

LTI

Language, Technology and the Internet

The Semantic Web

Francis Bond

Division of Linguistics and Multilingual Studies

<http://www3.ntu.edu.sg/home/fcbond/>

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 Redirections 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>