

# Example Based Machine Translation for Cross-lingual Information Retrieval

**Cristina Vertan**

vertan@informatik.uni-hamburg.de

## Abstract

Cross-lingual retrieval is a key issue in the development of current search engines, either in the web or in knowledge management systems (like for example eLearning platforms).

## Outline

Authors are requested to format their papers according to the guidelines specified in this style sheet. Cross-lingual retrieval is a key issue in the development of current search engines, either in the web or in knowledge management systems (like for example eLearning platforms).

One of the approaches to cross-lingual retrieval is the use of a machine translation engine for the query translation, followed by the usage of classical retrieval techniques. Especially when dealing with a big number of languages in which the retrieval has to be performed it is not cost effective to develop complex translations engines. Two arguments support this hypothesis:

- the syntax and semantic of the queries is usually simple
- the retrieval methods cannot consider fine semantic relations in the query,

Usually the queries follow syntactic and semantic patterns, therefore Example based machine translation is a promising technique. In this tutorial we aim to demonstrate how example based machine translation can be used in cross-lingual retrieval and illustrate this with help of a running system. This tutorial addresses to computational linguists, researchers in NLP in particular MT and IR, familiar with basic challenges and paradigms of Machine Translation..

## Contents

1. Example based machine translation
  - Presentation of the method
  - Matching: presentation of different distance measures
  - Recombination strategies
  - Enhancement of the example database with linguistic and ontological knowledge
2. Cross-lingual
  - Basic methods in cross-lingual retrieval
  - Evaluation of cross-lingual retrieval
3. Use of EBMT in cross-lingual retrieval
  - Strategy
  - Experiments with different parameters in matching and recombination phase
  - EBMT vs. Word by word translation in cross-lingual retrieval
  - Evaluation of results
  - Demonstration of running system

- A complete CV can be found at <http://nats-www.informatik.uni-hamburg.de/~criem>

## Presenter

Dr. Cristina Vertan is senior researcher at the University of Hamburg, Dept. of Natural Language Processing. She is coordinating a research group on Example based machine translation and is involved in the development of a cross-lingual retrieval engine for learning objects in 9 languages. She organized recently several workshops on Machine Translation, delivered a tutorial at ESSLI summer school about MT, and supervised two intensive hands-on seminars on the same topic. She is delivering regularly courses and seminars on Machine Translation and multilingual retrieval at the University of Hamburg