



Web Metriken

Autor: Bastian Grabski

Datum: 08.06.2005

Ort: G 29 / Raum 321



Inhalt

- Einführung
- Metriken-
Übersicht
- Graphen
Eigenschaften
- Bedeutung
- Ähnlichkeit
- Suche
- Verwendung
- Informationstheorie
- Zusammenfassung
und Schluss



Einführung

- Web-Metriken dienen zur Messung von Merkmalen und Eigenschaften von Web-Seiten und des gesamten WWW!
- Tiefer Einblick in Web als Ganzes und auf einzelne Seiten aus verschiedenen Perspektiven



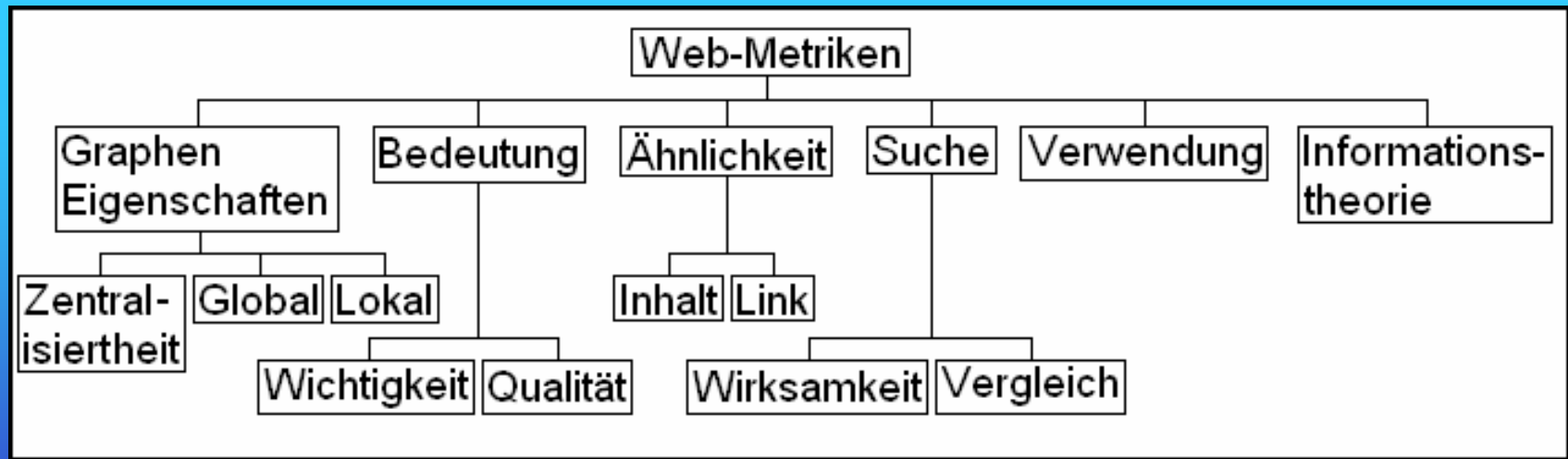
Einführung

- Messungen und Modelle um Attribute von Web-Seiten zu quantifizieren
- Ziel: Verbesserung von Informationsgewinnung, -Nutzung, -Qualität, Benutzbarkeit des WEB
- Bekanntesten Metriken mit Blick auf Messfunktionen, Anwendung, Bedeutung



Metriken-Übersicht

- Metriken-Gruppen basieren auf den gemessenen Attributen:





Metriken-Übersicht

- Graphen Eigenschaften:
 - WWW verkörpert durch eine Graphen Struktur mit Knoten (Web-Seiten) und Kanten (Links)
 - Graphen basierte Metriken quantifizieren strukturelle Eigenschaften des Web
 - Lesbarkeit und Navigationsfähigkeiten



Metriken-Übersicht

- Bedeutung:
 - Formalisiert absolute „Qualität“ und „Wichtigkeit“ (Informationsbedürfnis der Nutzer) von Web-Seiten
 - Suchmaschinen nutzen Metriken für ihr Antwortverhalten → z.B. genutzt von Google zur Antwort-Rangfolge



Metriken-Übersicht

- Ähnlichkeit:
 - Quantifiziert Umfang an Verwandtheit zwischen Web-Seiten
 - Existenz verschiedenster Indikatoren und Konzepte
 - Oft im Zusammenhang mit Clustern (Zusammenfassung der Informationen)



Metriken-Übersicht

- Suche:
 - Schätzen und Vergleichen der Leistung von Suchmaschinen u.ä. Diensten
 - Wirksamkeits- und direkte Vergleiche der einzelnen Services möglich



Metriken-Übersicht

- Verwendung:
 - Art und Weise wie Nutzer über Web-Ressourcen gehen (z.B. surfen)
 - Bietet viele Anhaltspunkte um Inhalt, Organisation und Präsentation zu verbessern
- Messungen in diesem Bereich



