

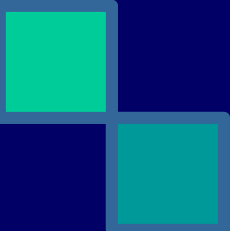

Zugriffsmechanismen und Standards für adaptive, personalisierte Dienste im Internet



Vorstellung des Konzeptes für die
Doktorarbeit von Wolfgang Wiese

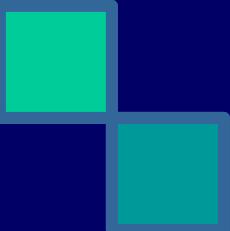



Überblick

- 
1. Einführung
 2. Anforderungen bzw. Lastenheft
 3. Beschreibung des Verfahrens
 4. Testfälle und Leistungsabschätzung für Suchmaschinen
 5. Zusammenfassung
- 

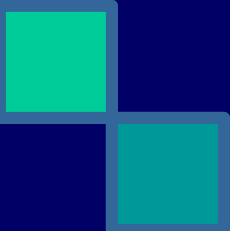



Einführung (1)

- 
- Wer sucht, der findet... oder auch nicht. Typische Suche mittels herkömmlicher Suchmaschine im WWW:
 - Eingabe eines Suchbegriffes kann zu mehreren Tausend Ergebnisseiten führen
 - Ergebnisse orientieren sich nur nach Wortsuche
 - Kontext eines Wortes meist unberücksichtigt
- 



Einführung (2)

- Wege aus dem Chaos:
 - Besser suchen durch logische Beziehungen
 - Suchmaschinen für spezielle Themengebiete
 - Kataloge
 - Mischformen
 - Personalisierung
- 
- 




Anforderungen (1)



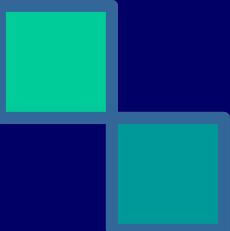

- Zielsatz:

„Wenn ein Informatiker nach dem Wort ‚Intelligenz‘ sucht, erhält er mehr Suchergebnisse aus dem Gebiet der AI. Wenn ein Biologe danach sucht, erhält er mehr Ergebnisse über biologische Gehirnfunktionen.“





Anforderungen (2)


- 
- Die Suche soll berücksichtigen:
 - Vorwissen des Suchenden: Das Benutzerprofil
 - Kontext des Suchbegriffes in einer Ergebnis-Seite
 - Kategorie des Suchbegriffes
- 



Anforderungen (3)



- Benutzerprofile


- Das Benutzerprofil soll automatisiert erstellt und verfeinert werden
 - Das Benutzerprofil wird transparent für den Benutzer verwaltet (z.B. als Browser-Plugin)
 - Benutzerprofile müssen anonymisierbar sein
 - Benutzerprofile verraten nicht mehr als nötig gegenüber Suchmaschinen
 - Profile werden nur dezentral gespeichert
- 



Anforderungen (4)

