Le CERN accélère la transmission de ses informations sur les mobiles

Dominik.Stankowski@eif.ch, EIA-FR,
Jean-Yves.Le.Meur@cern.ch,CERN &
Omar.AbouKhaled@eif.ch, EIA-FR

S'informer sur les activités du CERN tout le temps et de n'importe quel endroit est une nécessité pour ses membres et un avantage indéniable pour la communauté scientifique mondiale. Si vous êtes intéressés, allez consulter le bulletin hebdomadaire du CERN en ligne à l'adresse suivante: ***http://bulletin.cern.ch***, et si à l'instant vous n'avez pas d'ordinateur à disposition, alors peut-être pouvez-vous consulter le bulletin à l'aide de votre agent mobile à l'adresse suivante: ***http://www.eif.ch/cern***.

fig. 1 – site principal du Bulletin

fig. 2 – la catégorie des news

Quoi de neuf?

L'application de publication **Bulletin2PDA**[1] est le résultat d'une collaboration entre le groupe *CERN Document Server* et le *Mobile Information System Laboratory* de l'EIA-FR.

Dans cet article, l'accent est mis sur la face cachée de la scène pour vous expliquer les deux phases les plus importantes dans la réalisation d'une application mobile. Il s'agit d'une approche très connue dans le monde de systèmes d'information et qui devient de plus en plus utile:

- L'intégration de plusieurs sources d'information en se basant sur un format d'échange neutre et standardisé.
- La diffusion de l'information sur des dispositifs mobiles en tenant compte de différents profils des utilisateurs.

Le serveur du CDS *(CERN Document Server)*

Le serveur de documents du CERN[2] est le tout dernier maillon de la recherche fondamentale entreprise dans le plus grand centre mondial de la physique des particules. Tous les résultats des travaux issus de théories ou d'expériences y sont, en effet, rendus publics et archivés électroniquement à long terme (après bien sûr que les données brutes recueillies par les détecteurs aient passé les nombreux filtres d'acquisition, d'analyse, de tri, et de réflexion !). Ainsi, plus d'un millier de documents alimente chaque semaine la base de données de CDS... et la curiosité insatiable des physiciens du monde entier.

A coté de cette large quantité d'articles extrêmement spécialisés, CDS est devenu au fil des années Internet le lieu de stockage et de consultation de toute sorte de documentation destinée au long terme. Petit à petit ont émergé des collections de photos, de vidéos, de posters, d'articles grand public, d'objets de musée ou encore d'agendas de réunions. Près de 50 000 *hits* sont enregistrés quotidiennement.

Principalement deux logiciels développés par l'équipe CDS ont permis la propagation de collections (450 à ce jour) et l'adhésion de l'ensemble du CERN pour utiliser ce système: un logiciel de gestion de bibliothèque digitale (***http://cdsware.cern.ch***) et un logiciel de gestion d'agendas de réunions ou de conférences ***(http://cdsware.cern.ch/cdsagenda)***.

