



Web Tomography

Towards eMeasurement and eControl

Im Rahmen des Seminars „Web-Qualitätsmanagement“

Stefanie Langer

25.05.2005



Gliederung

- Motivation
- Web-Measurement
- Software eMeasurement
- eControl Software
- Web-Tomografie
- „WebTomix“
 - Das Tool WebTomix
 - Anwendungsbeispiel
- Auswertungen
 - Schritte
 - Arten der Auswertung / Formeln
- Ausblick



Motivation

- Schnelle Entwicklung und zunehmende Ausmaße der Bestandteile des Webs
 - Anpassung, Verbesserung, Wartung etc. mit immer höherem Aufwand verbunden
 - monetäre Bedeutung
- Technologien als Bestandteil des Webs
- Messung
 - Kennzahlen gewinnen und verwerten
 - Welche Institution benutzt welche Technologie?



Web Measurement I

Web Measurement

Qualitätsanalyse

Charakteristika der Webseite

z.B.:

- Textanteil:
 $Text\ Ratio = KB_{text} / KB_{total}$
- Referenzierungen:
 $Link\ Anchor_{br} = (Link\ Anchor - Link\ Anchor_w) / PageCount$
- Anteil an Elementen
 $Var_r = Var / PageCount$
- Zuverlässigkeit
 $MTBF = Number_{access} / Number_{failure_access}$

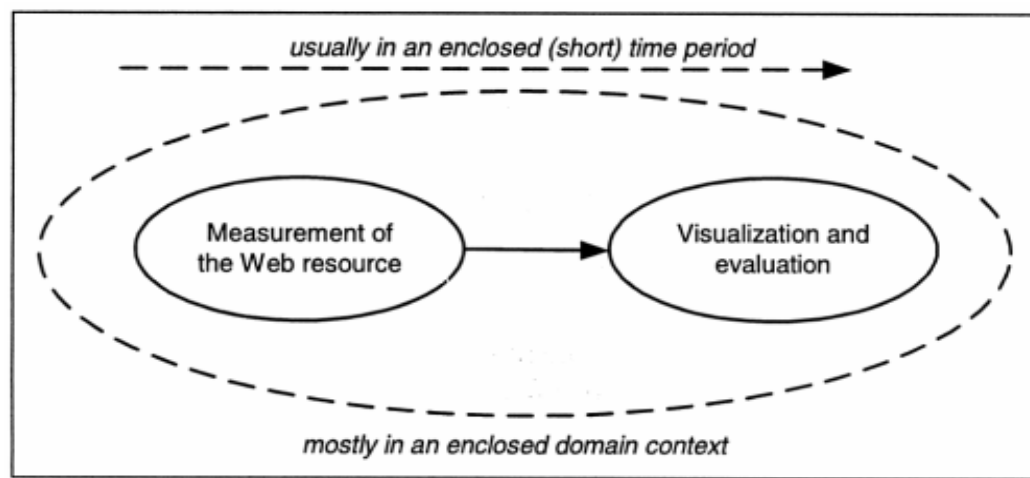
Analyse der Webnutzung

Webnutzung und Anwendungsform

z.B.:

- Anzahl der Zugriffe
in einer bestimmten Zeitspanne oder von speziellen Webdomänen
- Folge der Zugriffe
- Dauer/Häufigkeit der Besuche
durch (einen) bestimmte(n) Nutzer(gruppen)

Web Measurement II



■ Web Mining

„... die Übertragung von Techniken des Data-Mining zur (teil)automatischen Extraktion von Informationen aus dem Internet, speziell dem World Wide Web. Dabei können drei Untersuchungsgegenstände unterschieden werden:

- Die Inhalte (*Web-Content-Mining*) –Information-Retrieval
- Die Struktur der Verlinkung (*Web Structure Mining*) - Webometrie
- Das Benutzerverhalten (*Web-Usage-Mining*) - Analyse von Logfiles.“

http://www.wikipedia.org/wiki/Web_mining



Software eMeasurement

Beinhaltet:

- Software-Messaktivitäten
- die Messziele, -produkte, -werkzeuge, -ressourcen (Personal und Erfahrung etc.)

Basiert auf:

- Web-Technologien

„...die Anwendung und/oder Erweiterung von (standardisierten) Softwagemessmethoden durch web-basierte Technologien und Systeme.“

Aus: Lother, M., Dumke, R., Braungarten, R., Kunz, M.: Ein Portal zur funktionalen Größenmessung von Software – Anfänge eines Software eMeasurements



Web- Tomografie I

- **Tomografie** (*griech. tomos - Scheibe*)

Zusammenfassung verschiedener bildgebender Verfahren, bei denen durch schichtweise Aufzeichnung eines Objekts dessen räumliche Struktur ermittelt werden kann.