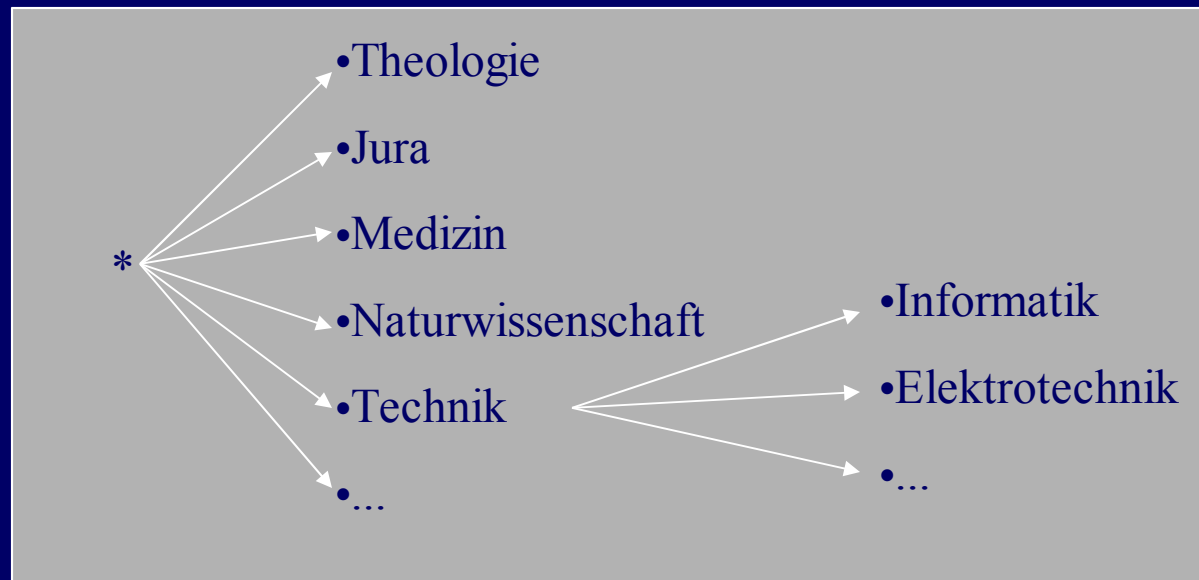


- Schnittstellen

- Es werden offene Standards verwendet
 - Daten werden über standardisierte Verfahren übertragen: XML, SOAP
 - Ergebnisseiten werden wie bisher für den Client zurückgegeben: HTML, XML, WML, ...
 - Austausch von Daten erfolgt „parallel“ zum normalen HTTP-Request/Response
- 


Beschreibung des Verfahrens

■ Definition Kategoriebaum



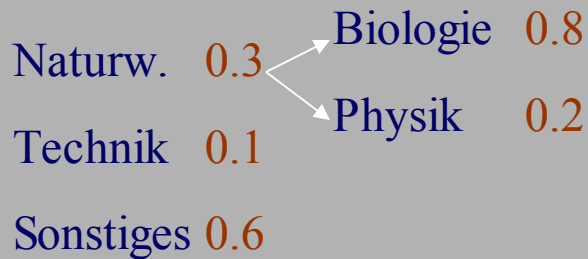


Verfahren (2)

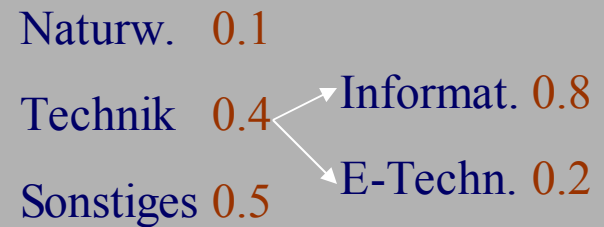
- Definition Benutzerprofil
 - Satz bestehend aus n Einträgen mit:
 - ID der Kategorie
 - Level: Zahl x mit $x \geq 0$ und $x \leq 1$
 - Timestamp: Letzte Modifikation des Eintrages
 - Anker: ID der Oberkategorie
 - Es gilt für alle n Einträge mit gleichen Anker:
 - $x_1 + x_2 + \dots + x_n = 1$
- 

Verfahren (3)

- Beispiel: 2 Profile mit minimalen Kategoriesatz:



