Josep Llados, Dimosthenis Karatzas, Joan Mas and Gemma Sanchez. 2008. *A Generic Architecture for the Conversion of Document Collections into Semantically Annotated Digital Archives*.