

Multivio: solution gratuite et performante pour visionner des documents numériques

Dans l'objectif d'améliorer la consultation d'objets numériques, RERO, le Réseau des bibliothèques de Suisse occidentale¹, a lancé en 2008 le projet Multivio². Le résultat de ce projet est une application web³ qui peut être couplée à des serveurs de documents en tant que couche de présentation de contenus numériques, fonctionnant directement à l'intérieur du navigateur.

Multivio : une solution open source

L'application est conçue comme un module indépendant mais adaptable au contexte dans lequel il s'intègre. Les résultats du projet sont disponibles publiquement en open source et peuvent être réutilisés par toute institution intéressée. Depuis début 2010, RERO l'exploite avec succès comme module de présentation des contenus de sa bibliothèque numérique RERO DOC¹.

Cet article commence par faire un état des limitations majeures touchant les méthodes traditionnelles d'accès aux documents numériques sur Internet. Il présente ensuite le fonctionnement de Multivio et sa manière de surmonter ces limitations, ainsi que les nombreux avantages que cette solution est en mesure d'offrir à toute institution mettant des contenus numériques à disposition sur Internet.

Les limitations des interfaces de consultation traditionnelles

A l'heure actuelle, la documentation disponible sur Internet existe soit directement sous forme de pages web au format HTML, «prêtes à consommer» dans l'environnement naturel du Web (le logiciel de navigation), soit sous forme de fichiers téléchargeables, dans la très grande majorité des cas en format PDF, pour ce qui est des contenus textuels. Lorsque les utilisateurs souhaitent consulter un fichier, ils/elles cliquent simplement sur le lien correspondant, ce qui entraîne son téléchargement. Ensuite, si le poste de consultation est ainsi configuré, le fichier s'ouvre automatiquement à l'aide d'un logiciel dédié, par exemple Adobe Reader® ou Aperçu (Mac OS), le premier se trouvant parfois sous forme de plugin externe rajouté au navigateur. Ce geste est devenu très courant et naturel, le même principe s'appliquant à d'autres types de fichiers que le PDF, chacun d'entre eux faisant appel à un outil approprié.

Toutefois, si les internautes, en tant que consommateurs de contenus, ont l'habitude d'accéder facilement à des documents disponibles sur la toile, de leur côté les fournisseurs de ces mêmes contenus ont des besoins spécifiques par rapport à la façon de les organiser et de les mettre à disposition. Pour un auteur (un chercheur, par exemple), il est relativement aisé de rajouter un lien sur sa page web personnelle menant à son article. Mais pour une institution telle qu'une haute école ou une bibliothèque, il s'agit de le faire pour des collections entières de document, qui doivent être disposées de façon systématique sur une plateforme dédiée, permettant aux internautes de rechercher et de naviguer aisément le long de ces contenus, ainsi que d'y accéder. Cela pose des défis particuliers, et l'approche consistant à simplement mettre les fichiers à disposition en téléchargement n'est pas optimale, pour les principales raisons décrites ci-après.

Les fichiers multiples

L'approche du téléchargement simple est relativement satisfaisante pour les *documents individuels*, composés d'un seul *fichier*. On distingue ici les termes document et fichier. Alors que ce dernier désigne spécifiquement un fichier informatique (PDF, JPEG, MPEG, TIFF etc.) le premier désigne plutôt une ressource documentaire, telle qu'un livre, une thèse ou un article, qui peut, elle, être composée de plusieurs fichiers. Mais elle présente des limites en présence de *documents structurés et composés de plusieurs fichiers*, comme des périodiques, collections, livres multivolumes et autres types de regroupements, obligeant l'utilisateur à télécharger séparément chaque fichier sans pouvoir les mettre en perspective.