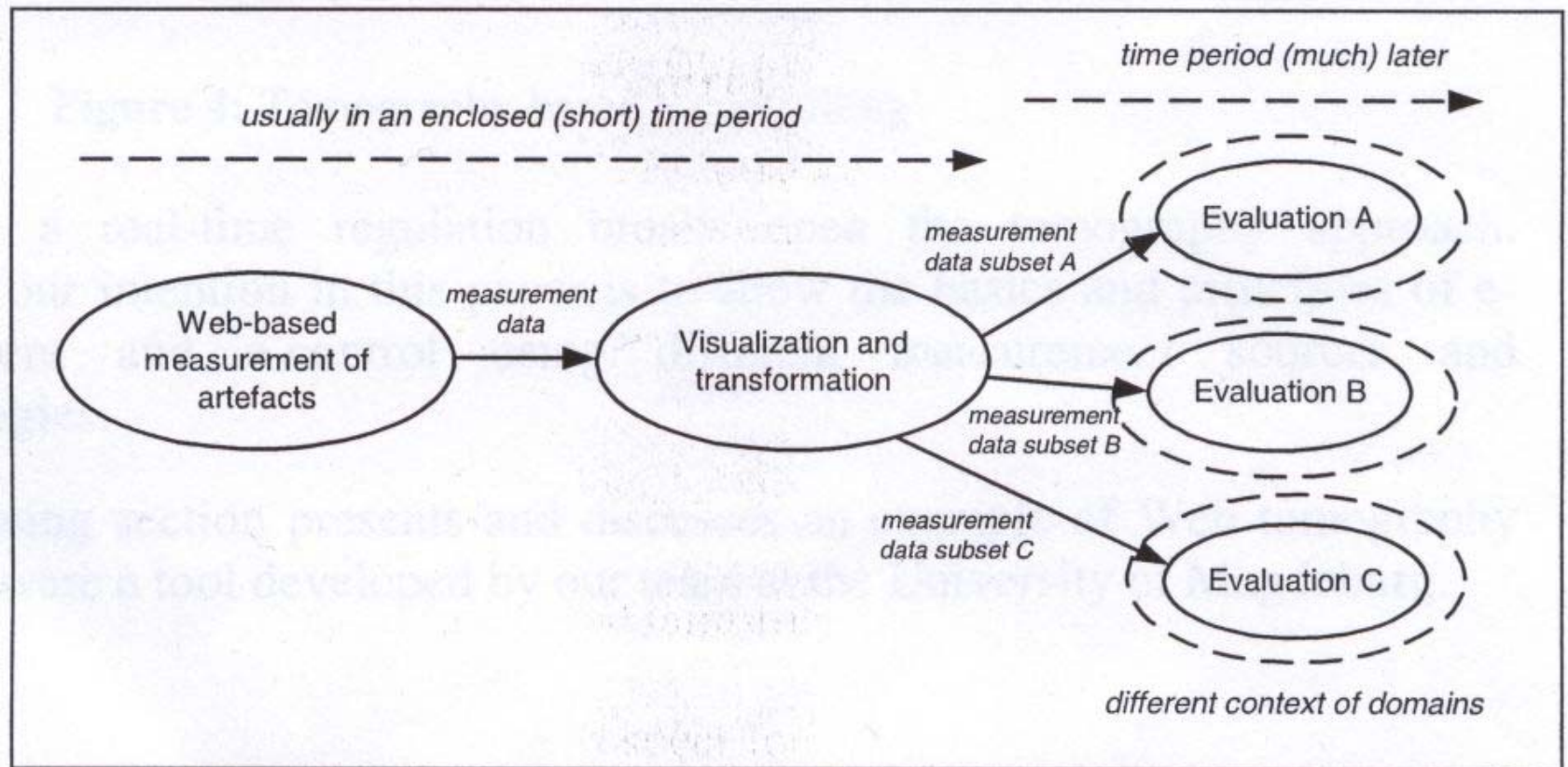
→ Medizin: Röntgenschnittverfahren

→ Internet-Tomografie

→ Web-Tomografie

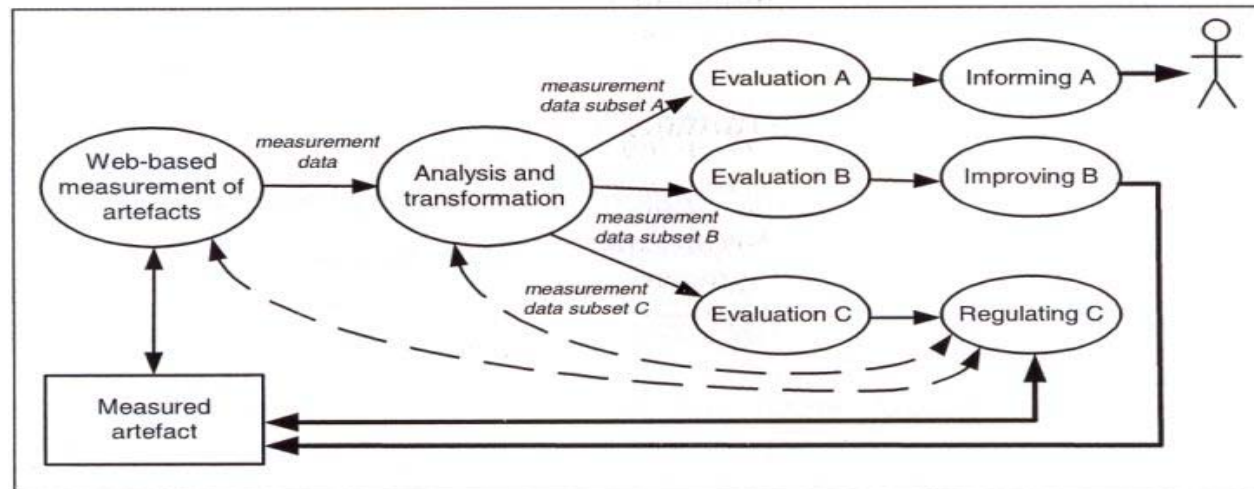


Web- Tomografie II



eControl Software

- Berücksichtigt die Gegenstände des Software eMeasurement und basiert wie dieses auf Web-Technologien.
- Bindet die Mess-Anwendung ein zur (automatischen)
 - Information des Benutzers
 - Verbesserung des Messgegenstandes
