

Automatische Schlagwortvergabe aus der SWD für Repositorien

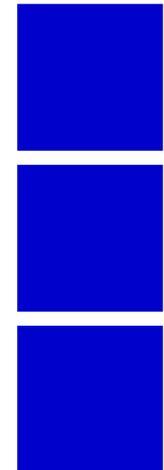
Ulrich Herb
Saarländische Universitäts- und Landesbibliothek (SULB)

Dr. Christoph Rösener
Institut der Gesellschaft zur Förderung der Angewandten Informationsforschung e.V. (IAI)



Die SULB - Aufgabenspektrum

- Universitätsbibliothek der Universität des Saarlandes (UdS)
- Sondersammelgebietsbibliothek für die Psychologie
- Landesbibliothek des Saarlandes



Die Repositories der SULB bis 2003

- **SOVA** – Saarbrücker Online Volltext Archiv
Beginn der Entwicklung: 1999
Ein Server für alle Aufgaben?
- DPI – Digitale Psychologie Information
Gemeinschaftsprojekt mit dem Zentrum für Psychologische Information und Dokumentation (ZPID), gefördert von DFG (2002 bis 2005)
u.a. Einrichtung des Psychologie-Volltextservers
PsyDok als separates Angebot für die psychologische Community



2003

Neustrukturierung der Server-Infrastruktur: Ein eigenes Repository pro Funktion

PsyDok – Service des SSG
Disciplinary Repository für die Psychologie
<http://psydok.sulb.uni-saarland.de>

SciDok – Service für die Universität
Institutional Repository der UdS
<http://scidok.sulb.uni-saarland.de>

SaarDok – elektronische Pflichtexemplare
Archivserver für landesspezifische Schriften
<http://saardok.bsz-bw.de/>



Technik

- **SciDok und PsyDok**
Software: OPUS, DINI-zertifiziert
-> PUSH-Service: Wissenschaftler spielen Dokumente ein
- **SaarDok**
Software: anfangs OPUS-basiert
später eine Instanz des BSZ-Online-Archivs BOA,
gehostet vom BSZ
-> PULL-Service: Bibliothekare harvesten Dokumente



Workflow

... in typischen Repository-Umgebungen (wie SciDok und PsyDok)

- Autor füllt Erfassungsmaske aus und erstellt dabei Metadaten
- Autor lädt Dokument zu den Metadaten auf den Server
- Metadaten und Dokument liegen im temporären Bereich des Repositoriums
- Bibliothekare überprüfen Dokument und Metadaten
- Bibliothekare schalten nach erfolgreicher Kontrolle das Dokument frei



Problem: Zahlreiche Repositories verlangen oder fordern eine sachliche Erschließung nach SWD

Leider ist den Autoren die SWD und ihre Nutzung unvertraut

Konsequenzen:

Autoren können das Dokument zwar sehr exakt beschreiben, aber ...

- sie vergeben unscharfe und zu allgemeine Schlagworte
- sie vergeben schlimmstenfalls falsche Schlagworte
- oder sie ignorieren die SWD und verwenden Begriffe, die in der SWD nicht abgebildet sind



Konsequenzen:

Bibliothekare verfügen über die notwendige SWD-Expertise und müssen die vorgenommene Verschlagwortung aufwändig nachbearbeiten oder – ohne Vorschläge durch Autoren – komplett selbst vornehmen, ...

können aber die Dokumente inhaltlich nicht so exakt beschreiben wie die Autoren.



Verallgemeinerbar?

Sommer 2007: SULB erhebt Status Quo unter Repository-Betreibern (OPUS-, DSpace-, EPrints- und MyCoRe-Systeme sowie Benutzer von Eigenentwicklungen)

Fazit:

Größtenteils analoge Erfahrungen ...

- Autoren vergeben falsche/unpassende Schlagworte
- (Nach-)Bearbeitung durch eigene Mitarbeiter
- Vergabe von SWD-Schlagworten im Repository wurde wegen Problemen/Aufwand eingestellt

...



Ist die Verwendung der SWD in Repositories sinnvoll?

- Sie garantiert die Erschließung aller Dokumente über ein fachübergreifendes, kontrolliertes Vokabular.
 - Sie sichert die nahtlose Integration der Metadaten in bibliotheksspezifische Retrievalsysteme und Anwendungen.
 - Auf aggregierter Ebene können Dokumente systematisch und unter Verwendung eines gepflegten und strukturierten Schemas recherchiert werden.
- > Vergabe von SWD-Schlagworten ist eine Empfehlung im Zertifikat für Dokumenten- und Publikationsservices der Deutschen Initiative für Netzwerkinformation e.V. (DINI)



Wünschenswert:

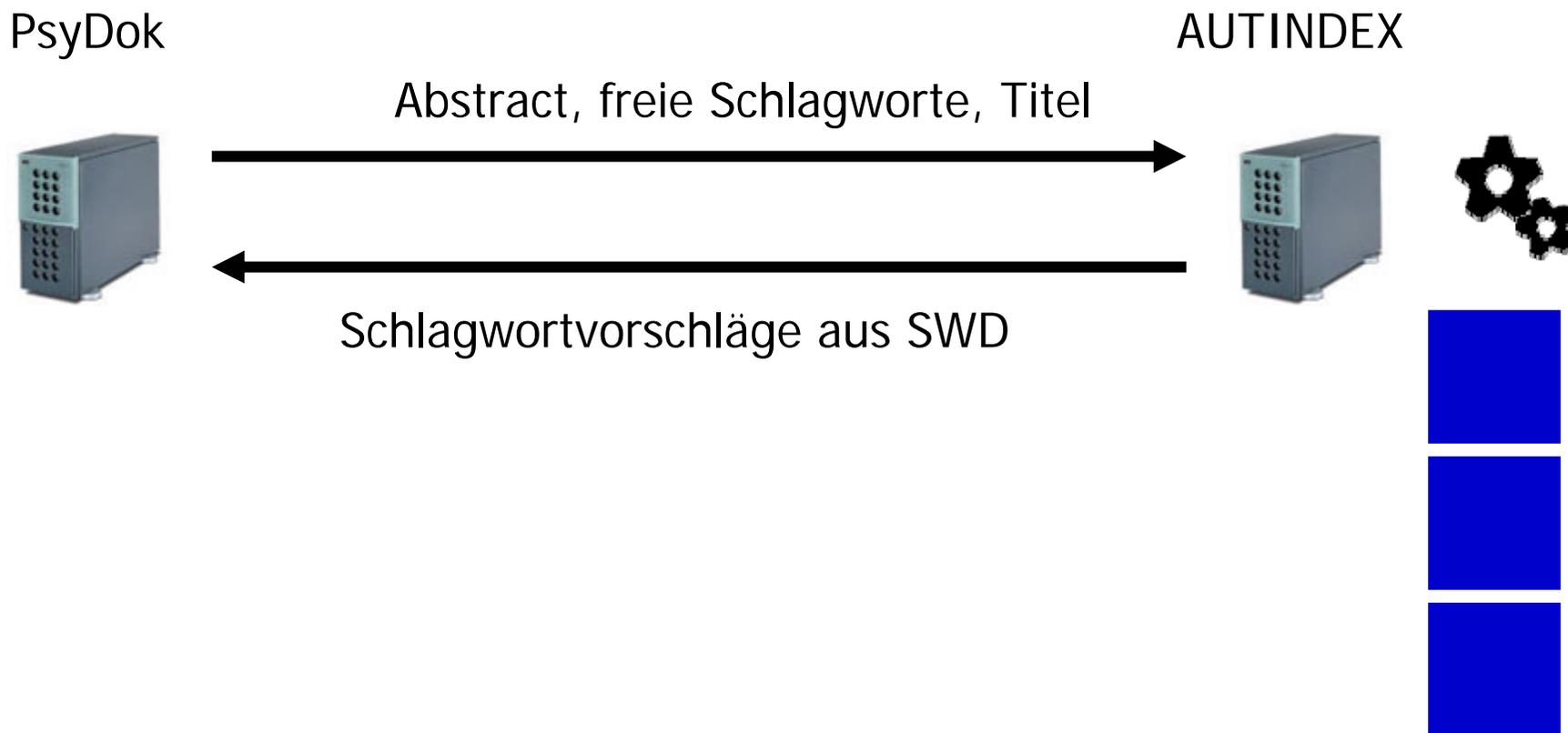
Technische Unterstützung der Autoren bei der Auswahl treffender SWD-Schlagworte, um

- a) über eine exakte Verschlagwortung ein optimiertes Retrieval zu ermöglichen und
- b) eine Erleichterung des Veröffentlichungs- und Bearbeitungsprozesses zu erreichen (easy submission)