L'obligation de télécharger le fichier complet

Un autre problème associé au téléchargement des fichiers se manifeste lorsque la taille du contenu à télécharger est considérable, tout particulièrement:

1. dans le cas où l'utilisateur ne possède pas une ligne Internet à haut débit (temps d'attente trop long);

2. lorsque les données téléchargées sont taxées au volume, comme c'est habituellement le cas pour les dispositifs mobiles (Smartphones) reliés à une connexion fournie par un opérateur de téléphonie. Il est courant d'avoir des fichiers PDF de 100 Mb et plus à télécharger, notamment lorsqu'il s'agit du résultat de numérisation de documents imprimés. De quoi faire exploser bon nombre de forfaits mensuels avec un ou deux téléchargements!

La suppression du terme et du résultat de recherche

Lorsque le document à consulter a été retrouvé à la suite d'une opération de recherche dans un catalogue ou moteur de recherche, l'expression recherchée devrait être présentée et mise en évidence à l'intérieur du document, à l'endroit où elle a été retrouvée, afin de permettre aux utilisateurs de s'y diriger directement. Or, avec l'approche du téléchargement simple, les utilisateurs sont obligés, une fois le document ouvert, de répéter la même opération de recherche, cette fois-ci dans le logiciel de consultation employé, et séparément pour chacun des fichiers qui composent le document.

Mode de fonctionnement de Multivio

Prenons un exemple pour illustrer le mode de fonctionnement de l'application: la bibliothèque numérique RERO DOC utilise Multivio comme outil de présentation de ses contenus. La liste de résultats de recherche sur RERO DOC présente, associée à chaque résultat, une vignette avec une miniature du contenu du document correspondant (celle de la première page, en général), offrant à l'utilisateur un repère visuel. Lorsque l'utilisateur clique sur l'une de ces vignettes, le contenu du document s'affiche immédiatement au centre de la fenêtre (il s'agit d'une opération instantanée).

Si l'utilisateur le souhaite, il/elle peut agrandir la zone d'affichage de Multivio pour disposer d'une place de consultation plus grande. Ou alors la fermer, ou cliquer sur un autre résultat pour accéder au contenu respectif. Dans les cas où le document est composé de plusieurs fichiers, la table des matières à l'intérieur de la fenêtre Multivio, relative à l'ensemble du document, le signale clairement, permettant à l'utilisateur d'y naviguer et de lancer des recherches textuelles sur tous les fichiers à la fois.

Contenus simples et composés

Chaque fois qu'elle est sollicitée, l'application Multivio reçoit un URL correspondant au document à présenter. Celui-ci peut être simple ou composé. Voici comment ces deux cas se distinguent.

Fichier isolé: c'est le cas le plus simple. Multivio convertit alors simplement le fichier (PDF, image, vidéo ou autre) en un format directement lisible par le navigateur (ex: un PDF est décomposé en images de façon optimale), analyse sa structure interne si elle est présente (table des matières) et extrait son contenu textuel pour permettre des opérations de recherche. Le contenu est ensuite présenté à l'utilisateur.

Document composé: au lieu d'un lien vers un fichier, Multivio reçoit un lien vers une fiche de métadonnées décrivant l'ensemble du document. La fiche doit alors contenir les URL des fichiers de contenu. Dans ce cas de figure, Multivio explore l'ensemble de ces URL, récupère les fichiers correspondants, fait les traitements nécessaires sur chacun d'entre eux (cf. ci-dessus), rassemble le tout selon une organisation logique et le présente à l'utilisateur de façon homogène.

Métadonnées comme point de départ

Les formats de métadonnées les plus couramment utilisés pour la description de documents sont: Dublin Core², MODS³ et MARC21⁴ (tous reconnus par Multivio), essentiellement des métadonnées bibliographiques mais qui peuvent contenir aussi des liens d'accès au contenu numérique, sous forme de liste. Par exemple, lorsqu'une thèse de doctorat est composée de fichiers différents pour le corps principal du document et pour les annexes, comme des données expérimentales ou des démonstrations multimédia, les métadonnées qui décrivent l'ensemble de la thèse contiendront la liste de fichiers en question.