 - (ggf.) Regulierung des gemessenen und ausgewerteten Prozesses

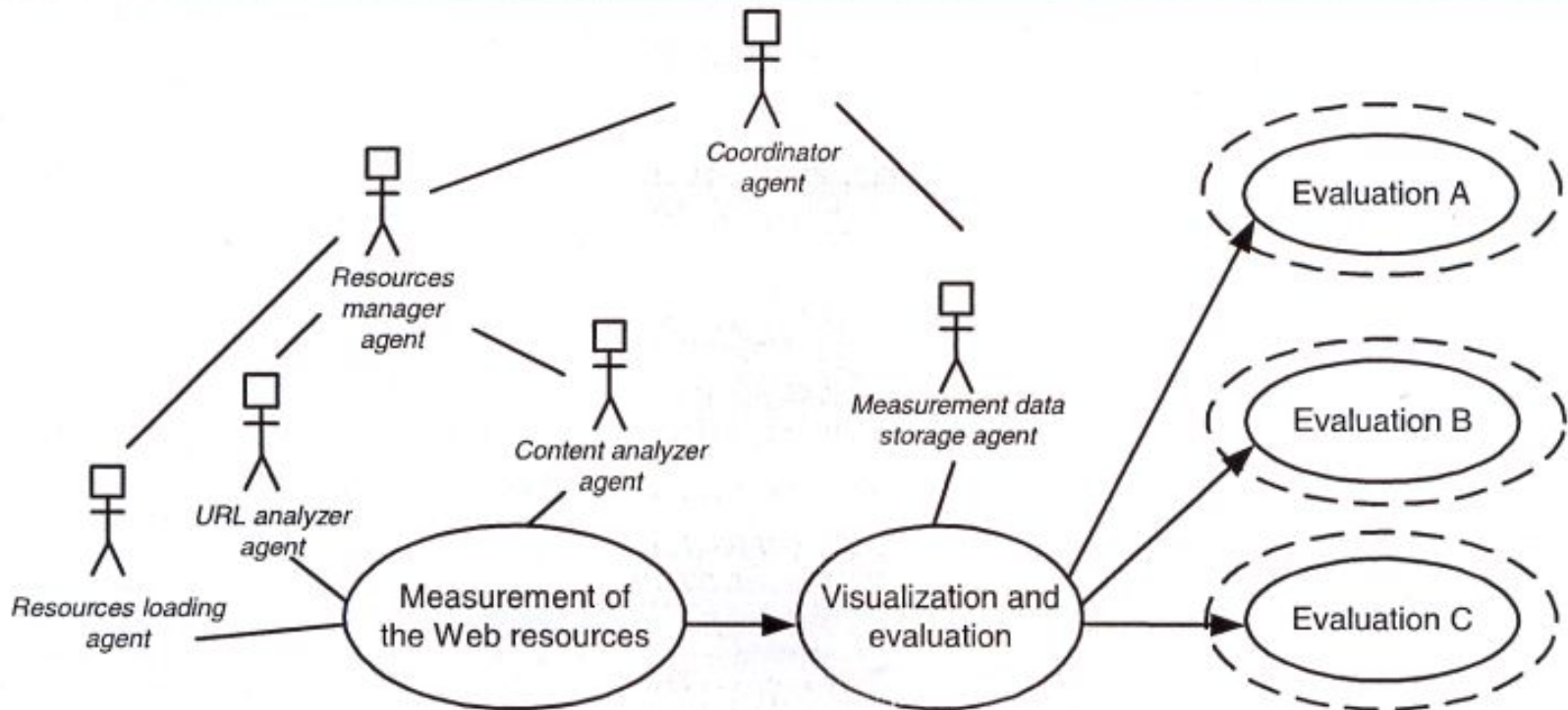




Der Webtomograf WebTomix

- systematisches Durchsuchen von Webpräsentationen
- Ausgangspunkt: eingegebene (Liste von) URL(s)
- Web-Tomograf sichtet alle vom Ausgangspunkt verlinkten Webseiten bis zu einer (zuvor) bestimmten Suchtiefe.
- Ausgabe: Verwendete Webtechnologien
 - XML-Format
 - Datenbank
- weitere Anforderung: erweiterbar

