

Proposed 10/22/08

# Word Sense Disambiguation for German

- **S: (n) bass** (the lowest part of the)
- **S: (n) bass, bass part** (the lowest)
- **S: (n) bass, basso** (an adult male)
- **S: (n) sea bass, bass** (the lean fish)

## Motivation

Word Sense Disambiguation (WSD) is the process of identifying the sense of a word in any given sentence, when the word has a number of distinct senses. At EGIZ, accurate WSD is needed for the detailed automatic analysis of German eParticipation documents.

## Project description

- Goals and tasks
  - Get an overview about existing WSD algorithms
  - Evaluation of these algorithms when applied to German documents
  - Compare quality and performance of the analyzed algorithms
  - Integration of the best algorithms in an existing classification framework
  - Task can be organized as project or as master thesis

## Literature

- Basics: [http://en.wikipedia.org/wiki/Word\\_sense\\_disambiguation](http://en.wikipedia.org/wiki/Word_sense_disambiguation)

## Deliverables

- Project files (Eclipse)
- Documentation (inline)
- Presentation (10 slides)

## Project schedule (master project or thesis)

- Start Immediately
- Month 1 Reading literature, find available algorithms, concepts and their application to German
- Month 2 Implementation, analysis
- Month 3 Implementation, final deliverables

## Master Project

Studies:  INF  SEW  TEL

## Prerequisites

- Java, Eclipse
- Machine learning basics

## Advisor / contact

[peter.teufl@egiz.gv.at](mailto:peter.teufl@egiz.gv.at)