Le Bulletin du CERN

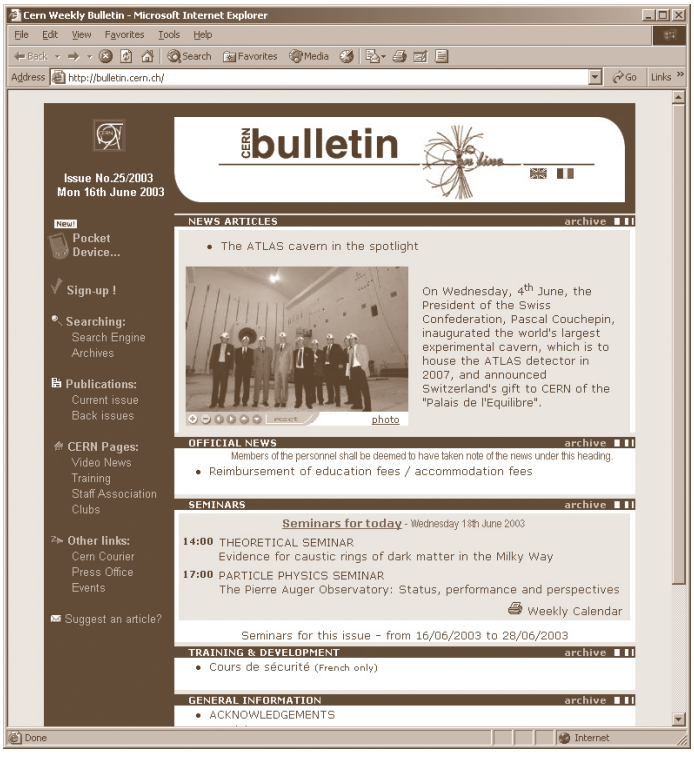

fig. 3 – le bulletin du CERN

Au CERN, un bulletin d'information hebdomadaire est diffusé depuis plus de vingt ans à tous les membres et ex-membres de l'organisation. Ce bulletin fut longtemps distribué sous forme papier uniquement. Cependant, dès 1997, la plupart des articles ont commencé à être mis en ligne et en octobre 2000 ils sont entrés dans une nouvelle collection du Serveur de Documents, s'organisant ainsi dans une base de données standardisée et rendant la recherche plus facile. Des articles de fond aux informations officielles en passant par les activités culturelles et sociales, l'intégralité du bulletin est devenue morceau par morceau électronique. Les efforts *politiques* pour alors mettre en place un bulletin électronique (avec la perspective d'une suppression du bulletin papier coûteux à imprimer et diffuser) ont finalement vu le lancement de ***http://bulletin.cern.ch***, avec un système d'abonnement électronique permettant aux abonnés de ne plus dépendre de la version papier.

Avec 700 abonnés et près de 170 visiteurs différents par jour (***http://bulletin.cern.ch/webstat***), le site a vite prouvé son utilité mais une enquête d'opinion à l'intérieur de l'organisation a également révélé qu'une grande partie des lecteurs (***http://cdsweb.cern.ch/search.py?recid=45917***) restaient néanmoins tributaires du journal dans son format traditionnel: papier A4, facile à feuilleter... et à transporter !

L'idée d'un format électronique portable était alors lancée... et sa réalisation a vu le jour à travers la collaboration entre le CERN et l'EIA-FR.

Intégration des données

Les articles du bulletin sont donc stockés dans la base de données documentaire du CERN, qui contient 600'000 entrées bibliographiques dont 250'000 documents complets, soit une énorme quantité d'information. L'application **Bulletin2PDA** met en évidence les techniques permettant de mettre à disposition une telle quantité riche de données pour un grand nombre d'utilisateurs. La base de données documentaire supporte via un module d'extension la conversion des données en format d'échange XML *(eXtensible Markup Language)*. Une autre source importante pour le bulletin est la base de données des agendas. Il s'agit d'un *calendrier* contenant tous les événements, les séminaires, les cours, etc. liés au CERN. Pour mieux pouvoir adapter des changements éventuels dans cette base de données, nous avons mis à disposition un module de transformation de données basé sur XLE *(XML Lightweight Extractor)*[3] (cf. plus bas) qui permet de définir de manière abstraite la conversion de données relatives au format XML.

Les différentes données sont fusionnées dans un modèle XML propre au bulletin. Un module s'occupe de l'intégration complète et examine les résultats obtenus, afin de garantir la persistance des données. Ainsi, un fichier contenant toute l'information est créée et sauvegardé pour chaque numéro de bulletin.

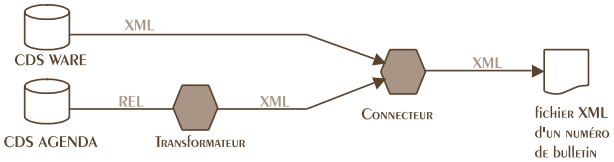

fig. 4 – intégration des données

Personnalisation de l'information

Le fichier XML ainsi créé est très riche. Il contient les articles de toutes les catégories du bulletin en deux langues (anglais/français), les informations sur les différentes images et les autres indications qui peuvent être utilisées pour des buts d'archivage, les tables de matières et les extensions futures.