WebTomix: Agentenbasierte Struktur



WebTomix: Bedienung

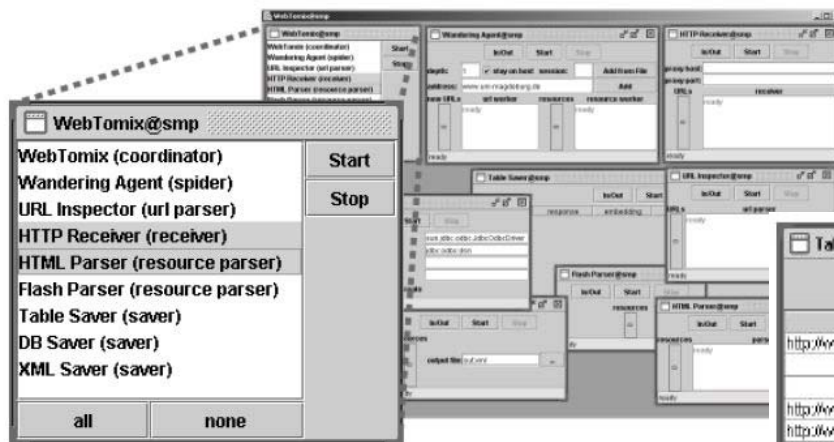


Abbildung B.1: WebTomix

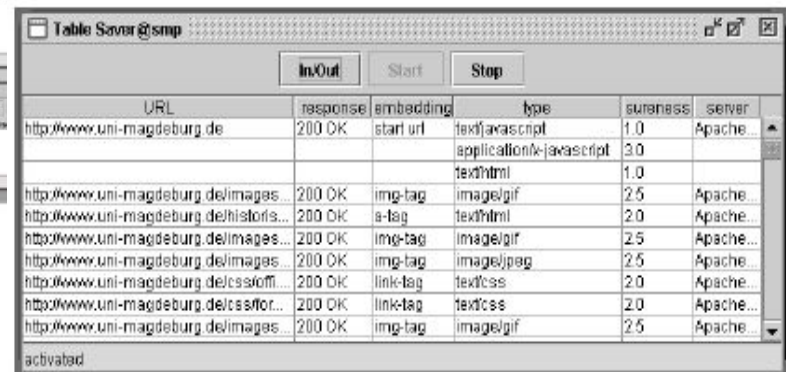
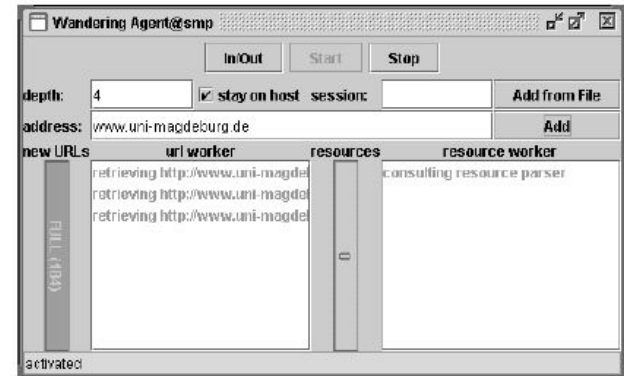


