



Spécifications de l'utilitaire mk- doc

Logilab S.A.
10, rue Louis Vicat
F-75015 PARIS
Tél. : (+33) 1 45 32 03 12
Web : www.logilab.fr
Adel : contact@logilab.fr

Olivier CAYROL

Spécifications de l'utilitaire mkdoc

Olivier CAYROL

Spécifications de l'utilitaire mkdodoc

1. Objet

L'utilitaire mkdodoc permet à partir d'un fichier au format DocBook XML ou ReSt d'obtenir un fichier au format DocBook XML, au format PDF, ou au format HTML.

2. Besoins

2.1. Aperçu général

Les transformations du XML dans le format de sortie (PDF ou HTML), sont faites en utilisant les feuilles de style XSLT de LOGILAB. Il existe différentes feuilles de style suivant le résultat souhaité. Ces feuilles de style se trouvent dans des sous-répertoires distincts :

- `fo/standard` : transformation standard pour PDF
- `fo/reponse-ao` : transformation pour PDF pour les réponses aux appels d'offre
- `fo/rapport-admin` : transformation pour PDF pour les rapports administratifs
- `fo/publi-technique` : transformation pour PDF pour les publications techniques
- `fo/publi-commerciale` : transformation pour PDF pour les publications commerciales
- `fo/lettre` : transformation pour PDF pour les lettres (correspondance)
- `fo/contrat` : transformation pour PDF pour les contrats
- `fo/catalogue-formation` : transformation pour PDF pour le catalogue de formations
- `html/standard` : transformation standard pour HTML (balises simples)
- `html/single-file` : transformation pour HTML créant un seul fichier
- `html/multi-files` : transformation pour HTML créant plusieurs fichiers
- `html/web-site` : transformation pour HTML créant plusieurs fichiers composant un site Web

Dans chacun des sous-répertoires, la feuille de style se trouve dans le fichier `root.xml` (e.g. `fo/lettre/root.xml`).

Il est nécessaire d'indiquer à l'utilitaire quelle feuille de style il doit utiliser pour effectuer la mise en forme. Afin d'éviter de fastidieuses copies, seul le nom du sous-répertoire sera spécifié (e.g. `web-site` ou `rapport-admin`).

D'autre part, il arrive de plus en plus souvent qu'il faille faire un pré-traitement sur le fichier XML pour obtenir le DocBook qu'on veut transformer. Les pré-traitements sont effectués par une transformation XSLT se trouvant dans des sous-répertoires de `pre-process/` (e.g. `pre-process/session2prg-cours`). Il est donc, là encore, nécessaire d'indiquer à l'utilitaire quelle transformation de pré-traitement on souhaite appliquer. Cette indication se fait en spécifiant le nom du sous-répertoire (e.g. `session2prg-cours`).

2.2. Indication de la feuille de style à utiliser

Afin de pouvoir conserver une grande simplicité d'utilisation de l'utilitaire *mkdoc* (e.g. *mkdoc mon-fichier.xml*), les feuilles de style à utiliser seront indiquées par des instructions de traitement directement dans le fichier XML.

Il existe deux types d'instructions de traitement :

- les instructions de traitement spécifiant la feuille de style à utiliser pour mettre en forme le DocBook XML. La cible de ces instructions est *logidoc-style* et les données sont composées d'un champ *target* ayant pour valeur *pdf* ou *html*, et d'un champ *xslt* contenant le nom de la feuille de style XSLT.
- les instructions de traitement spécifiant la feuille de style à utiliser pour pré-traiter le XML en vue d'obtenir du DocBook XML. La cible de ces instructions est *logidoc-preprocess* et les données sont composées d'un unique champ *xslt* contenant le nom de la feuille de style XSLT.

Quelques exemples d'instructions de traitement :

- `<?logidoc-style target="pdf" xslt="reponse-ao"?>`
- `<?logidoc-style target="html" xslt="multi-files"?>`
- `<?logidoc-preprocess xslt="session2prg-cours"?>`

Lorsque, dans un fichier XML, il y a plusieurs instructions de traitement *logidoc-style* avec la même valeur pour le champ *target*, c'est la première qui est utilisée.