Metriken-Übersicht

- Informationstheorie:
 - Metriken messen Eigenschaften bezüglich Informationsbedürfnissen, Produktion, Konsum
 - Beziehungen zwischen einer Anzahl von Regelmäßigkeiten bei der Informationserzeugung im Web



Graphen Eigenschaften

- Zentralisiertheit:
 - Maße bestimmen Ausmaß an Verbundenheit zwischen Knoten im Graph
 - Hierarchien zwischen Web-Seiten mit einer Wurzel können definiert werden
 - Ziel: einfach zu navig. Seitenhierarchien, Graph als querverlinkte Baum-Struktur



Graphen Eigenschaften

- Globale Metriken:
 - Hypertext als Ganzes – keine Knoten
 - Hierarchisch organisierten Hypertext mit gekennzeichneten Links
 - Kompaktheit (Maß Quer-Referenzierung)
 - Schicht (Maß für lineare Anordnung bzw. Struktur des Web-Graph)



Graphen Eigenschaften

- Lokale Metriken:
 - Eigenschaften indiv. Knoten des Graphen
 - Tiefe (Wurzel-Distanz, Zusammenhang zwischen Bedeutung-Distanz-Entfern.)
 - Unbalanciertheit (jeder Knoten eine Idee, Links sind Entwicklungen, Hypertext ein balancierter Baum, unbalancierte Knoten)



Bedeutung

- Wichtigkeit:
 - Ähnlichkeit und Vergleich von Such-Seiten
 - Verschiedene Modelle genutzt
 - Information Retrieval Techniken angepasst um Wichtigkeit von Web-Seiten für Schlüsselwortabfragen zu bestimmen
 - Gewichtung der Suchanfragenergebnisse



Bedeutung

- Qualität:
 - Link Informationen zur Unterscheidung von häufig genutzten Seiten
 - Qualität einer Web-Seite abhängig von der umgebenden Hyperlink Struktur
 - Link ist ein positives Urteil über eine Seite
 - Nutzung von Link-Struktur-Analysen



Ähnlichkeit

- Inhalts-basierte:
 - Durch Textvergleiche der Dokumente
 - Inhaltsvergleiche von z.B. Teilfolgen als „Subsequence Matching“ und Wort-Vorkommen-Statistiken
 - Ähnlicher Inhalt der Seiten → Zuordnung zum selben Cluster von Web-Seiten



Ähnlichkeit

- Link-basierte:
 - Links (Verbind.) zwischen Seiten erlauben Aussagen über inhaltlichen Zusammenhang
 - Basieren auf Hyperlinkstruktur eines Web Graph → verbundene Seiten sind ähnlich, gleicher Kern des Inhalts, selbe Autoren
 - Abgeleitet von Zitat-Analysen (Referenzen)



Suche

- Wirksamkeit:
 - Information Retrieval Techniken für schlüsselwortbasiertes Suchen
 - Qualität der Antworten durch Precision (Anteil relevanter Seiten) und Recall (Anteil aller relevanten Seiten) gemessen
 - Reindexieren → Suchmaschine aktuell



Suche

- Vergleich:
 - Viele Studien über Suchmaschinen und ihre Leistung bzw. Antwortverhalten
 - Relevanz der Ergebnisse für den Nutzer
 - Performance, Antwortzeit, Größe, Überlappung, Qualität, Ergebnismenge
 - z.B. Zufallsanfragen, URL-Indizierungstests



Verwendung

- Modellieren und Vorhersagen von Web-Seiten Zugriffen → Schätzungen
- Vorhersagen mit Hilfe alter Daten (Logs)
- Vorhersagen: Zugriffshäufigkeit, Browsing, Zeitpunkte, Beliebtheit von Web-Seiten
- Trends bei Seitenzugriff, Navigationsmuster → verbessert Organisation, Inhalt



Verwendung

- Web-Seiten Zugriffsmodellierung in Beziehung zur Vorhersage von effizientem Information Retrieval im WWW (Inhalt)
- Basis für verbesserte Qualität der Information, der Aufbereitung, des Zugriffs
- Adaptive, sich verändernde Web-Seiten können geschaffen werden



Informationstheorie

- Erwünschtheit (Informationen werden benötigt) und Überlebensfähigkeit (Zeitpunkt der Löschung bestimmter Teile) von Web-Dokumenten werden bestimmt
- Änderungsraten von Web-Seiten
- Alter der Web-Seiten bzw. der Informationen



Schluss

- Existenz verschiedenster Metriken für nahezu alle Perspektiven und Eigenschaften des Web und seiner Seiten
- Schaffung von Modellen zu Erklärung und Beschreibung von Zusammenhängen
- Führen zur Verbesserung des Web-Inhalts, der Organisation und des Aufbaus



Fragen?!