Abbildung B.7: Table Saver

WebTomix: Anwendung I

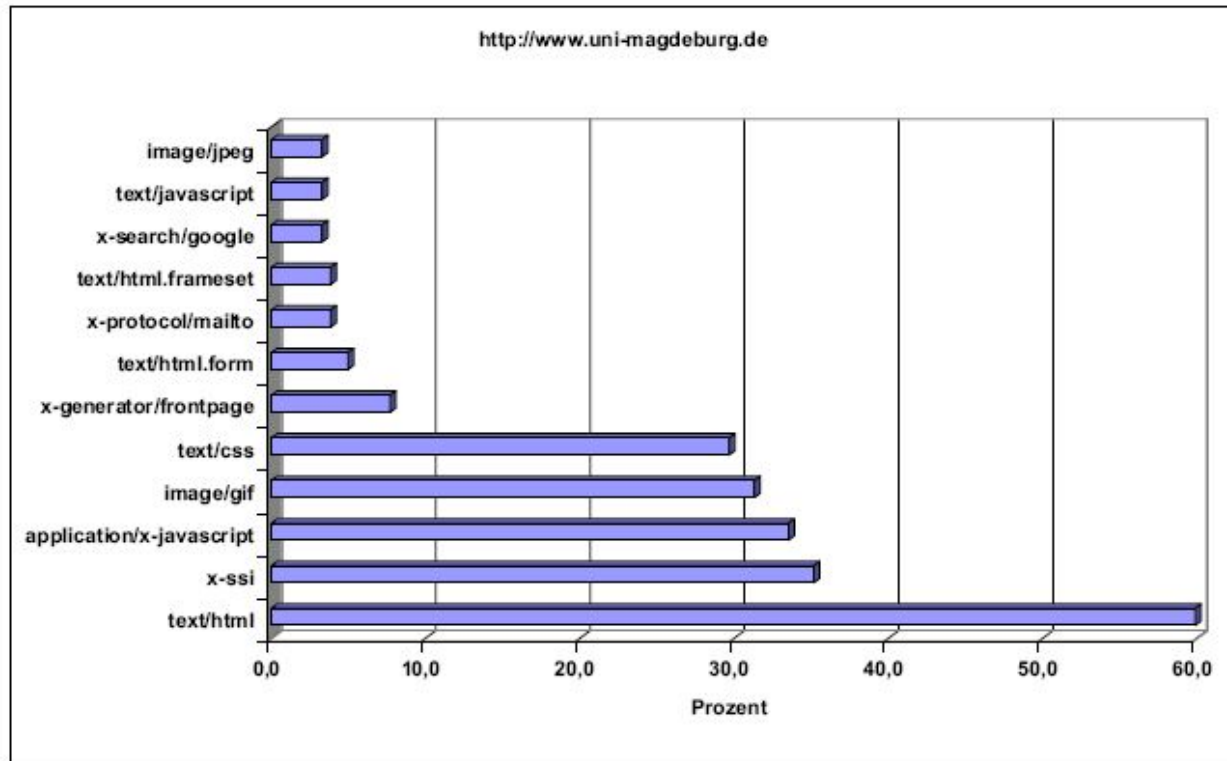


Abbildung 7.5: Otto-von-Guericke-Universität Magdeburg (Suchtiefe 3, 182 URLs)

WebTomix: Anwendung II

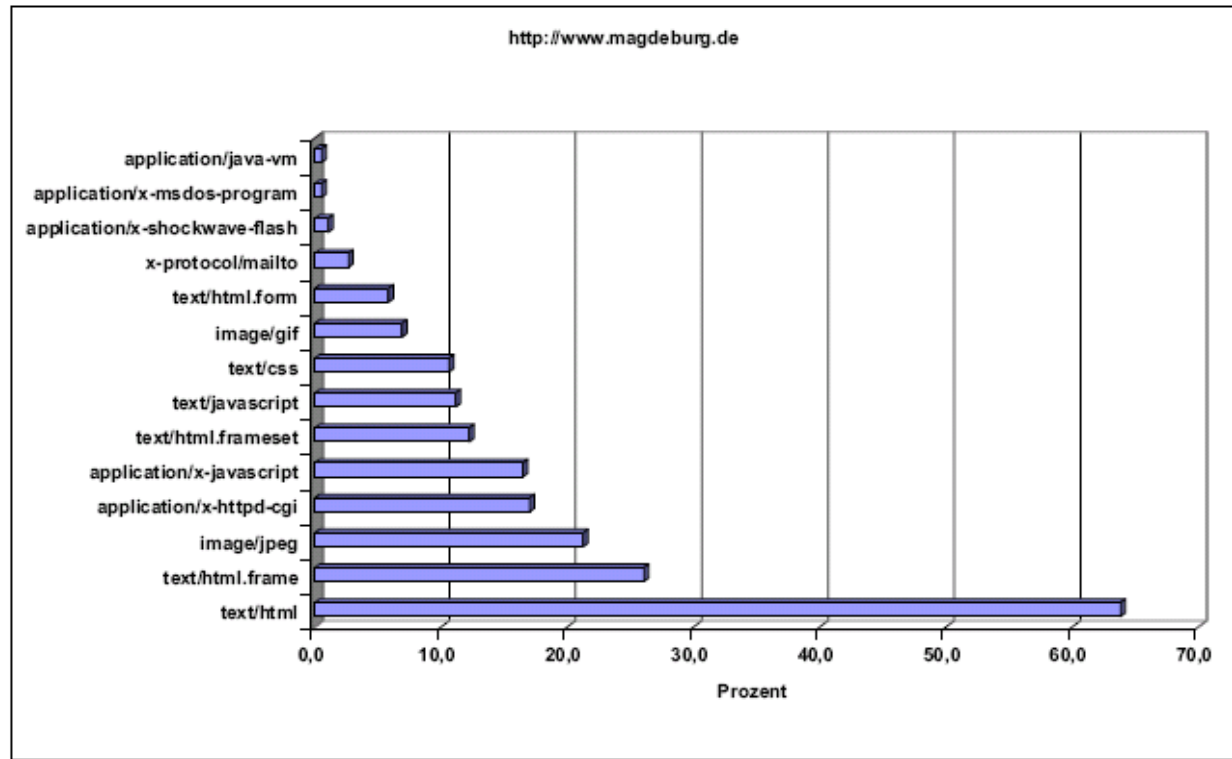


Abbildung 7.3: Stadtauftritt Magdeburg (Suchtiefe 5, 188 URLs)



eMeasurement – Auswertungen I

- Schritte:

