «Nous donnons aux chercheurs les moyens de trouver des indices de mesures efficaces de prévention et de traitement dans toutes les disciplines, chez toutes les espèces et à partir de tout type de données pour empêcher le diabète de se déclarer et de se développer.»

À la recherche d'un outil adapté à l'objectif, le Dr. Alexander Jarasch, à la tête du management de la bioinformatique et des données du centre, a eu par le passé l'occasion de travailler sur un projet du Helmholtz Zentrum de Munich (centre de recherche en santé environnementale) qui reliait les données de séquences d'ADN avec des données métabolomiques. Il avait ainsi exécuté des requêtes en utilisant Neo4j, ce qui l'a décidé à l'employer au DZD également. «Créer les premiers modèles de données avec Neo4j a été très rapide. La première semaine, nous pouvions déjà connecter les métadonnées récupérées par nos scientifiques dans un modèle de données, le tester et montrer les avantages de la base de données de graphes.»

La solution

Avec l'aide de Neo4j, le Dr. Jarasch a mis en œuvre un nouvel outil interne nommé DZDconnect. La base de données de graphes sert de couche au-dessus des diverses bases de données relationnelles reliant les différents systèmes et silos de données du DZD. DZDconnect n'est pas encore complètement déployé, mais les chercheurs peuvent déjà accéder aux

ÉTUDE DE CAS



« Créer les premiers modèles de données avec Neo4j a été très rapide. La première semaine, nous pouvions déjà connecter les métadonnées récupérées par nos scientifiques dans un modèle de données, le tester et montrer les avantages de la base de données de graphes. »

– Dr. Alexander Jarasch,
Management de la bioinformatique
et des données, DZD, Allemagne

métadonnées des études cliniques dans un prototype. La visualisation des graphes et la facilité de requête que permet Neo4j les impressionnent particulièrement.

Les chercheurs veulent savoir comment certaines informations deviennent rapidement plus pertinentes : Le patient était-il à jeun ou venait-il de consommer du glucose ? Comment le niveau de glucose a-t-il évolué ? Et cette évolution peut-elle être attribuée à un régime, un traitement médical ou un facteur encore inconnu ? Selon les prévisions du DZD, il sera bientôt possible de répondre à ces questions importantes et complexes avec les graphes.

À l'avenir, les données de mesure seront également automatiquement transmises à DZDconnect. Un premier cas d'usage concerne les analyses de sang : si un groupe de recherche X mesure le niveau A de glucose de l'échantillon de sang 0001, cette valeur est stockée dans la base de données, ainsi que le lieu de conservation et l'heure de la mesure. D'autres chercheurs peuvent, à leur tour, utiliser ces informations dans leur propre travail sans devoir réitérer les mesures déjà prises ni gâcher de précieux échantillons. Il en découle un cadre d'information dans lequel les scientifiques peuvent mener des recherches parallèles sans poser de nouveau les mêmes questions ou refaire des essais.

L'outil de visualisation Neo4j Bloom permettra à l'avenir de poser des questions en langage naturel : combien d'échantillons de sang de patients masculins âgés de moins de 69 ans avons-nous reçus ? D'où viennent ces échantillons ? Quels sont les paramètres de mesure ? Plus les informations sont détaillées, plus il est facile d'identifier des relations et des modèles.

Les résultats

Les premiers résultats sont positifs. «Avec la technologie de graphes de Neo4j, nous avons pu combiner et interroger des données réparties en différents endroits. Même si les données n'ont été intégrées qu'en partie, les requêtes ont déjà révélé des relations intéressantes que nos scientifiques vont pouvoir étudier plus en détail. À long terme, autant de données du DZD que possible vont être intégrées dans la base de données de graphes» indique le docteur Jarasch.

Il indique également que la prochaine étape va consister à voir comment les données humaines des études cliniques peuvent être complétées avec les données hautement standardisées de modèles animaux - par exemple des souris - pour trouver des points communs et d'autres informations. Comment des facteurs tels qu'un régime alimentaire ou l'inactivité physique affectent-ils la maladie ? Quels gènes jouent un rôle dans le type de diabète ?

Ces sortes de différences, de modèles ou de relations peuvent facilement être identifiées dans la base de données de graphes.

L'apprentissage machine et le traitement du langage naturel vont jouer un rôle majeur pour la suite. Il est particulièrement intéressant de mettre en place un système capable de "lire" automatiquement les textes scientifiques et de les intégrer dans la base de données, prêts à être analysés. «Grâce à la technologie, il est plus facile de considérer les enjeux médicaux sous différents angles et selon toutes les indications,» souligne le docteur Jarasch. «Il est également possible d'identifier des corrélations entre le diabète et d'autres maladies très répandues.»

Neo4j est le leader de la technologie des bases de données de graphes. Avec le plus grand nombre de déploiements au monde, Neo4j aide des entreprises mondiales - comme Airbus, [Michelin](#), [NASA](#), [Crédit Agricole](#) et [Volvo Cars](#) - à prédire et identifier la façon dont les personnes, les processus, les lieux et les systèmes sont interconnectés. Grâce à cette approche par les relations, les [applications](#) mises au point en utilisant Neo4j relèvent les défis associés aux données connectées, tels que [l'analytique et l'intelligence artificielle](#), la [détection de fraude](#), les [recommandations](#) en temps réel et les [graphes de connaissance](#). Pour en savoir plus, merci de consulter [Neo4j.com](#) et [@Neo4jFr](#).

Des questions sur Neo4j ?

Contactez-nous :
info@neo4j.com
neo4j.com/contact-us