
PELLEAS :

la gestion des documents électroniques sur l'intranet du Polytechnicum de Marne-la-Vallée et de l'université de Reims

par Christian Lupovici*

PELLEAS est un projet de plate-forme technique qui vise à installer sur un campus universitaire un environnement de création, de gestion, de recherche et de distribution de documents électroniques multimédias. Pour le compte du Polytechnicum de Marne-la-Vallée, la bibliothèque de l'UMLV a soumissionné en 1997 à un appel à proposition du ministère de l'Éducation nationale ; elle a été retenue pour réaliser un projet pilote qui est contrôlé par un comité de pilotage tripartite (ministère, UMLV, personnalités extérieures).

L'université de Reims s'est depuis jointe au projet, dans le cadre d'un programme de coopération entre l'UMLV et l'URCA, et un commencement de réalisation va pouvoir avoir lieu grâce au financement de ce programme.

Le projet PELLEAS part des constatations suivantes :

- Les universités et les établissements d'enseignement supérieur sont inévitablement amenés à travailler dans un environnement électronique. Ils ont nécessairement besoin de construire une infrastructure d'édition électronique.
- Les établissements d'enseignement supérieur sont producteurs de documents, même quand il n'existe pas à proprement parler de presses universitaires. Ces documents sont très variés et vont de la revue scientifique au support de cours, de la thèse au didacticiel pour l'autoformation.
- L'offre de documents électroniques des éditeurs se fait plus consistante et remplacera bientôt l'offre papier.

• Contrairement à ce que les pratiques actuelles d'accès distant aux périodiques électroniques pourraient laisser croire, il est évident que, techniquement, l'exploitation de la documentation électronique ne peut se faire de façon efficace que par l'exploitation locale des documents les plus utiles et par un recours marginal à l'accès distant (comme on le fait aujourd'hui en partageant nos acquisitions entre abonnements et fourniture de documents).

• Les bibliothèques sont équipées de systèmes intégrés de gestion qui n'ont pas vocation à gérer des documents électroniques. Il est donc indispensable de construire un système de gestion des documents électroniques qui puisse s'interfacer avec les SIGB du marché, pourvu que ceux-ci utilisent un protocole de communication normalisé (Z 39.50).

L'enjeu du projet PELLEAS est de bâtir un système qui puisse tenir compte à la fois des documents achetés (en nombre) chez les éditeurs et de documents structurés produits localement, de documents textuels mais aussi de documents multimédias, et d'interfacer ce système avec les SIGB standards.

Production des documents

Les documents créés peuvent être saisis par les auteurs et reformatés par l'atelier d'édition électronique, ou obtenus par numérisation en mode image avec éventuellement, si l'on veut ensuite pouvoir mettre à jour le document ou effectuer une indexation automatique, une transformation en mode caractères (par OCR). Des documents multimédias autonomes ou destinés à être insérés dans des documents textuels sont actuellement créés par l'équipe multimédia sur une base expérimentale.

* Université de Marne-la-Vallée

Néanmoins, la plupart des documents proviennent de saisies effectuées sur des logiciels de traitement de texte type Word, par les enseignants ou les thésards. Le projet consiste donc à définir des « feuilles de style » utilisables avec un petit nombre de logiciels éditeurs standards et d'imposer des feuilles de style aux créateurs d'information. Ainsi, par exemple, une feuille de style est en cours de définition pour les mémoires de maîtrise, les DEA et les thèses. Le travail de l'atelier d'édition électronique consistera ensuite à reformater automatiquement les documents provenant de logiciels de traitement de texte, dans un format standard SGML/XML (pris comme format pivot de stockage et de conservation).

C'est également en XML que se fera principalement (mais non exclusivement) la diffusion sur le Web. Ce balisage doit correspondre aux normes en vigueur dans le monde de l'édition et de la documentation, comme la DTD (description de type de document) ISO 12083 et la TEI (Text Encoding Initiative) utilisée dans le milieu universitaire (aux États-Unis et au Canada) en lettres et en sciences humaines.

Chargement et stockage

Le chargement de documents peu nombreux mais quelquefois très volumineux (documents multimédias réalisés en interne), ou moins volumineux mais nombreux (les articles de périodiques achetés à Elsevier, Kluwer, Academic Press etc.), pose des problèmes spécifiques qui nécessitent la réalisation d'un chargeur puissant (donc rapide) avec une indexation de la base après chargement.

La gestion de la base des documents électroniques doit donc être autonome par rapport aux systèmes de gestion de bibliothèque, tout en offrant une interface commune pour que l'utilisateur ait à ignorer l'architecture du système d'information. Il est prévu un stockage relativement massif afin d'éviter la saturation des réseaux et des serveurs par le recours trop fréquent aux sources éloignées.

Il est prévu également que les documents soient stockés dans le format le plus riche possible en vue de garantir une conservation et une réexploitation future aux meilleures conditions techniques, tenant compte du fait que la distribution de l'information se fera en fonction de l'état de l'art des visualiseurs et des logiciels de mise en forme et d'impression du moment.

Visualisation et impression

L'environnement universitaire est un environnement « contrôlé », où le parc des postes clients est identifiable. Il est donc possible de distribuer à tous les postes de l'université (ou presque) les logiciels clients qui sont nécessaires pour interroger, visualiser et imprimer les documents. L'utilisateur ne doit rien avoir à spécifier pour voir et imprimer son document. L'interface Web est devenue un modèle normatif, et c'est dans cet environnement que sera réalisée l'interface utilisateur si elle ne se trouve pas déjà incorporée dans les outils standards du marché.

Les choix techniques de PELLEAS

Les universités de Reims et de Marne-la-Vallée se sont équipées de systèmes de gestion de bibliothèque aux normes internationales (enregistrements en format MARC et interface aux protocoles HTTP et Z 39.50) comme première étape d'un système d'information électronique de campus. Elles se sont également équipées de serveurs HP 9000 avec leur environnement d'accès et de stockage.

Le choix de l'architecture est la construction d'un système totalement indépendant du SIGB. Chacun des deux systèmes peut donc évoluer à son rythme. L'intégration fonctionnelle est effectuée à l'interrogation par une solution de type antéserveur qui lance une requête en parallèle sur les deux systèmes. Les métadonnées des documents électroniques ne sont pas redondantes avec les notices du SIGB réservées à la gestion des documents physiques (livres, revues, cédéroms, cassettes vidéo, DVD...).

Le choix du format pivot des documents électroniques est SGML/XML. En ce qui concerne le stockage, il faut choisir le format le plus riche possible et ne stocker les documents qu'une seule fois, en reformatant à la volée pour la distribution. L'interface utilisateur est en mode Web.

Un appel d'offres a été lancé en février 2000 pour une réalisation à partir du second semestre 2000. Dans cette attente, nous travaillons sur l'alimentation du système, en particulier sur la diffusion de feuilles de style pour la structuration des documents électroniques produits par l'université, les thèses au premier chef afin de pouvoir les gérer et les conserver en SGML/XML.