1. Analyse der Web-Technologien
2. Analyse der technologischen Struktur der Webpräsentation
3. Aufstellen von Formeln/Schemata zur empirischen Auswertung
4. Gebrauch von bestehenden Erfahrungswerten zur Verbesserung der zuvor definierten Formeln/Schemata
5. Ableitung der endgültigen Formeln/Schemata zur empirischen Auswertung



eMeasurement – Auswertungen II

- Bsp. Lineare Interpretation (<-> strukturelle I.)
 - Einteilung der Technologien in Klassen

$T_{text} = \{html, msword, pdf, shtml\}$

$T_{dynamicParts} = \{cgi, css, dtd, flash, frameset, frontpage, htdig, javascript, ms-exel, ms-powerpoint, perl, php, ssi, xml\}$

$T_{communication} = \{authorization, cookies, form, mailto, ssl\}$

$T_{multimedia} = \{gif, image, jpeg, ms-wvx, png, pn-realaudio, wav\}$

- Aufstellen der Formeln

Performance $P = T_{text} [\%] / (T_{dynamicParts} [\%] + T_{communication} [\%] + T_{multimedia} [\%])$

Wartbarkeit $M = (T_{text} [\%] + T_{multimedia} [\%]) / (T_{dynamicParts} [\%] + T_{communication} [\%])$

Bedienbarkeit $U = (T_{communication} [\%] + T_{multimedia} [\%]) / (T_{text} [\%] + T_{dynamicParts} [\%])$

- Ableitung der Ergebnisformeln

Performance $P_{communities} \triangleleft P_{universities} \triangleleft P_{WebServices}$

Wartbarkeit $M_{universities} \triangleleft M_{communities} \triangleleft P_{WebServices}$

Bedienbarkeit $U_{WebServices} \triangleleft U_{communities} \triangleleft U_{universities}$



Ausblick

Erweiterung des Web-Tomografen zur weiterführenden Auswertung:

- automatisierte Eingabe von Start-URLs
- Erhöhung der Erkennungsvollständigkeit
- Erhöhung der Erkennungssicherheit
- Erkennung der Version u.ä. Details
- Mitarbeit bzw. Unterstützung von Herstellern



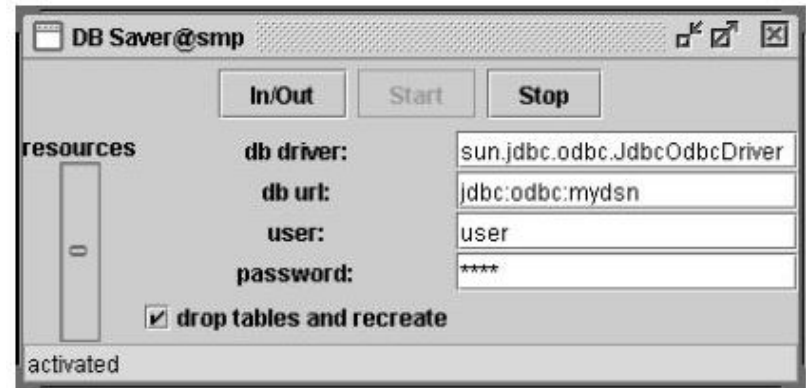
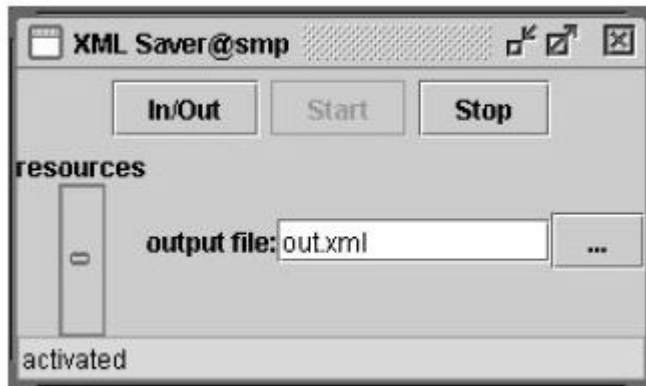
Informationsquellen

