ESSIR Summer School 2011
August 31, 2011, Koblenz, Germany

SYMPOSIUM ON BIAS AND DIVERSITY IN IR
(aka LIVINGKNOWLEDGE SUMMER SCHOOL)

LivingKnowledge Consortium
What are the different perspectives?

What types of images are used?

Are some sources biased?

50,000 Wissenschaftler gegen Guttenberg

In der Debatte um die Plagiatsvorwürfe gegen Verteidigungsminister Karl-Theodor zu Guttenberg (CSU) gerät Bundeskanzlerin Angela Merkel in die Kritik.
VISION

Diversity and bias in the Web today:
• High diversity of content and content creators
• But: no systematic support to explore the diversity

Vision of Diversity-related research:
• Make diversity, bias and evolution traceable, understandable and exploitable

Aim and scope of this symposium:
• Give an overview of leading-edge technology for dealing with diversity and bias in IR
CONTEXT: LIVING KNOWLEDGE PROJECT

- Part of the presented material are relevant results from the LivingKnowledge project
- EU FET Project since February 2009
- **Project Goal:**
  Improve navigation and search in large cross-media datasets and the Web by considering diversity, time, opinions and bias
- Relevant project results:
  - Testbed with more than 40 components (will be shown and used today)
  - Innovative methods and algorithms, e.g. sentiment extraction from images, image forensics, contextualization of facts
  - First applications of diversity-aware technology, e.g. Time explorer
CONTEXT: LIVING KNOWLEDGE PROJECT

- Relevant project results:
  - Testbed with more than 40 components (will be shown and used today)
  - Many innovative methods and algorithms, e.g., sentiment extraction from images, image forensics, contextualization of facts;
  - First applications of diversity-aware technology, e.g., Time explorer Bias and Diversity in IR, August 31, 2011

Impact of Contrast adjustment on opinion perceived by the users.

Contrast adjustment impacts on opinion perceived by the users.

**Module** | **Status**
---|---
HTML Content Analyser | In testbed
Codebook Aggregator | In testbed
URL Extractor | In testbed
Document Layout Extractor | In testbed
Places Annotator | In testbed
Entity Disambiguator | In testbed
EXIF Image metadata Extractor | In testbed
Image Colourfulness | In testbed
Image Naturalness | In testbed
Image Seniment Annotator | In testbed
Quote Extractor | In testbed
Phrase Extractor | In testbed
TimeML Annotator | In testbed
DicFonary Annotator | In testbed
Fact Annotator | In testbed
'Best Image' Annotator | In testbed

**Legend**
- Green: Untampered
- Red: Tampered
- Blue: Not Analyzable

**Intensity of contrast modifications**
- Heavy negative: -35
- Light negative: -20
- Original: 0
- Light positive: +20
- Heavy positive: +35
SYMPOSIUM STRUCTURE & OVERVIEW

• Structured into 3 parts

• Part I: Discovering Facts and Opinions and their Evolution over Time

• Part II: Discovering Dias and Diversity in Multimedia Content

• Part III: A Testbed for Diversification in Search
  – Demonstration of selected technology
  – Time for further in depth discussions
Part I: EXTRACTION OF OPINIONS FROM THE WEB

• Introduces basic concepts in opinion extraction
• Overview of standard coarse- and fine-grained opinion extraction methods
• Introduction to domain adaptation
• Overview of opinion resources
• Shows some recent research results in the area
PART I: TEMPORAL FACT EXTRACTION, DISAMBIGUATION, AND EVOLUTION

• Factual knowledge evolves over time

Goal: Exploring the temporality of entities

- Extract entities from Web documents
- Disambiguate them to canonical entities
- Aggregate facts to investigate their dynamics
PART II: A BRIEF INTRODUCTION TO EXTRACTING INFORMATION FROM IMAGES

• Why look at images?
• Image representation and understanding
• Information extraction techniques

INFORMATION EXTRACTION FROM MULTIMEDIA CONTENT ON THE SOCIAL WEB

• How to exploit combined information from visual data and meta data
• Photos on the social web: attractiveness, maps and sentiment
• Videos on the social web: tagging
Part III: A BRIEF INTRODUCTION TO THE DIVERSITY ENGINE TESTBED
FROM THEORY TO PRACTICE

Learn how to run these tools and visualize the results
Add your own tools
Build diversity-aware applications