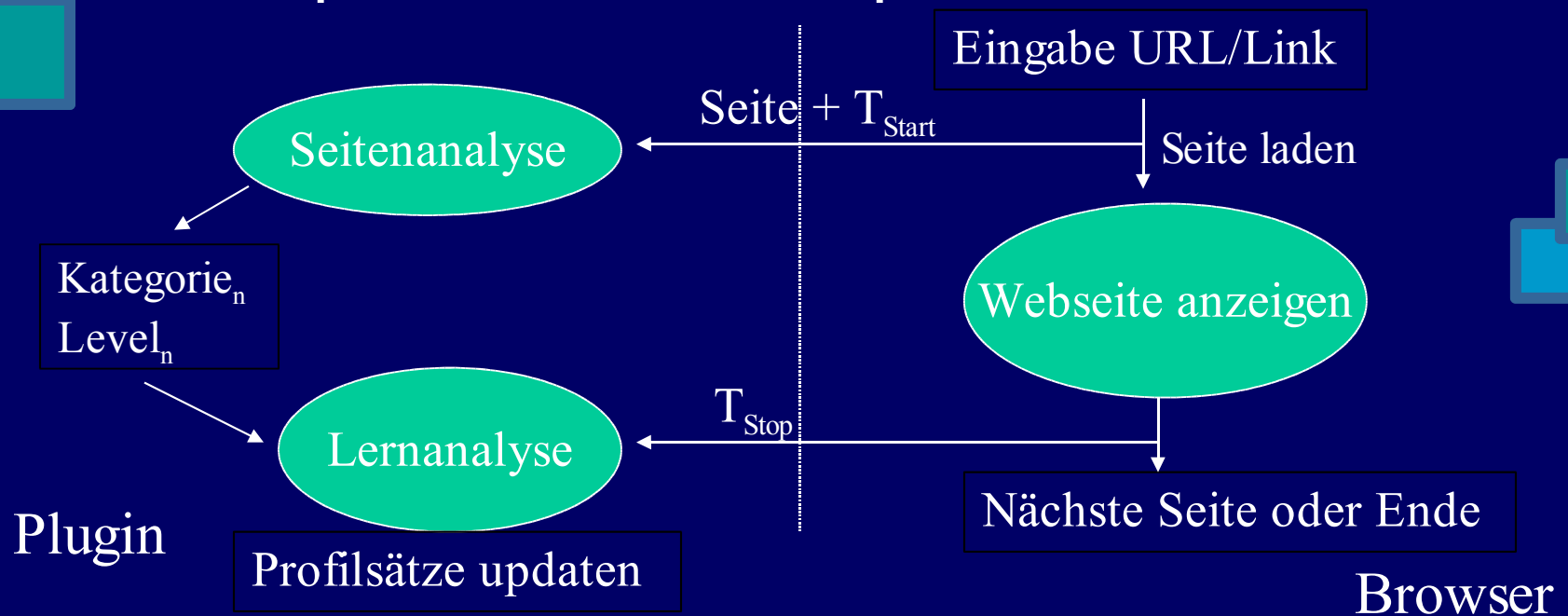
Biologe



Informatiker

Verfahren (4)

Adaption der Benutzerprofile






Verfahren (5)



- Lernanalyse

- Verschiedene Möglichkeiten:

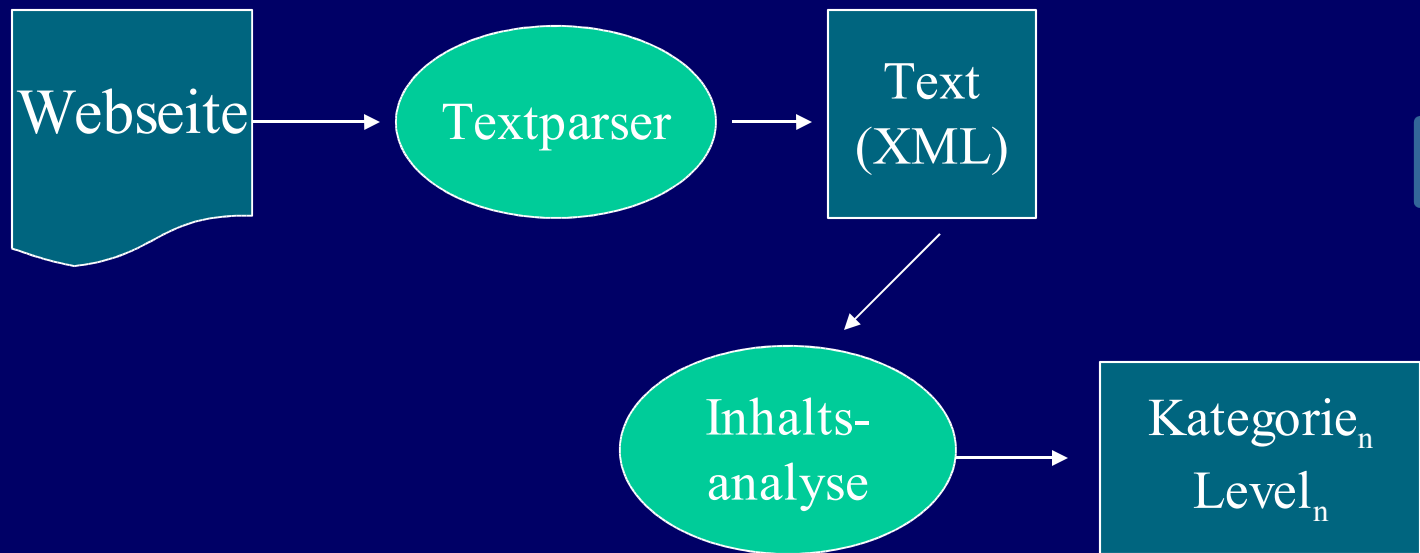
- Trivial: Was man sieht, lernt man
 - Verfahren aus Test-Theorie, KI, Mustererkennung
 - Sprachentests: C-Test, Sprachdiagnose, Marburger Sprachtest, usw.
- 

- Reliabilität

- Crombach-Alpha
- Spearman-Brown

Verfahren (6)

■ Seitenanalyse






Verfahren (7)