Il existe aussi des formats de métadonnées structurales, dont METS⁵ est l'exemple le plus courant, qui permettent de décrire la structure interne d'un document et, à partir de points spécifiques de cette structure, pointer vers des fichiers de contenu numérique. Cette approche est couramment utilisée pour décrire des livres numérisés, en assemblant les images obtenues à partir de l'opération de scanning selon une structure logique.

Indépendamment de ce qu'il reçoit en entrée – contenu simple ou composé – Multivio explore la structure interne du document et détecte automatiquement le type de contenu de chacune des sous-parties, s'il n'est pas indiqué explicitement.

Fonctionnalités

Dans sa version 1.0, Multivio permet la consultation de documents basés sur les formats PDF et image (JPEG, GIF, PNG, TIFF, etc. – plus de 30 formats d'image en tout). Il offre aussi la possibilité de lancer des recherches textuelles (y compris multi-fichiers), avec surbrillance des résultats, ainsi que la sélection de texte (pour copier-coller). Il supporte la navigation de structures hiérarchiques, permet d'afficher une liste ou une grille avec les vignettes des pages et permet de faire zoom et rotation sur le contenu affiché.

Au cours de l'année 2011 sont en développement notamment le support de formats de contenu dynamique (audio, vidéo) et le contrôle d'accès à des documents protégés.

Les points forts de Multivio

L'application proposée dans le cadre de ce projet présente un nombre d'avantages qui en font une solution très séduisante:

- Elle offre un *support élégant de documents composés de plusieurs fichiers*, dont les contenus peuvent être consultés de façon unifiée.
- *Recherche à la volée*: lorsque Multivio est appelé à présenter un document, il est possible de lui fournir en même temps un terme de recherche textuelle. Les résultats de l'opération de recherche, à l'intérieur du document en question, et du terme fourni sont alors présentés au moment de son ouverture, sous forme de liste et aussi avec surbrillance du terme recherché dans le texte de la page.
- *Légèreté et rapidité*: l'application est très rapide et légère, capable de servir facilement des documents de très grande taille (> 1 GB), même avec des vitesses de connexion limitées. Le transfert de données est minimal.
- Elle est basée sur des *technologies standards* du Web: HTML + JavaScript + CSS. Cela signifie qu'elle est indépendante de la configuration du poste client. Le seul navigateur web lui suffit, ce qui *évite l'installation de plugins dédiés* et permet son fonctionnement sur des dispositifs mobiles.
- Elle *s'intègre aisément* dans toute bibliothèque numérique ou serveur de documents.
- Elle supporte de *nombreux formats de fichiers et de métadonnées*, étant conçue pour pouvoir facilement s'adapter à de nouveaux formats.

– *Faible encombrement*: selon besoin, les données à consulter peuvent faire l’objet d’un prétraitement par la couche logicielle serveur, de façon très optimisée. Cette opération est effectuée dans tous les cas à la volée, au moment où les contenus sont demandés, *rendant ainsi inutile toute mise en place de copies des fichiers* spécialement dédiées à la présentation sur Internet.

– *Gratuité*: le code source de Multivio est disponible à l’adresse <https://www.multivio.org/> sous licence open source GPL⁶. La participation communautaire à l’extension de l’application sera la bienvenue. A cette même adresse se trouvent les instructions permettant d’installer Multivio localement.

– *Modularité et extensibilité*: l’architecture de Multivio a été conçue dès le départ pour être ouverte et modulaire, avec l’objectif de permettre le plus facilement possible d’étendre ses fonctionnalités à l’avenir, sans bouleverser la structure de base de l’application. Sont notamment visés le support de nouveaux formats de fichiers de contenus et de types de métadonnées, ainsi que l’ajout de nouvelles composantes à l’interface utilisateur.