Etant données les limites en capacité de mémoire de la plupart des PDA, il apparaît logique d'ordonner l'information entière en la réduisant au strict nécessaire et de pouvoir l'adapter au besoin individuel de l'utilisateur.

Différentes possibilités sont offertes à celui-ci pour personnaliser son abonnement au bulletin: le choix du numéro de bulletin, le choix de la langue, le choix d'afficher des images, et le choix des rubriques à afficher.

Les paramètres choisis sont stockés sur le PDA (ou en utilisant le service AvantGo[4] sur un serveur AvantGo) via le principe des *Cookies*. Cette approche revêt un rôle beaucoup plus important au niveau des PDA en comparaison avec des ordinateurs classiques du fait de l'aspect personnel et privé de son utilisation.

fig. 5 – un article de news (avec image)

fig. 6 – un article de news (sans image)

La personnalisation des données est réalisée par un système de filtre modulaire. Celui-ci permet, par indication de différents paramètres venant du profil de l'utilisateur, de filtrer les informations complètes *on-the-fly* et même de prendre en considération les informations sur les capacités du dispositif qui se connecte au système. On arrive ainsi à un fichier XML personnalisé qui a une taille nettement inférieure au fichier XML initial.

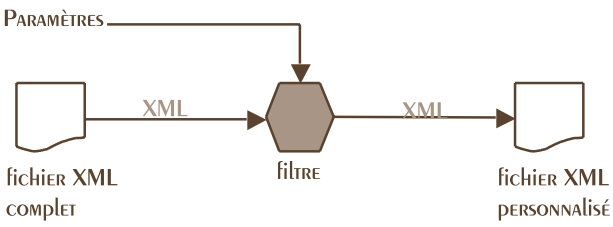

fig. 7 – Module de filtre pour personnaliser l'information

Publication de l'information

Pour la publication de l'information, les possibilités et les limites du média de destination doivent être connues. C'est pourquoi **Bulletin2PDA** intègre un système de reconnaissance de dispositif en se basant sur les headers HTTP. Aussitôt que les caractéristiques d'un appareil sont connues, les paramètres du système sont adaptés et le XSLT-Stylesheet *(eXtensible Stylesheet Language Transformations)* approprié est choisi visant le formatage du fichier XML personnalisé. Ces feuilles ne gèrent pas uniquement le formatage des données, mais elles mettent aussi en place un système de navigation adaptée au dispositif. Pour cela, un dictionnaire avec la langue demandée est fusionné avec le fichier XML personnalisé avant la transformation.

Plus rien ne fait obstacle au formatage du bulletin du CERN dans différents formats. Actuellement les formats suivants sont supportés: HTML normal, HTML pour PDA, HTML pour AvantGo. Il devient désormais envisageable de créer une sortie PDF qui pourrait remplacer la version imprimée du bulletin actuel du CERN, et de mettre en place une sortie spécifique pour les navigateurs WAP, et finalement d'ajouter une sortie SMS et MMS.