■ Inhaltsanalyse

- Text wird gemäß Sprachsyntax zu Einzelworte gesplittet.
 - Nur Objekte, Subjekte und Namen werden berücksichtigt
 - Wortanalyse auf bekannte Worte.
 - Kappung von Kategorie-Level $<$ Unschärfe
 - Mittelwertbildung über Ergebnisprofile aller Worte
 - Unbekannte Worte über Word-Digging oder Manuell abfragen
- 

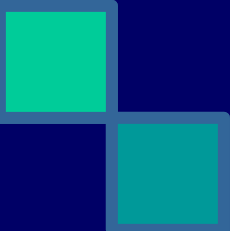



Verfahren (8)

- Wortanalyse und Wortdatenbank
 - Gesucht wird nur nach Subjekten, Objekten und Namen; Unbekannte Worte werden nach herkömmlicher Suche bearbeitet
 - Wortdatenbank (Client + Server):
 - $\text{Wort}_i, (\text{Kategorie}_n/\text{Level}_n)_i, \text{Timestamp}_i$
 - Wort-Kategoriezuordnung:
 - „Word-Digging“ (Peering)
 - Manuell
- 



Verfahren (9)

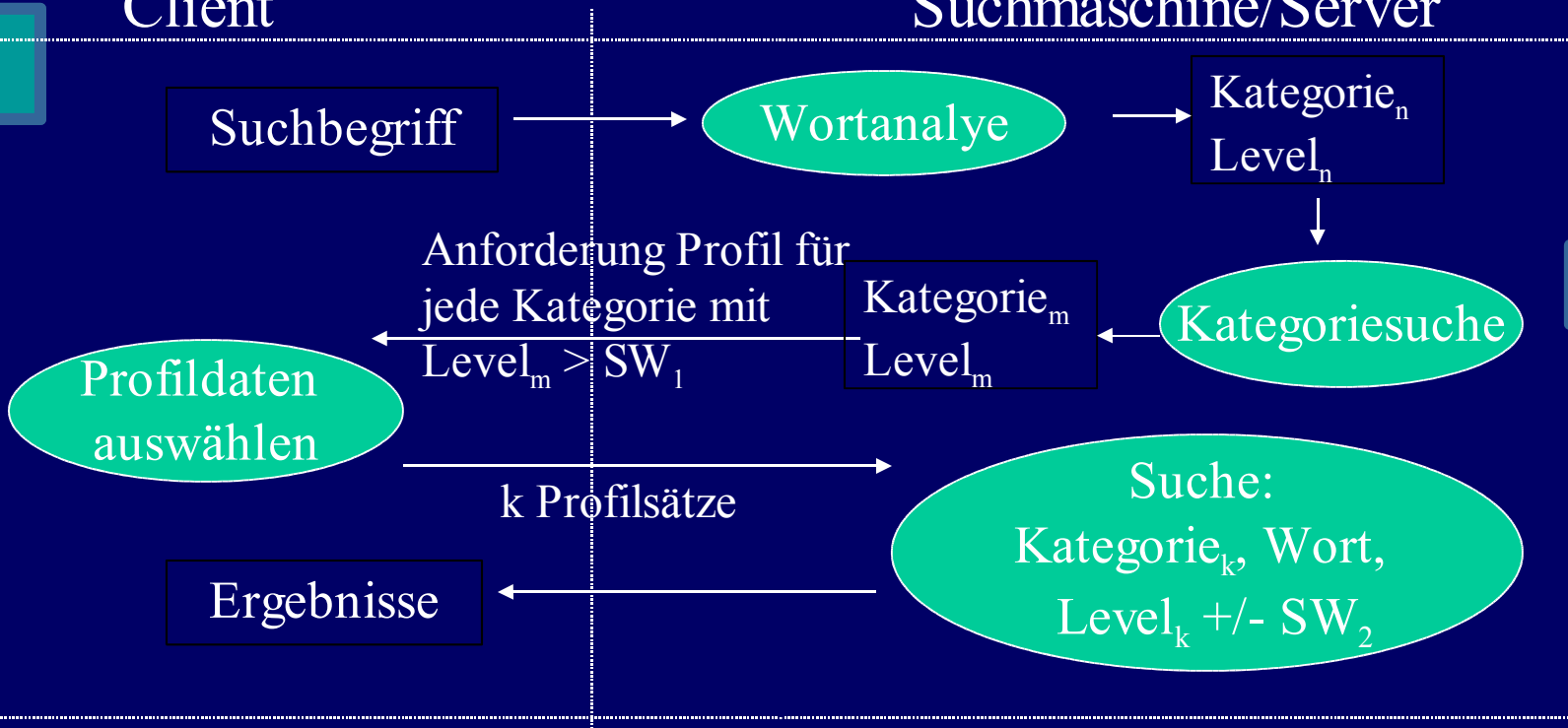
- 
- **Word-Digging: Dezentrale Verteilung von Wort-Kategorie-Zuordnungen mittels peer-to-peer Netzwerke**
 - Server (Suchmaschinen) erhalten im Gegenleistung zu Suchergebnissen Wort-Kategorie-Zuordnungen von unbekanntem Worten oder Worten mit einem alten Timestamp
 - Server tauschen ihre Wortdatenbanken untereinander aus
 - Clients können Server nach unbekanntem Worten fragen
- 

