



3803ICT  
Big Data Analysis

## **Lab 06 – Data Analytics for Texts (cont'd)**

## Table of Contents

<b>1. Topic Modelling.....</b>	<b>3</b>
<b>2. Named Entity Recognition .....</b>	<b>3</b>

# 1. Topic Modelling

This exercise is about modelling the main topics of a database of News headlines.

The tasks include:

- Text pre-processing
- Topic modelling using LSA - LSI
- Topic modelling using LDA
- Topic visualization

Open the News-Topic-Modelling.ipynb notebook and follows the instructions.

# 2. Named Entity Recognition

This exercise is about named entity recognition and its application in hiding personal information.

The tasks include:

- Text pre-processing (if needed)
- Named Entity Recognition
- Censor any person names

Open the GDPR-Compliance.ipynb notebook and follows the instructions.