

# Die Rolle von Metadaten, Semantik und Ontologien im Web

Vortrag zum Seminar „Web-Qualitätsmanagement“

von Frank Wischeropp

# Gliederung

- Motivation
- Metadaten
- InfoHarness System
- Metadata Reference Link (MREF)
- Markup Formate
- Ontologien
- Fazit

# Motivation

- Web: große Mengen von Daten in strukturierten, unstrukturierten u. sequenziellen Formaten
- Semantik Flaschenhals
- Bedarf der Unterstützung der Zusammenhänge zwischen Inhalten verschiedener Formate und Medien
- Semantik Web:
  - Ziel: Nutzung von semantischen Informationen zur Unterstützung von Beziehungen
  - Maschinen lesbare Informationen für automatisierte Dienste
  - Metadaten und Vokabulare (Ontologien)

# Metadaten

- Metadaten: „Daten über Daten“
- Typen:
  - Inhaltsunabhängig (z.B. Ort)
  - Inhaltsabhängig (z.B. Seitenzahl)
    - Direkt Inhaltbasiert (z.B. Volltext Index)
    - Inhaltsbeschreibend
      - Domain unabhängig
      - Domain spezifisch
        - Intra Domain
        - Inter Domain

# Metadaten zur Kapselung von Informationen

- InfoHarness System: Metadaten basiertes Objekt Modell für digitale Inhalte

## InfoHarness Object (IHO)

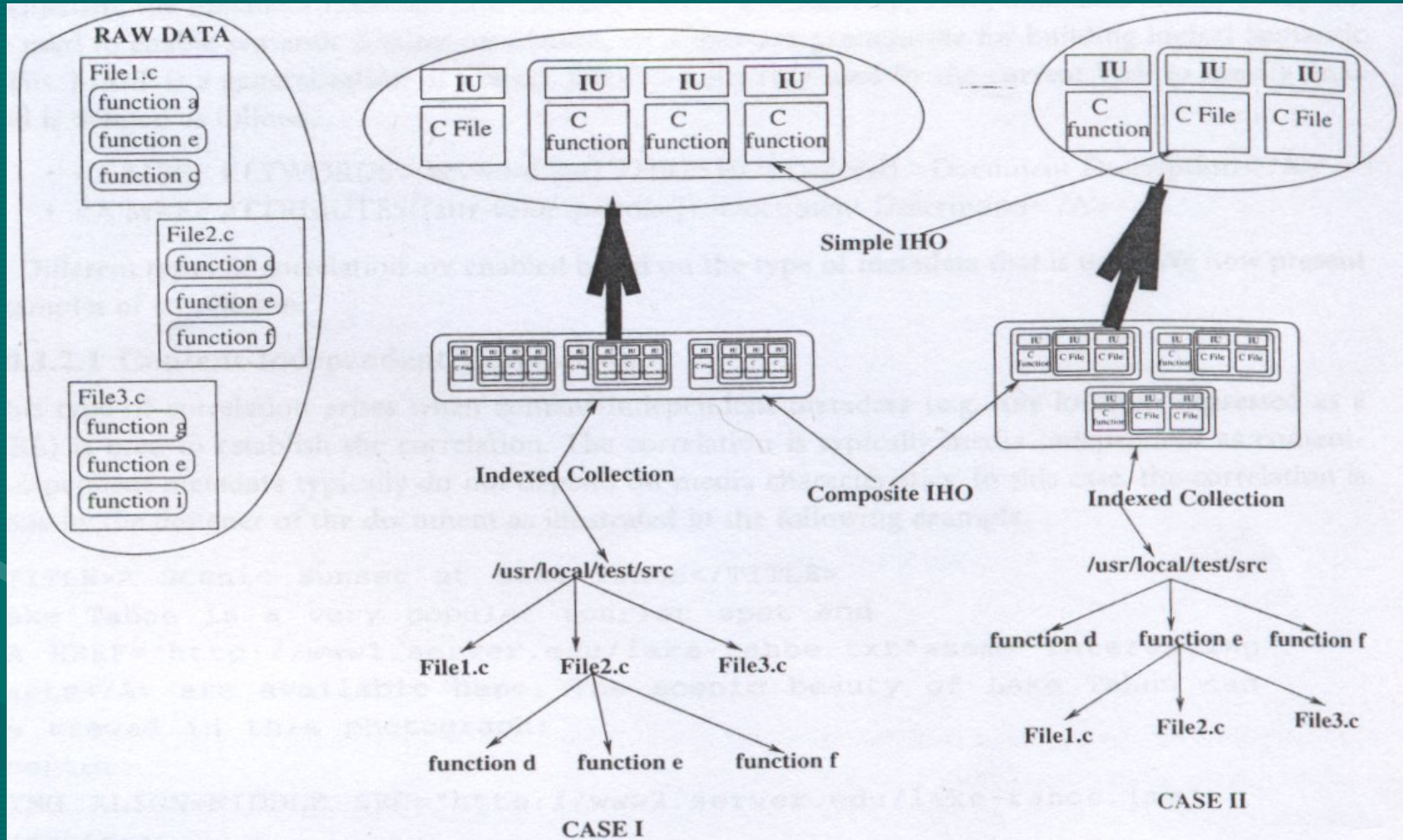
- Information Unit
  - Type
  - Location
  - Other Attributes
- Collection of IHO's

Metadata  
Extraction

(portion of) text file,  
bitmap, ...

- InfoHarness Repository Definition Language (IRDL)

# Beispiel Objekt Modell



# Metadata Reference Link (MREF)

- Potentielle Erweiterung von HTML
- Generalisierung von HREF
- mächtige Möglichkeit, Daten zueinander in Beziehung zu setzen unabhängig von ihrem Ort, Format und Medium

# MREF II

- Inhaltsunabhängige Korrelation
  - `<A HREF=„http://www.waterfall.org“>Waterfall</A>`
- Korrelation durch direkt inhaltsbasierte Metadaten
  - `<A MREF ATTRIBUTES (keyword=„waterfall“; color=„blue“)>Waterfall</A>`
- Domain spezifische Korrelation
  - `<A MREF ATTRIBUTES(population > 500; area > 50; regiotype=block; landcover=urban; relief=moderate)>regions having population greater than 500 and area greater than 50 sq feet having an urban landcover and moderate relief</A>`

# Markup Formate

- XML – Meta-Sprache zur Beschreibung von Markup-Sprachen
- XSLT & XPath – Sprachen zur Transformation von XML-Spezifikationen
- XQuery – Sprache zur Anfrage an XML-Daten
- RDF – Format zur Darstellung maschinenlesbarer Metadaten im Web
- DAML+OIL: DARPA Agent Markup Language

# Ontologie

- Ontologie:
  - beschreibt Wissensbereich mit Hilfe standardisierten Terminologien u. Beziehungen
  - Erleichtert Teilen und Wiederverwendbarkeit von Wissen
- Sprachen zur Beschreibung von Ontologien:  
RDF, DAML+OIL
- Terminologische Verpflichtung – Brücke zwischen Content-Providern und Nutzern

# Fazit

- Vorhandene Probleme in Bezug auf Skalierbarkeit und Informationsverlust im Web
- Erzeugung und Extraktion von maschinenlesbaren Metadaten als Komponente des Semantik Web als Lösungsansatz
- Herausforderung: Sinnvolle Definition von Vokabularen und Ontologien

# Literatur

- Munindar P. Singh: „The Practical Handbook of Internet Computing“



Vielen Dank für Ihre Aufmerksamkeit!