Verfahren (10)

Suchmaschine

Client

Suchmaschine/Server

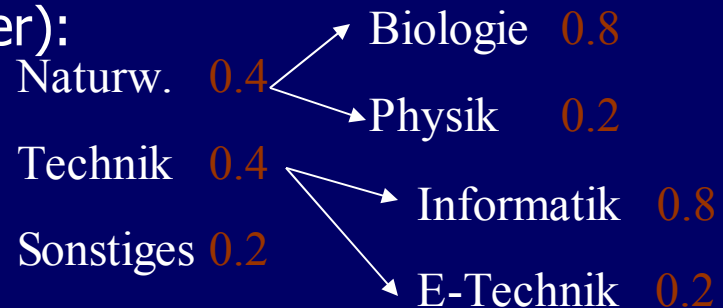


Verfahren (11)

■ Beispiel: Suche

■ Suchbegriff ‚Intelligenz‘

■ Wortanalyse (Server):

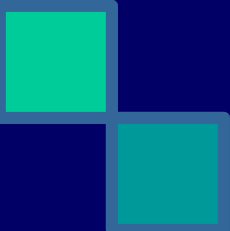



- Kategoriesuche (Server): Seiten mit allen Kategorie vorhanden. Ignoriere sonstige Kategorien

- Unschärfe 0.3: Frage Client nach Kategorien
Naturw. -> Biologie und Technik->Informatik

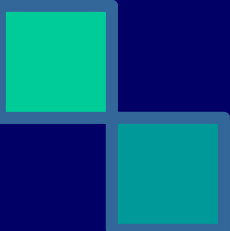



Verfahren (12)

- 
- Beispiel: Suche (cont.)
 - Userprofil „Biologe“ (Client):
 - Kategorie Naturw.: 0.3; Naturw->Biologie: 0.8
 - Andere Kategorien: $Level_n < \text{Schwellwert } 0.3$
 - Rückgabe: Naturw.->Biologie, (k=1)
 - Suchmaschine (Server):
 - Suche nach ‚Intelligenz‘ nur in Kategorie Naturw.->Biologie
- 

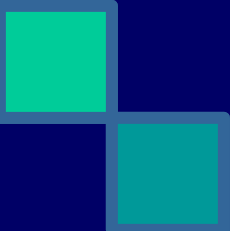



Testfälle

- 
- Ergebnisse obiger Verfahren müssen validiert werden:
 - Suche nach einer Liste von vorgegebenen Worten mit und ohne Benutzerprofile
 - Nachanalyse der Ergebnisse einer vorhandenen Suchmaschine: Vergleich des „Abfalls“ und der wirklichen Treffer
 - Suche nach gleichen Worten von unterschiedlichen Personengruppen
- 

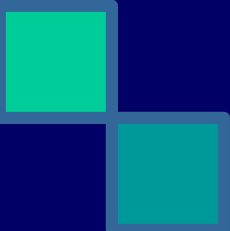



Zusammenfassung (1)

- 
- Anforderungen / Lastenheft definiert
 - Grundkonzeption
 - Erweiterungen / Verfeinerungen
 - Suche: Rückgabe von Kategorien nur mit $\text{Level}_n - U \leq \text{Level}_{\text{Suchmaschine}} \leq \text{Level}_n + U$
 - Kategoriebaum wie IP-Adressen verteilen:
„ID-Oberkat . ID-Unterkat . ID-Thema . ID-Fachgebiet“
- 



Zusammenfassung (2)

- 
- **Ausblick**
 - Implementation und Tests
 - Standards definieren / RFCs
 - OpenSource / Veröffentlichungen
 - Kontakte knüpfen
- 



Vielen Dank für Ihre Geduld



...Fragen offen?