

Insolite : Abbyy, L'héritage imprimé européen en numérique

Insolite

Posté par : JulieM

Publié le : 8/6/2011 15:00:00

ABBYY met son logiciel de **reconnaissance optique des caractères** (OCR) et son expertise dans la reconnaissance des polices de caractères historiques à la disposition du **projet de recherche européen** dont l'objectif est de transformer l'héritage imprimé européen en ressources disponibles sous forme numérique.

Riche de l'expérience acquise dans le cadre de ce projet quadriennal, ABBYY propose désormais des produits pourvus de la technologie de reconnaissance de texte la plus avancée pour les documents historiques imprimés.



ABBYY, un fournisseur leader dans la reconnaissance de documents, la capture de documents et les technologies et services linguistiques, annonce les premiers résultats de sa participation au projet IMPACT (IMProving ACcess to Text) de la Commission européenne.

Le consortium IMPACT réunit vingt-six bibliothèques nationales et régionales européennes, des instituts de recherche et des partenaires technologiques qui partagent leurs connaissances, leurs meilleures pratiques et développent des outils innovants destinés à améliorer les capacités des moteurs OCR. Dès la fin de l'année 2011, le projet améliorera considérablement l'accès aux textes historiques produits avant 1900. Les efforts de recherche et développement communs ont permis de franchir une étape importante dans la numérisation de l'héritage culturel européen.

Depuis 2008, ABBYY joue un rôle majeur dans le projet IMPACT en fournissant sa technologie OCR de pointe et son expertise pour numériser les polices de caractères historiques et extraire le texte des documents anciens. La technologie OCR d'ABBYY est idéale pour reconnaître du texte sur des images de documents imprimés avec différentes polices de caractères historiques. Il est ainsi possible de numériser l'intégralité de ces ressources, d'y effectuer des recherches et de les intégrer à des systèmes utilisés pour la création de lexiques. Les équipes de recherche et de développement d'ABBYY travaillent en étroite collaboration avec des membres majeurs des équipes du projet IMPACT afin de gérer les principaux problèmes technologiques associés aux grands projets de numérisation de bibliothèques.

S'appuyant sur des échantillons historiques pertinents collectés par de grandes bibliothèques européennes, ABBYY a réalisé des innovations technologiques importantes dans le domaine du pré-traitement des images et de l'analyse de la mise en page des documents afin d'obtenir de meilleurs résultats dans la reconnaissance des caractères.

« Auparavant, les logiciels OCR ne disposaient pas des fonctions avancées nécessaires à une transformation satisfaisante des pages scannées en texte intégral ou XML, en particulier lorsqu'il s'agissait de livres, de magazines ou de journaux anciens », explique **Aly Conteh**, membre du Comité de direction du projet IMPACT et directeur du programme de numérisation à la British Library. « En participant au projet IMPACT, nous avons pu accéder à la technologie OCR innovante d'ABBY qui permet aux membres du projet IMPACT de traiter plus efficacement des documents dont la qualité est souvent mauvaise. Nous savons que nous pouvons nous appuyer sur cette collaboration car ABBY travaille en étroite collaboration avec les bibliothèques et les autres partenaires en vue d'améliorer le courant technologique. »

« Non seulement la numérisation peut aider à préserver l'héritage intellectuel européen pour les futures générations, mais elle permet aussi de dévoiler ces trésors et les mettre à la disposition de millions de chercheurs, d'étudiants et de lecteurs à travers le monde », explique **Andrey Isaev**, directeur du département Produits SDK chez ABBY. « ABBY est très heureux de contribuer à cet objectif constructif par l'innovation OCR et le partage des meilleures pratiques. Sans compter que nous apprécions véritablement cette collaboration étroite et productive avec des personnes qui se passionnent tout autant que nous pour la reconnaissance de texte. »

La participation d'ABBY au projet IMPACT s'appuie sur les précédents projets de numérisation européens dans lesquels la société s'est impliquée avec succès, notamment le projet METAe, dans le cadre duquel ABBY a développé FineReader XIX, et conçu spécialement pour ce projet un OCR omnipolice pour reconnaître la police de caractères historiques Fraktur, ou Black Letter, que l'on retrouve typiquement sur des textes publiés entre 1800 et 1938, ou encore le projet Digitisation-on-Demand (Numérisation à la demande) dont le but est de numériser des millions d'ouvrages afin de les proposer au format électronique.

L'Institut Fraunhofer de communication multimedia (IMK) en Allemagne a également numérisé avec succès les archives du journal Neue Zürcher Zeitung à l'aide de la solution ABBY FineReader SDK avec fonction OCR pour polices de caractères historiques. Les dernières améliorations apportées aux technologies de reconnaissance ABBY sont désormais présentement disponibles sur le marché pour les instituts et les organismes du monde entier. ABBY FineReader Engine SDK et ABBY Recognition Server 3.0 intègrent plusieurs des dernières améliorations d'ABBY en matière de reconnaissance de textes et de documents spécialisés.

Vous trouverez de plus amples informations sur [le projet IMPACT](#).

Pour plus d'informations sur [la technologie OCR d'ABBY](#) pour la reconnaissance de textes historiques