The research of methods for analyzing material on the Web can be split up into three steps:

1. **Data acquisition**
   - Crawling or creation of development data using the Web
   - Definition of relevant scenarios and data sources with support from the authorities
   - ISPs, social network providers etc. will assist in providing interfaces, metadata etc.
   - Cleaning and preprocessing, e.g. treatment of typos, slang...

2. **Data analysis**
   - Development/application of state of the art Natural Language Processing (NLP) techniques for:
     - Semantically enriched document retrieval
     - Keyphrase Extraction
     - Topic Clustering
     - Named Entity Recognition / Disambiguation
     - Relationship Extraction
     - Automatic Summarization
     - Example: Identification of key persons in an extremist forum, analysis of their relationships and the content of their posts

3. **Presentation of results**
   - Development of user interfaces for:
     - Visualizing and highlighting relevant results
     - Interactive exploration of the result space
     - Assistance for transferring results into evidence usable in court

**EXPECTED RESULTS**

The police and other authorities are challenged by the new forms of communication in the Web 2.0, which are increasingly used for preparing, organizing or committing crimes such as:

- Sexual harassment of children
- Distribution of illegal and dangerous material
- Planning of unauthorized demos, terror acts etc.
- Announcement of rampages and suicides
- Weapon, drug or human trafficking

To make information on the Web manageable for manual inspection, we aim to research methods for processing natural language documents.

**Goals**

- Create tools which aid in investigating crimes on the Web
- Find relevant documents using a semantic search
- Identify relevant information bits (persons, places, times)
- Analyze the relations between them

**RELATED WORK AND COLLABORATION**

**Existing collaboration agreements**

- CCICADA (Command, Control and Interoperability Center for Advanced Data Analysis)
  - Research Director: Prof. Eduard Hovy
  - Various high-profile projects in this area
  - Department of United States Homeland Security
- FIVES/INBEKI
  - EU-Projects for fighting child pornography
  - Coordinator: Dr. Armin Stahl, DFKI

**Application partners**

- Landeskriminalamt Hessen
- Bayerisches Landesamt für Verfassungsschutz

**Team:**

Iryna Gurevych, Michael Matschek