Demo

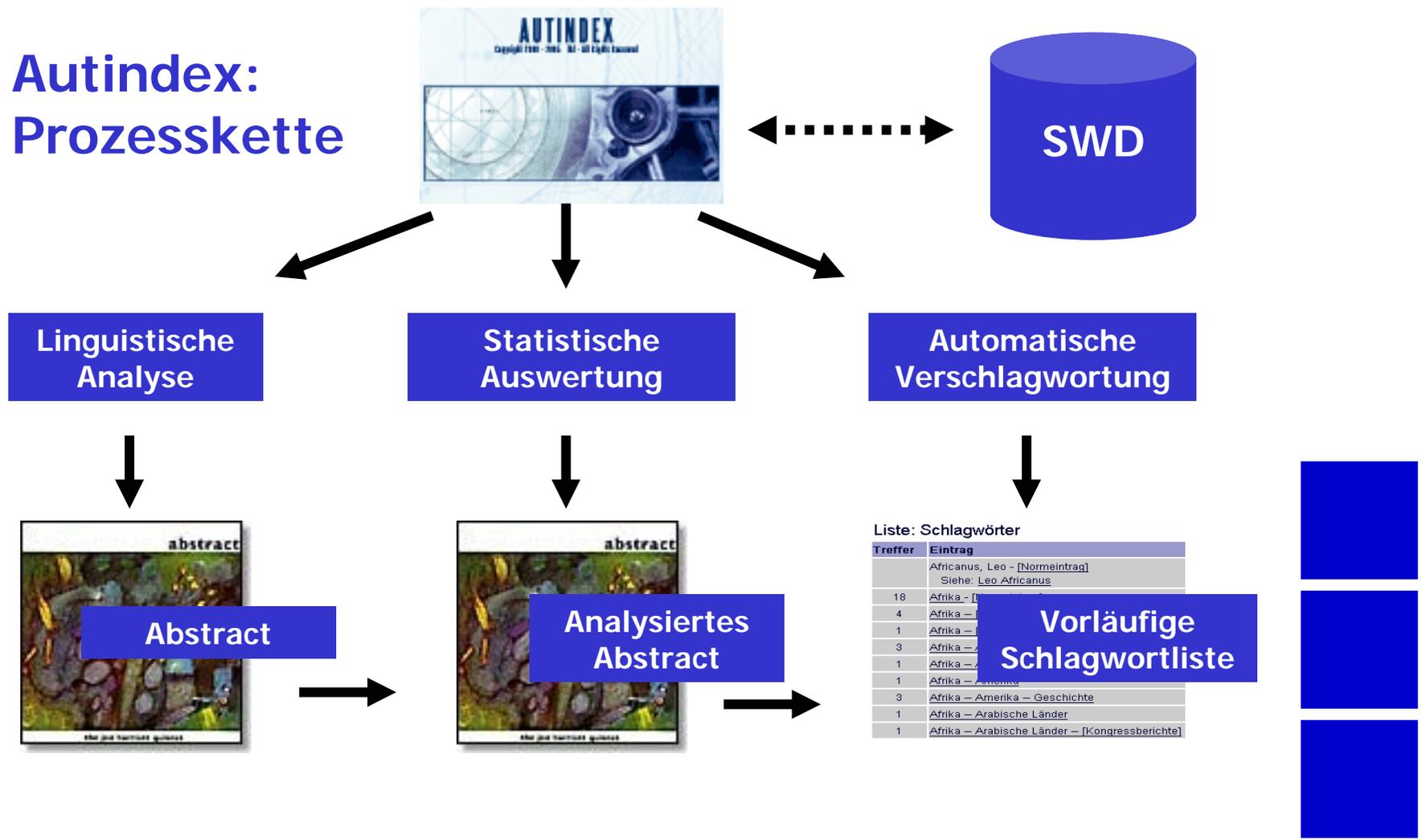


Was passiert?



Automatische Schlagwortvergabe aus der SWD für Repositorien

Autindex: Prozesskette



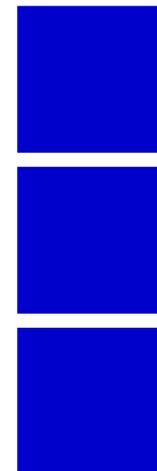
Autindex: Vorteile des linguistischen Verfahrens bei der Textanalyse

- Erkennung morphologischer Varianten (Wittgenstein / wittgensteinsche Sprachphilosophie, Haus / Häusern)
- Kompositazerlegung: Bessere Information über Semantik des Texts: Rentenversicherungsangestellter → Rente, Versicherung, Angestellter / Rentenversicherung, Versicherungsangestellter
- Erkennung syntaktischer Varianten: (Reduktion von/der Kosten / Kostenreduktion, Reduzierung von Kosten, Kostenreduzierung)
- Semantische Varianten (Synonyme: Müllwirtschaft / Abfallwirtschaft)
- Kombination mit syntaktischen Varianten: (Kostenverminderung / Verminderung von Kosten / Einsparungen der Kosten)



Geplante Realisierung

Im Rahmen eines gemeinsamen bei der Deutschen Forschungsgemeinschaft DFG beantragten Projekts des Institutes der Gesellschaft zur Förderung der Angewandten Informationsforschung e.V. an der Universität des Saarlandes ([IAI](#)) und der Saarländischen Universitäts- und Landesbibliothek ([SULB](#))



Vielen Dank für ihre Aufmerksamkeit

Ulrich Herb
u.herb@sulb.uni-saarland.de

Dr. Christoph Rösener
chrisr@iai.uni-sb.de

