

ARABASE : Base de données Web pour l'exploitation en reconnaissance optique de l'écriture Arabe

Noura Bouzrara*, Nacéra Madani Aissaoui**, Najoua Essoukri Ben Amara***

*Ecole Nationale d'Ingénieurs de Tunis (ENIT)

bouzrara_noura@yahoo.fr

**Faculté des Sciences de Monastir (FSM)

Aissaoui.Nacira@fsm.rnu.tn

***Ecole nationale d'Ingénieurs de Sousse (ENISO)

Najoua.Benamara@enim.rnu.tn

Nous proposons dans ce travail ARABASE une base de données d'images d'échantillons d'écriture arabe pour l'exploitation en reconnaissance optique de l'écriture (OCR-Optical Character Recognition). Cette base est implémentée sur un réseau à longue distance.

L'étude des particularités morphologiques de l'écriture arabe, dans sa forme imprimée et manuscrite (en ligne et hors ligne), et les différents problèmes liés à l'OCR arabe, nous ont conduit aux choix retenus au niveau de notre base de données.

Le contexte de ARABASE est diversifié (montant littéraires, noms de villes, texte libres, ligatures, nombres, signatures...), il correspond aux différents modes d'écritures : imprimé et manuscrit (hors ligne et en ligne). Pour chacune des classes du contexte correspondent des sous classes associées aux mots, pseudo-mots et aux caractères qui composent l'entité considérée. Des informations relatives à l'origine du document source et aux différents modes d'acquisition des données sont également disponibles dans ARABASE. Un document est produit par un périphérique d'entrée (un scanner, une tablette graphique, une imprimante...).

Dans le cas du manuscrit, nous considérons un contexte multi-scripteurs (Essoukri Ben Amara, 2005).

L'ensemble des informations de ARABASE est organisé dans le diagramme de classe statique selon la méthode orientée objet UML- Unified Modelling Language (Roques, 2002), la figure 1 donne un extrait de ce diagramme.

Plusieurs fonctionnalités sont offertes par cette application, nous citons en particulier :

- La consultation des différentes entités du contexte et des informations relatives aux outils d'acquisition.
- Les recherches selon des critères spécifiés par l'utilisateur, qui peut être administrateur ou client.
- La possibilité d'effectuer diverses statistiques relatives aux différents types d'informations

En plus de ces fonctionnalités classiques, ARABASE offre la possibilité d'enrichir le contexte de la base de données par l'ajout d'une nouvelle classe au modèle conceptuel, c'est-à-dire l'ajout de nouveaux vocabulaires au contexte de la base.

L'application est réalisée sous l'environnement SQL Server- Microsoft Structured Query Language Server (Spennik et Sledge, 2001) ce qui assure la sécurité des données.