

Génération de documents Word à partir de R : utilisation du package R2DOCX dans une plateforme statistique en milieu industriel.

David Gohel ^a

^a Consultant R

Lysis Consultants

david.gohel@lysis-consultants.fr

Mots clefs : génération de document MS docx, R.

L'automatisation du reporting est un point d'optimisation essentiel dans le milieu industriel, à la fois pour diminuer le temps considérable consacré à l'édition de rapports mais également pour tendre vers une recherche reproductible et assurer la qualité et l'homogénéité des analyses. Il est de plus incontournable, pour assurer la communication des résultats au sein de l'entreprise, de présenter des analyses statistiques éditables, avec des mises en forme spécifiques.

Aujourd'hui plusieurs solutions de reporting éditable existent via l'utilisation de R. Cependant, elles ne sont pas pleinement satisfaisantes : difficulté d'intégration dans un contexte industrialisé, mises en forme parfois limitées, difficultés d'intégration des modèles de document (template Microsoft Word).

Afin de répondre au besoin d'un client, la société Lysis a développé le package R 'R2DOCX' qui permet la génération de document MS Word à partir de R dans un contexte industrialisé. Ce package s'appuie sur la bibliothèque open source java docx4j[1] qui offre la possibilité de créer et de manipuler des documents au format Microsoft Open XML[2]. Aucun autre composant logiciel n'est requis.

La première version de R2DOCX offre la possibilité d'automatiser la production de tableaux, de graphiques, de textes et de tables des matières. Le document final peut être personnalisé grâce à l'utilisation d'un modèle défini par l'utilisateur. Le document final est réalisé via l'ajout successif d'éléments à éditer, produits par R (tables, graphiques, textes) à partir de ce modèle et des styles prédéfinis.

La mise en forme et le contenu des tableaux sont facilement configurables par l'utilisateur. Il est possible de modifier les entêtes, de fusionner des cellules, de spécifier les formats des contenus (numériques, pourcentages, etc.) et des cellules (bordures, espace entre le contenu et les bordures, etc.).

L'insertion de paragraphes, qui permet d'ajouter du texte, permet également la mise en forme conditionnelle de sous-paragraphes. Finalement, il est également possible de remplacer des mots clefs dans le document modèle, comme par exemple le nom de l'auteur, le titre, la date etc.

L'ensemble de ces fonctionnalités sera illustré au travers d'exemples simples d'analyses descriptives.

Références

[1] docx4j. <http://www.docx4java.org/>

[2] Standard ECMA-376 Office Open XML File Formats. <http://www.ecma-international.org/publications/standards/Ecma-376.htm>