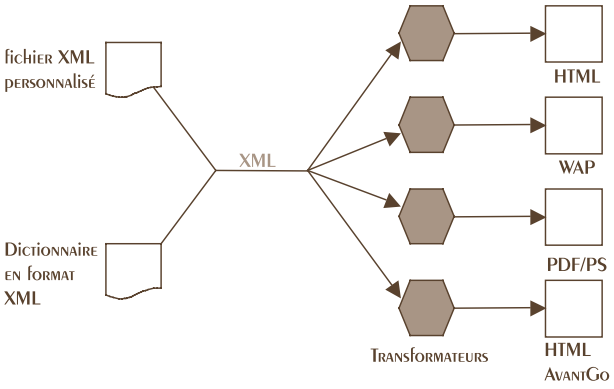

fig. 8 – publication de l'information en différents formats

Techniques employées

Chaînage des transformations XML/XSL

Toute transformation de données décrite (avec l'exception du transformateur **Relationnel/XML**) se passe en utilisant des feuilles de styles XSLT et le parseur **Xalan**[5]. Celui-ci offre le raccordement de plusieurs transformations ce qui influence le niveau de performance. La liaison du module de filtre et du module de publication en utilisant le principe de chaînage permet de débuter la deuxième transformation dès que la première commence à livrer des résultats, même si celle-ci est encore loin d'avoir fini tout le traitement.

Pré-compilation des feuilles de style

La pré-compilation des feuilles de style est une technique qui permet de générer des templates de transformateurs

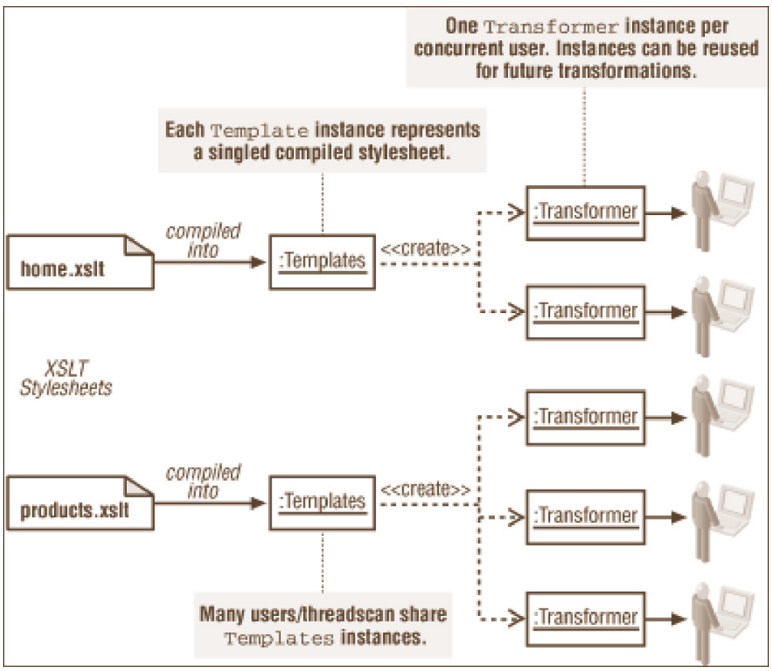

fig. 9 – templates de feuilles de style

XML/XSL en parsant les feuilles de style au préalable et en les stockant dans la mémoire. Ainsi, dès qu'il y a une demande de transformation, un nouveau transformateur est généré à partir du chablon sans avoir besoin d'accéder à la feuille de style sur le disque. Cette technique accélère considérablement le traitement..

XLE (XML Lightweight Extractor)

Dans beaucoup d'applications, il est nécessaire de générer des documents XML à partir des sources existantes de façon qu'ils se conforment à des DTD *(Document Type Definition)* ou des XML Schemas. Pour ce faire, XLE propose un langage permettant d'exprimer la structure logique de chaque document XML cible depuis diverses tables d'une base de données.

Conclusion

Le projet **Bulletin2PDA** montre la possibilité d'intégrer plusieurs sources d'information provenant de plusieurs bases de données en les traitant et en les fusionnant dans une source cible avec un format neutre et standard. Par la suite, cette source sera traitée pour personnaliser la publication en la diffusant sous différents formats. Pour plus d'informations sur ce projet, visitez l'adresse suivante: ***http://www.eif.ch/cern/info***

Si vous avez des questions ou des remarques, n'hésitez pas à nous contacter. ■

1 *http://www.eif.ch/cern/info*
2 *http://cds.cern.ch*
3 *http://www.alphaworks.ibm.com/tech/xle*
4 *http://www.avantgo.com*
5 *http://xml.apache.org/xalan-j*