Le projet

Le projet a été cofinancé par RERO et par un programme d’impulsion de la Confédération, puisqu’il est l’un des 20 sous-projets «e-lib.ch: Bibliothèque électronique suisse⁷», projet de coopération et d’innovation de la Conférence universitaire suisse pour la période 2008–2012. A noter que l’EIA-FR, l’Ecole d’ingénieurs et d’architectes de Fribourg⁸, a aussi participé à une partie du développement, dans le cadre de son accord de partenariat avec RERO.

Dans le contexte d’e-lib.ch, le projet s’est achevé avec succès en avril 2011, ayant atteint pleinement ses objectifs. Actuellement, RERO poursuit son développement pour sa bibliothèque numérique RERO DOC, selon le modèle open source traditionnel.

Multivio n’est pas le seul produit existant dans ce domaine. Citons comme exemples répondant à des besoins un peu différents, DFGViewer⁹ et l’outil «Quick View» intégré par Google dans les résultats de son moteur de recherche, ainsi que plusieurs autres solutions propriétaires et fermées, intégrées dans des logiciels commerciaux de gestion de documents numériques. Multivio innove et se distingue de ces solutions en offrant une combinaison de fonctionnalités et de caractéristiques unique, non seulement par sa disposition naturelle à traiter aisément des contenus composés de fichiers multiples, aspect véritablement novateur, mais aussi par sa performance et sa flexibilité.

Serveur Multivio public

RERO met à disposition depuis avril 2010 un serveur de démonstration de Multivio à usage public, à l'adresse <http://demo.multivio.org>. Ce démonstrateur permet à chacun de tester Multivio avec ses propres documents, soit manuellement à l'aide d'un petit formulaire acceptant un URL, soit directement depuis son propre catalogue ou serveur de documents, à l'aide d'un URL paramétrable qui actionne l'appel à Multivio.

L'utilisation du serveur public est entièrement libre, avec comme seules réserves l'absence de garantie de support et de disponibilité du service, ainsi qu'une limitation dans la taille des documents admis. Les institutions intéressées peuvent installer et paramétrer leur propre instance, si elles le souhaitent. Quoi qu'il en soit, cette instance publique leur offre désormais une solution d'interface de présentation intéressante et performante pour leur serveur de documents.

1 <http://doc.rero.ch/>.

2 <http://dublincore.org>.

3 <http://www.loc.gov/standards/m...>

4 <http://www.loc.gov/marc/>.

5 <http://www.loc.gov/standards/m...>

6 <http://www.gnu.org/licenses/gp...>

7 <http://e-lib.ch/>.

8 <http://eia-fr.ch/>.

9 <http://dfg-viewer.de/>.



Miguel Moreira

Chef de projet RERO

Résumé

Deutsch

Multivio ist ein Teilprojekt von e-lib.ch und bietet eine verbesserte Open-Source-Lösung für die Ansicht digitaler Inhalte am Bildschirm. Die Anwendung, die aus diesem Projekt resultiert, ist als ein unabhängiges Software-Modul gedacht, das mit Dokumentenservern und digitalen Bibliotheken gekoppelt werden kann. Es zielt darauf ab, den Benutzern den direkten und unmittelbaren Kontakt mit den ihnen zur Verfügung stehenden digitalen Inhalten zu ermöglichen und dadurch nicht nur den Benutzerkomfort für Erstere, sondern auch die Auswertung Letzterer zu verbessern.

Multivio zeichnet sich gegenüber der einfachen Online-Stellung zum Download und auch gegenüber anderen vergleichbaren Produkten durch zahlreiche Vorteile aus. Es berücksichtigt nicht nur einfache Dateien (was der häufigste Fall ist), sondern auch und vor allem strukturierte Dokumente, die in mehreren Dateien gespeichert sind (was demgegenüber der interessantere Fall ist). Die Benutzerfreundlichkeit, die Schnelligkeit und die Leichtigkeit der Bedienung sowie die Tatsache, dass es sich um eine gratis zur Verfügung stehende Software handelt (Open Source), machen Multivio zu einem sehr interessanten Werkzeug.