

	JSTOR/Harvard Object Validation Environment (JHOVE) 1.5	
	File Format Identification and Validation Tool	
	System Requirements: JRE version 1.5.0 ou plus récente Minimum 512 MB RAM recommandé	
	Développé par	JSTOR/Harvard University Library http://hul.harvard.edu/jhove/index.html
Licence GPL	Personne à contacter	Arnaud.Hulstaert@smals.be

Fonctions

Dans le cadre de l'archivage à long terme d'objets numériques, il est nécessaire pour l'organisation en charge de la conservation de disposer d'une connaissance approfondie des formats de fichiers et de leurs propriétés pour qu'elle puisse déterminer au mieux comment les conserver.

A cette fin, la plate-forme d'archivage doit :

1. identifier le format de l'objet à archiver (*format identification*)
2. en vérifier la conformité avec la norme ou le standard (*format validation*) afin de s'assurer d'archiver un objet valide et exploitable selon les possibilités du format
3. générer des informations sur l'objet afin que le système (et ses responsables) puisse valider son entrée dans le système et le gérer (métadonnées).

Développé par JSTOR et la bibliothèque universitaire d'Harvard, Jhove exploite les informations stockées dans le *Global Digital Format Registry*, un répertoire d'informations sur les formats les plus couramment utilisés dans le contexte de l'archivage. Ce logiciel open source écrit en JAVA permet d'effectuer les trois étapes citées ci-dessus.

Le logiciel peut être utilisé via une GUI, des lignes de commandes ou des API. Il est possible d'exporter les résultats sous forme de fichier TXT ou XML. Ce fichier peut ensuite être archivé avec l'objet numérique et les informations générées réutilisées comme métadonnées.

La version 1.5 de Jhove permet de gérer les formats suivants (ainsi que certaines variantes de ces formats) : AIFF, ASCII, GIF, HTML, JPEG, JPEG 2000, PDF (1.0-1.6, /X, /A), TIFF (4.0, 5.0, 6.0, IT, EP, GeoTIFF, ...), UTF-8, WAVE et XML. Il couvre donc un large éventail des besoins en matière d'archivage.

La version 2.0 de Jhove est actuellement en cours de développement, e.a. pour en réviser l'architecture afin d'en faciliter l'intégration dans d'autres systèmes et permettre l'export des métadonnées selon les différents standards actuels en matière de préservation et d'archivage (PREMIS, Dublin Core, EAD, ...).

Conclusions et Recommandations

Jhove est un outil open source en JAVA qui permet d'identifier et de valider le format d'un objet numérique en vue de son archivage (ce qui explique que les formats gérés – 11 formats génériques au total – relèvent du domaine de l'archivage). Cette validation est nécessaire pour s'assurer que l'objet archivé est valide, ce qui en facilite la gestion et la préservation.

Les tests effectués indiquent que la validation de Jhove est fiable. Les informations générées en output sont utiles pour préserver l'objet puisqu'il s'agit de métadonnées techniques générées automatiquement. L'outil est extrêmement simple et peut être utilisé via une GUI, des API's et des lignes de commande, ce qui en facilite l'intégration avec des plates-formes d'archivage.

Tests et Résultats



Deux tests ont été effectués en se basant sur les deux formats les plus couramment utilisés pour l'archivage à long terme d'objets numériques : le format PDF-PDF/A et le format TIFF.

L'interface GUI de Jhove est extrêmement simple et intuitive. Pour analyser le fichier, il suffit de le sélectionner ou d'indiquer l'URL de l'objet.

Dans le cas du document PDF/A-1a, nous avons utilisé un document paramétrisé comme PDF/A-1a mais ne respectant pas la norme (absence de tags de structure alors qu'ils sont requis afin que le document puisse être affiché correctement sur des gsm, PDA, ...). Jhove a correctement détecté que le document n'était pas conforme à la norme 1a mais uniquement à la norme 1b (qui ne nécessite pas les tags).

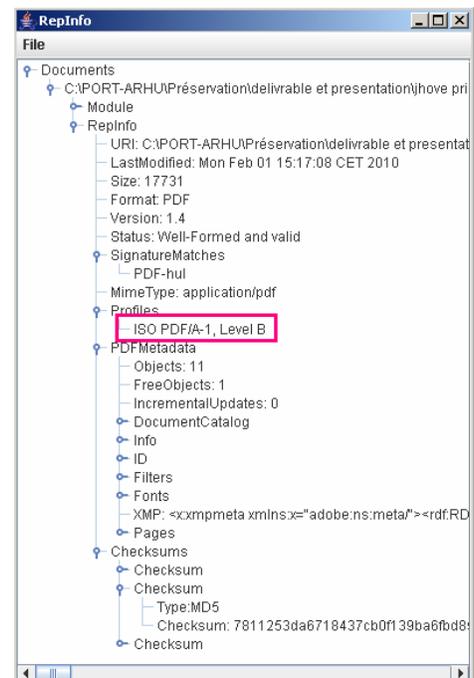
Nous avons ensuite effectué un test avec des fichiers TIFF 5.0 et 6.0 conformes à la norme et un test avec un fichier non-conforme (balise obligatoire corrompue). Après analyse, les fichiers conformes ont été validés par Jhove, ce qui ne fut pas le cas du fichier non conforme.

La validation des fichiers par Jhove est donc fiable.

Jhove peut également être exécuté en ligne de commande. Nous l'avons donc exécuté sur un dossier de 4000 images TIFF d'environ 3.5 Mo chacune, provenant de la numérisation de microfilms. Notre objectif était de tester la conformité des fichiers par rapport à la norme 5.0 ou 6.0 du format TIFF, cette exigence étant mentionnée dans le cahier des charges. Nous avons donc lancé la requête suivante « `jhove -c conf/jhove.conf -k "C:\xxx" -o test_3.txt` »

L'output final délivré sous forme de texte (dans ce cas-ci le fichier texte 'pèse' 75 Mo) ou sous forme XML permet de réutiliser les informations générées pour documenter la procédure de validation mais également comme métadonnées de gestion et de préservation. Jhove a ainsi récupéré le texte contenu dans la balise `ImageDescription` qui n'est pas une balise obligatoire. Par contre, il ne récupère pas les balises privées (*private tag*) contenue dans les fichiers et permises par la norme TIFF.

Le logiciel génère également sur demande les checksums de l'objet selon trois algorithmes différents : CRC32, MD5, SHA-1. Ces checksums sont surtout utiles ici pour vérifier la qualité de la transmission davantage que pour des raisons cryptographiques.



Budget

Aucun, l'outil Jhove est gratuit.