Lorsque, dans un fichier XML, il y a plusieurs instructions de traitement *logidoc-preprocess*, elles sont exécutées les unes après les autres dans l'ordre du document.

Note

Remarque de mise en oeuvre

Un simple analyseur SAX est suffisant pour récupérer les instructions de traitement. L'analyse SAX du document XML d'entrée permet de vérifier qu'il est bien formé, voire de le valider si besoin est.

2.3. Indication du format de sortie

Plusieurs formats de sortie sont possibles : XML DocBook, PDF, ou HTML. Le format de sortie par défaut est PDF. Un autre format de sortie peut être spécifié sur la ligne de commande de *mkdoc*.

Partant d'un fichier XML, l'utilitaire effectue les opérations suivantes afin d'obtenir le format de sortie demandé :

- **XML DocBook** : les instructions de traitement spécifiant les pré-traitements (*logidoc-preprocess*) sont exécutées dans l'ordre du document.
- **PDF** : après avoir effectué les étapes concernant XML DocBook, une transformation du XML DocBook en XSL-FO est faite à l'aide de la feuille de style spécifiée dans l'instruction de traitement *logidoc-preprocess* dont la cible (*target*) vaut *pdf*. S'il n'y a une telle instruction de traitement, c'est la feuille de style *standard* qui est utilisée. Le XSL-FO ainsi obtenu est transformé en PDF à l'aide d'un formateur FO.
- **HTML** : après avoir effectué les étapes concernant XML DocBook, une transformation du XML DocBook en HTML est faite à l'aide de la feuille de style spécifiée dans l'instruction de traitement *logidoc-preprocess* dont la cible (*target*) vaut *html*. S'il n'y a une telle instruction de traitement, c'est la feuille de style *standard* qui est utilisée.

Lorsque le fichier de départ est au format ReSt, un paramètre particulier est passé à l'utilitaire sur sa ligne de commande. L'utilitaire va alors transformé le fichier initial en XML DocBook avant d'effectuer les opérations décrites ci-dessus lorsque le fichier de départ est du XML.

2.4. Utilitaire en ligne de commande

2.4.1. Comportement standard

En standard, l'utilitaire (*mkdoc*) transforme un fichier XML au format PDF. Les opérations effectuées pour ce faire sont décrites ci-dessus.

L'utilitaire utilise le processeur XSLT *xsltproc* et le formateur FO *fop*.

2.4.2. Options

Il est possible de spécifier un certain nombre d'options en ligne de commandes :

- `-t format` ou `--target format` pour spécifier le format de sortie (pdf, html, dcbk),
- `-r` ou `--rest` pour spécifier que le fichier d'entrée est au format ReSt,
- `-v` ou `--validate-xml` pour demander la validation du XML d'entrée (la validation s'effectue lors du chargement du fichier pour récupérer les instructions de traitement),
- `-i` ou `--ignore-pi` pour demander à l'utilitaire d'ignorer les instructions de traitement se trouvant dans le XML d'entrée,
- `-s name` ou `--stylesheet name` pour spécifier à l'utilitaire la XSLT de mise en forme à utiliser (l'utilitaire continue toutefois à effectuer les pré-traitements),
- `-p name value` ou `--param name value` pour passer des valeurs particulières aux paramètres de la XSLT de mise en forme (ce mécanisme n'existe pas pour les XSLT de pré-traitement),
- `-k` ou `--keep-fo` pour demander à l'utilitaire de ne pas effacer le fichier temporaire au format FO (valable uniquement pour le PDF),
- `-h` ou `--help` pour demander à l'utilitaire d'afficher un message d'aide.

2.4.3. Configuration

L'utilitaire est configuré en donnant :

- le chemin de l'exécutable *xsltproc*,
- le chemin de l'exécutable *fop*,
- le répertoire contenant les feuilles de style LOGILAB (e.g. `/home/logilab/lib/xslt`).