French Ministry of Defense Selects A2iA FieldReader to Process World War One Archives.

La technologie A2iA au service des Archives du Ministère de la Défense

Latest Version of A2iA's FieldReader Handwriting Recognition Engine Reads Images from World War I Archives and Enables People to Access the Records on the Internet

Pour permettre à tous les citoyens d'accéder à la Mémoire française, le Ministère de la Défense a confié à A2iA et au prestataire de service SAFIG la numérisation, l'indexation et le traitement du fichier des 1.300.000 morts pour la France de la Guerre 1914-18. L'ensemble de ces fiches est en ligne sur le site **www.memoiredeshommes.sga.defense.gouv.fr,** depuis le 11 novembre 2003.

The French Ministry of Defense has selected the advanced recognition engine A2iA FieldReader Version 2.0 to read and process World War I archives.
The Ministry of Defense chose to partner with A2iA for its ability to accurately and efficiently read cursive handwriting. A2iA adapted its recognition engines to improve its performance on the handwriting style of that era.
On November 11, 2003, which is the 85th anniversary of the cease fire that ended World War I, the Archive Department of the Ministry of Defense made World War I patriots' records available on the Internet, and among the 1,300 000 records, many contain medical information that must remain confidential in order to meet French legislation requirements.
A2iA FieldReader determines which records contain medical information and therefore need to remain undisclosed and enables the Ministry of Defense to effectively post only the details that are considered public information for people to access on the Internet **(www.memoiredeshommes.sga.defense.gouv.fr)**.

The detection of medical information contained on the records was done in four steps:

1. The Recognition Zone is located using A2iA FieldReader's key word detection system. The field is automatically located even though its position varies on each form.
2. The "Cause of Death" field is recognized using A2iA's Intelligent Word Recognition Technology, by matching the handwritten information to a vocabulary of approximately 2,000 words in a dictionary to describe a cause of death.
3. Search for medical key words, which are matched against a dictionary containing 260 medical words.
4. Classification/Confidentiality: Depending on the presence or lack of a medical key word, the record remains classified or is posted on the website.

With millions of archives, records and old documents worldwide, there are numerous projects that A2iA's advanced recognition technology could handle in various international languages.

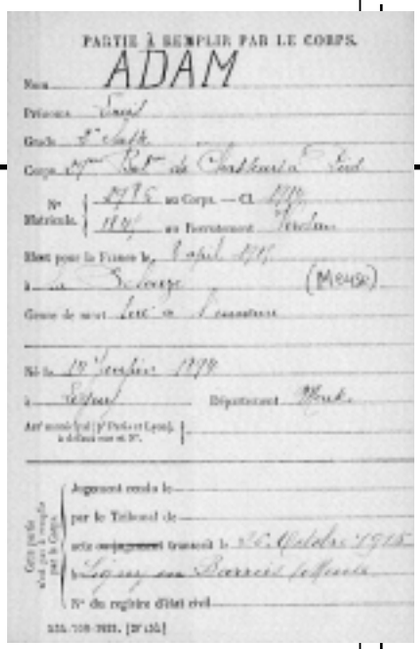
"We are very proud that A2iA FieldReader has enabled citizens to access information that remained unavailable for so long. We look forward to working on other archive recognition projects around the world," said Jean-Louis Fages, CEO of A2iA Corp.

« Nous espérons que notre expérience dans la lecture automatique des archives pourra contribuer à faire perpétuer de nombreux documents anciens dont les intéressés ignorent l'existence », conclue Jean-Louis Fages, Directeur Général d'A2iA.

Selon la législation, les fiches contenant des informations à caractère médical devaient rester confidentielles. Grâce à la technologie de reconnaissance de l'écriture manuscrite d'A2iA, le fichier du Ministère de la Défense a pu être traité de manière automatique, et le champ « genre de mort » analysé, pour extraire les fiches à caractère médical.

Dans le cadre de ce projet, la DMPA (Sous-direction des Archives et des Bibliothèques) a lancé en 2000, un vaste programme de recherche scientifique avec A2iA et l'Université Paris V René Descartes.
A2iA a pu ainsi travailler à la lecture d'écritures anciennes et adapter son logiciel pour atteindre des résultats de recherche satisfaisants.

En application de la loi 79-18 du 3 janvier 1979 relative aux archives dont l'article 7 stipule que « le délai au-delà duquel les documents d'archives publiques peuvent être librement consultés est porté à cent cinquante ans à compter de la date de naissance pour les documents comportant des renseignements individuels de caractère médical », il a fallu procéder à un second traitement des fiches de décès des soldats pour extraire celles contenant des mentions médicales.



Pour détecter les fiches contenant des mentions médicales dans les délais qui lui étaient impartis, A2iA a procédé en plusieurs étapes :

1. Localisation de la zone de reconnaissance : cette zone n'est pas toujours située au même endroit sur l'ensemble des 1.300.000 fiches traitées.
2. Reconnaissance de la mention « genre de morts» : en s'appuyant sur un dictionnaire de 2000 mots représentant l'ensemble des mots possibles écrits dans la zone de reconnaissance (i.e. le contexte linguistique de l'application).
3. Recherche de mots clés médicaux : sur l'ensemble des hypothèses de reconnaissance des mentions, à l'aide d'un dictionnaire de 260 mots clés.
3. Enfin la classification en mention normale ou médicale selon l'absence ou la présence d'au moins un mot clé médical.

Le ministère de la Défense exigeait une qualité de détection assurant un taux d'erreur de 0.5% parmi les documents acceptés. A2iA a su maintenir la qualité de ses résultats en dessous de ce seuil en atteignant un taux de substitution de 0,2% des fiches traitées tout en assurant une économie de saisie de l'ordre de 2/3.

La solution de reconnaissance de l'écriture cursive d'A2iA ouvre de grandes perspectives d'avenir : il existe des millions d'archives pouvant faire l'objet de numérisation, de reconnaissance et d'indexation pour perpétuer durablement ces informations sous forme électronique.