



UNIVERSITÉ
DE GENÈVE



Contribution ID: 1

Type: **not specified**

Tutorial 1. MarcXimiL : near duplicates detection (and similarity analysis)

Wednesday, June 22, 2011 9:00 AM (2h 30m)

MarcXimiL is an open source tool which works on MARCXML records and calculates similarity indices between these records. After a short theoretical introduction, the tutorial will focus on how to install, parametrize and use the tool.

This tool can be implemented in order to :

- prevent creation of duplicates (similar records are shown during the validation process)
- identify duplicates into batch files before ingest
- find duplicates inside a collection
- suggest to users similar records to the one found after a request
- match related documents eg. preprints and articles
- and so on.

<http://marcximil.sourceforge.net/>

Your affiliation/institution

University of Geneva & Ecole Polytechnique Fédérale de Lausanne

Your name

Jan Krause & Alain Borel

Your email

jan.krause@unige.ch

Presenters: Dr BOREL, Alain (Ecole Polytechnique Fédérale de Lausanne (EPFL)); Mr KRAUSE, Jan (University of Geneva)