

Andreas Fischer

*Centre for Pattern Recognition and Machine Intelligence, Concordia University*

**The HisDoc Project**  
**Automatic Analysis, Recognition, and Retrieval of**  
**Handwritten Historical Documents for Digital Libraries**

**Résumé :**

This talk gives an overview of the HisDoc project, which aims at developing tools to support cultural heritage preservation by making historical documents, particularly handwritten medieval manuscripts, electronically available for access in digital libraries. The project intends to provide generic tools that can be adapted, without large effort, to different scripts and languages. HisDoc is a synergy project that brings together experts from three universities in Switzerland, namely the universities of Fribourg, Bern, and Neuchâtel. It consists of three major components. The first component is layout analysis, which aims at providing a structural description of the scanned document pages. The second component is handwriting recognition. Here flexible and robust recognition systems are developed that are suitable to automate the transcription of historical manuscripts. In this context, flexibility means that the system can be adapted to new writing styles without great effort, and robustness means that the recognizer should attain a high rate of correctly recognized words. The third component of HisDoc is information retrieval. Here suitable methods are being developed that allow effective information retrieval from non-perfectly transcribed text. The talk will give an introduction to the HisDoc project, describe its present state, and report on various experiments and their results.

Joint work with Horst Bunke, Nada Naji, Jacques Savoy, Micheal Baechler et Rolf Ingold.

