

Cinq secondes pour retrouver un document dans une base de 1 Téraoctet

Par *admin*

Créé le 13/12/2003 - 00:00

Cinq secondes pour retrouver un document dans une base de 1 Téraoctet

Vendredi, 12/12/2003 - 23:00 [0 commentaire](#)

- [Diminuer la police](#)
- [Augmenter la police](#)
- [Imprimer](#)
- [Version PDF](#)

-
- [Tweeter](#)
-
-

0 avis :



[zoom](#)

Cinq secondes pour retrouver un document dans une base de 1 Téraoctet : c'est l'exploit réalisé par XediX, la base de données native XML développée par le CEA, dont les performances surpassent largement celles de tous les outils équivalents du marché. AM2 Systems, une start-up de l'INRIA, vient de lancer la commercialisation. Premiers utilisateurs visés : l'archivage, les médiathèques, les projets coopératifs de recherche ; et à terme, tous les gestionnaires de grandes bases de données. Un Téraoctet, cela correspond à plus d'un millier de fois la collection complète de l'Encyclopaedia Universalis. Un océan, une immensité de données, au milieu desquelles la recherche d'un document ou d'une référence précise devient une gageure. A ce jour, aucun produit du marché ne s'y risque : les plus avancés annoncent 300 à 400 Go avec des temps de réponse modestes, et la norme se situe plutôt vers 200 Go, y compris chez les grands éditeurs mondiaux. ?C'est d'ailleurs après les avoir rencontrés que nous avons décidé de développer notre propre outil, précise Didier Courtaud, de la Direction des Applications militaires du CEA. Nous voulions de la pérennité et de la performance, pour de grands volumes, toutes choses bien présentes dans XediX?. ?L'entrepôt de documents numériques?, pour reprendre la terminologie en vigueur, est utilisé depuis deux ans par le CEA pour ses besoins propres. Sa grande

spécificité : il stocke et gère les données textuelles exclusivement en XML*, langage de description indépendant des logiciels de bureautique et de leurs incessantes évolutions ; quant aux fichiers images ou vidéo, ils sont stockés dans la base et répertoriés en XML sous forme de métadonnées décrivant le sujet, la date de tournage, les personnages ou tout autre critère fixé par le gestionnaire de la base. ?L'utilisateur détermine librement les modèles de structuration de ses données et peut les faire évoluer, explique Didier Courtaud. Il ne subit pas les contraintes des bases de données objet ou relationnelles, dont les schémas sont figés?. Cette philosophie ?tout XML? est la clé des incroyables performances de recherche de XediX. Car le moteur de recherche, intégré à l'outil et non interfacé, n'a pas à explorer toute la base de données : il sélectionne directement les éléments correspondants à la requête. ?C'est le nombre d'éléments pertinents, et non la taille de la base, qui détermine le temps de réponse?. Celui-ci est de 5 secondes en moyenne pour une base de 1 Téraoctet et en théorie, le même score serait possible avec 5 ou 10 Téraoctets, voire davantage. Le ?tout XML? permet aussi une gestion ultra-fine des droits d'accès, qui peuvent être définis à l'extrême par document et par utilisateur ! Avec de tels atouts, XediX intéresse déjà les secteurs de l'archivage et des médiathèques, ainsi que les communautés de recherche (projets européens, projets recherche-industrie). Il vient d'être commercialisé par AM2 Systems, une start-up de l'INRIA dirigée par Alain Michard, expert XML de notoriété européenne. Des signes qui ne trompent pas : ce ?super-outil? a un bel avenir devant lui.

CEA : <http://www.cea-technologies.com/infocom/68-201.html>

Noter cet article :

Recommander cet article :

-
- [Tweeter](#)
-

- **Nombre de consultations :** 103
- **Publié dans :** [Informatique](#)
- **Partager :**
 - [Facebook](#)
 - [Viadeo](#)
 - [Twitter](#)
 - [Wikio](#)

[Informatique](#)

URL source: <https://www.rtfash.fr/cinq-secondes-pour-retrouver-document-dans-base-1-teraoctet/article>