- Dumke, R., Lothar, M., Schäfer, U., Wille, C.: *Web Tomography Towards e-Measurement and e-Control*. In: Software Measurement – Research and Application, IWSM/MetriKon 2004, Königs Wusterhausen, Germany
- Schäfer, U.: *Analyse und Auswertung von Webpräsentationen auf der Grundlage eines implementierten Web-Tomographen*. Diplomarbeit, Universität Magdeburg, Fakultät für Informatik, März 2004
- Lothar, M., Dumke, R., Braungarten, R., Kunz, M.: *Ein Portal zur funktionalen Größenmessung von Software – Anfänge eines Software eMeasurements*. . Softwaretechnik Trends, 25(2005)1, S. 39-44
- http://de.wikipedia.org/wiki/Web_Mining, /~Tomografie, ~/Data Mining, ~/Information-Retrieval, ~/Webometrie, gelesen 15.05.2005
- <http://www.computerbase.de/lexikon/MTBF>, gelesen 23.05.2005
- <http://www.nature.com/nature/webmatters/tomog/tomog.html>, gelesen 11.05.2005
- http://www.html-world.de/program/ssi_1.php, 22.05.2005



Zusätzliches

XML Saver und DB Saver



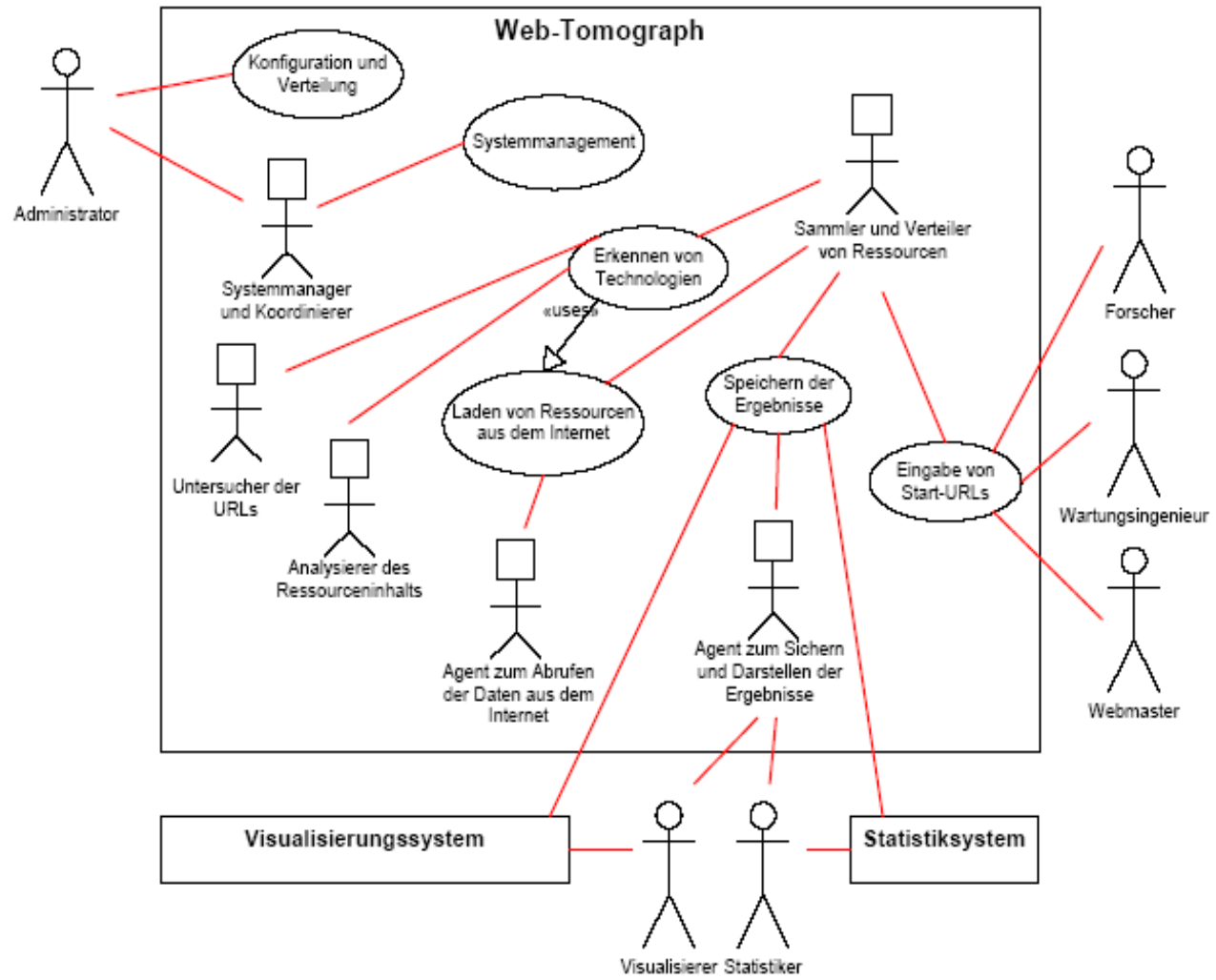
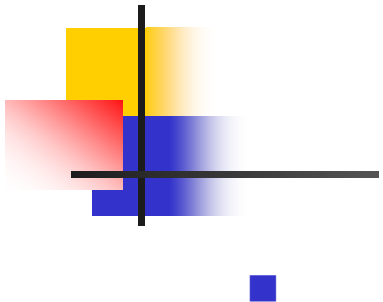


