

LTI

Language, Technology and the Internet

The Semantic Web

Francis Bond

Department of Asian Studies

Palacký University

<https://fcbond.github.io/>

bond@ieee.org

Lecture 10

Revision of Text and Meta-text

Text and Meta-text

- Explicit Meta-data
 - Keywords and Categories
 - Rankings
 - Structural Markup
- Implicit Meta-data
 - Links and Citations
 - Tags
 - Tables
 - File Names
 - Translations

Explicit Metadata

- You can get information from metadata within documents
 - When they are accurate they are very good
 - They are often deceitful
- HTML, PDF, Word, ...Metadata
- Keywords and Tags
- Rankings
- Links and Citations
- Structural Markup

Implicit Metadata

- You can get clues from metadata within documents
 - as they are non-intended, they tend to be noisy
 - but they are rarely deceitful
- HTML tags as constituent boundaries
- Tables as Semantic Relations
- File Names (content type and language)
- Translations — Bracketed Glosses; Cross-lingual Disambiguation
- Query Data
- Wikipedia Redirects and Cross-wiki Links

LLM

ChatGPT

If ChatGPT Can Do It, It's Not Worth Doing

<https://www.insidehighered.com/opinion/blogs/just-visiting/2023/09/20/chatgpt-shows-way-toward-our-own-humanity>

AI from a legal perspective LWN By Jake Edge September 26, 2023

Chatbot Hallucinations Are Poisoning Web Search Wired, ILL KNIGHTBUSINESSOCT 5, 2023 12:00 PM (accessed 2023-10-06)

Signs of undeclared ChatGPT use in papers mounting *Retraction Watch* October 6, 2023 Frederik Joelsing (accessed 2023-10-03)

An Evaluation of a Zero-Shot Approach to Aspect-Based Sentiment Classification in Historic German Stock Market Reports (2023) Janos Borst, Lino Wehrheim, Andreas Niekler, Manuel Burghardt Preprints of Communication Papers of the of the 18th Conference on Computer Science and Intelligence Systems pp. 51–60