Abbildung 6.1: Anwendungsfälle für WebTomix

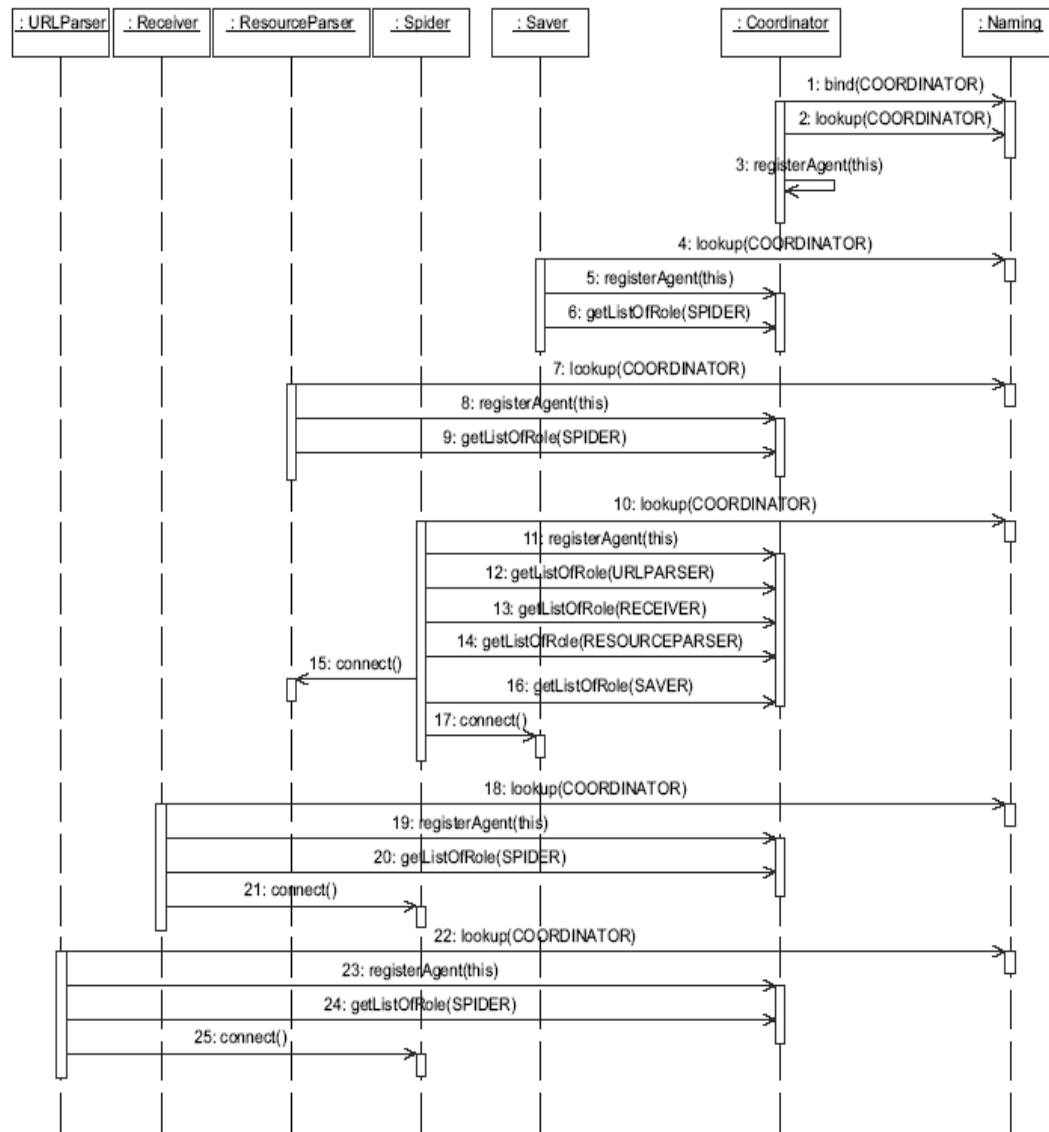
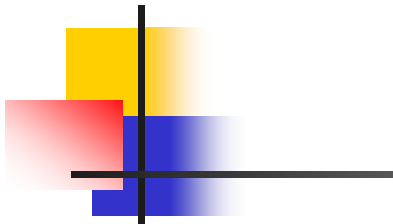


Abbildung 6.4: Eintreten der Agenten in das Agentensystem (Sequenzdiagramm)