

BPS-Suche

Akzeptanzworkshop

*„Optimierung von Suchmöglichkeiten in Modulen zum
E-Learning von grundständigen Bachelor- und
Master-Studiengängen auf der Basis von Dublin
Core Metadaten“*

2004-12-16

Prof. Dr. G. Heyer,

Dipl.-Inf. M. Barth, Dipl.-Inf. P. Mairif

Übersicht

- ◆ Anforderungen
- ◆ Beispiel
- ◆ Technischer Hintergrund
- ◆ Perspektiven

Anforderungen: Zielgruppen

- ◆ Lehrender
 - ◆ Hochschulübergreifende Nutzung von Lehrmaterial
 - ◆ Vergleichbarkeit von Modulen
 - ◆ Vermeidung von Doppelarbeiten
- ◆ Lernender
 - ◆ Suche nach Dokumenten
 - ◆ Suche nach Informationen

Anforderungen: Relevanz

- ◆ Inhaltliche Priorisierung
 - ◆ Extraktion und Bewertung relevanter Stichwörter
- ◆ Granularität der Ergebnisse
 - ◆ Drei-Ebenen-Hierarchie

Beispiel

- ◆ Suche nach *Soziologie, Organisations-systeme*
- ◆ Google: 180-seitiges Dokument als Treffer





Einführung in die Soziologie

Skript

Prof. Dr. Hajo Weber
Universität Kaiserslautern
FG Soziologie

Beispiel: BPS

The screenshot shows a Mozilla browser window titled "Mozilla <2>". The main content area displays the "BPS Search Engine" interface. At the top, there is a search bar containing the text "Organisationssysteme" and a "search" button. To the right of the search bar are several radio buttons: "pages" (selected), "whole documents", and "whole units". Further right are two dropdown menus, both set to "all types" and "all languages". Below the search bar, the text "search results - there are 2 results" is displayed. A table with two columns, "unit->document->part" and "content", lists the search results. The first result is "Einfuehrung_Soziologie -> Skript_einf_Soziologie.pdf -> 2" with the content "Gliederung Begriff der Soziologie Entwicklung der ...". The second result is "Einfuehrung_Soziologie -> Skript_einf_Soziologie.pdf -> 117" with the content "Organisation als geschlossenes System selbstrefere...". The browser's status bar at the bottom shows "Done" and various navigation icons.

Organisationssysteme search pages whole documents whole units all types all languages

search results - there are 2 results

unit->document->part	content
Einfuehrung_Soziologie -> Skript_einf_Soziologie.pdf -> 2	Gliederung Begriff der Soziologie Entwicklung der ...
Einfuehrung_Soziologie -> Skript_einf_Soziologie.pdf -> 117	Organisation als geschlossenes System selbstrefere...

Mozilla <2>

document view

prev next

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60
61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90
91	92	93	94	95	96	97	98	99	100	101	102	103	104	105	106	107	108	109	110	111	112	113	114	115	116	117	118	119	120
121	122	123	124	125	126	127	128	129	130	131	132	133	134	135	136	137	138	139	140	141	142	143	144	145	146	147	148	149	150
151	152	153	154	155	156	157	158	159	160	161	162	163	164	165	166	167	168	169	170	171	172	173	174	175	176	177	178	179	180

7.3 Organisation als offenes/geschlossenes System (selbstreferentielles, autopoietisches System)

- Folge: Keine direkte Verbindung
zwischen Umwelt/Organisation
Bearbeitung nach organisationsinternen
Operationen
- **Autopoietische Systeme**
 - Organisationssysteme bestehen aus Elementen
z.B. Entscheidungen
 - Organisationen erzeugen diese selbst

Funktionalität

- ◆ Generierung von Meta-Daten (Schlüsselwörter)
- ◆ Nutzung ex. Meta-Daten
- ◆ Suche in
 - ◆ Meta-Daten
 - ◆ Volltext
- ◆ „Zielgenaue“ Suchergebnisse

Funktionalität (2)

- ◆ Suche über
 - ◆ Standorte/ Bildungseinrichtungen
 - ◆ Sprache der Inhalte (Deutsch, Englisch, ...)
 - ◆ Dokumentarten: Folien, Paper, *Übungsaufgaben*, ... (im Gegensatz zu Google)

Mozilla <3>

BPS Search Engine

pages
 whole documents
 whole units
 slides

search results - there are 122 results

unit->document->part	score	content
ESSUM01A.pdf -> 3	284	Prof Dr Ing Wilhelm Spruth Raum Hauptgebäude Uni L...
schwarz_diss -> follen.pdf -> 1	72	erzeugte Modelle logischer Programme Sibylle Schwa...
rnvs_lehre_praktikum_kom -> serie_iv.pdf -> 21	58	Aufgabenserie Aufgabe Recherchieren und charakt...
ESSUM01A.pdf -> 4	56	Gliederung der Vorlesung Wirtschaftliche und techn...
go.pdf -> 15	20	Informatik Uni Leipzig Nachbearbeitung von Datein ...
go.pdf -> 20	18	Informatik Uni Leipzig Ein Beispiel mit verlaufend...

Mozilla <3>

Mozilla <2>

BPS Search Engine

pages
 whole documents
 whole units
 script

search results - there are 1178 results

unit->document->part	score	content
ordnungen -> inf-2nf.pdf -> 8	4005	Anlage zur Studienordnung Nebenfach Informatik Stu...
ordnungen -> inf-2hf.pdf -> 8	2580	Weitere Bestimmungen Studienangebot Das Studienang...
theo_cps -> absberV2H.pdf -> 66	2112	Anlagenverzeichnis Diplomprüfungsordnung Informati...
theo_cps -> absberV2H.pdf -> 30	1760	Verlauf Ergebnisse In Bild ist als Beispiel das mo...
ordnungen -> inf-2nf.pdf -> 5	1740	Das Grundstudium wird durch die Zwischenprüfung da

Mozilla <2>

Technischer Hintergrund

- ◆ Idee: Hierarchischer Index
- ◆ Architektur
- ◆ Grundlagen: Terminologie Extraktion
- ◆ DC-Meta-Daten

Hierarchischer Index

Kollektion I

Dokument a

Seite 1

Seite 2

Dokument b

Seite 1

Seite 2

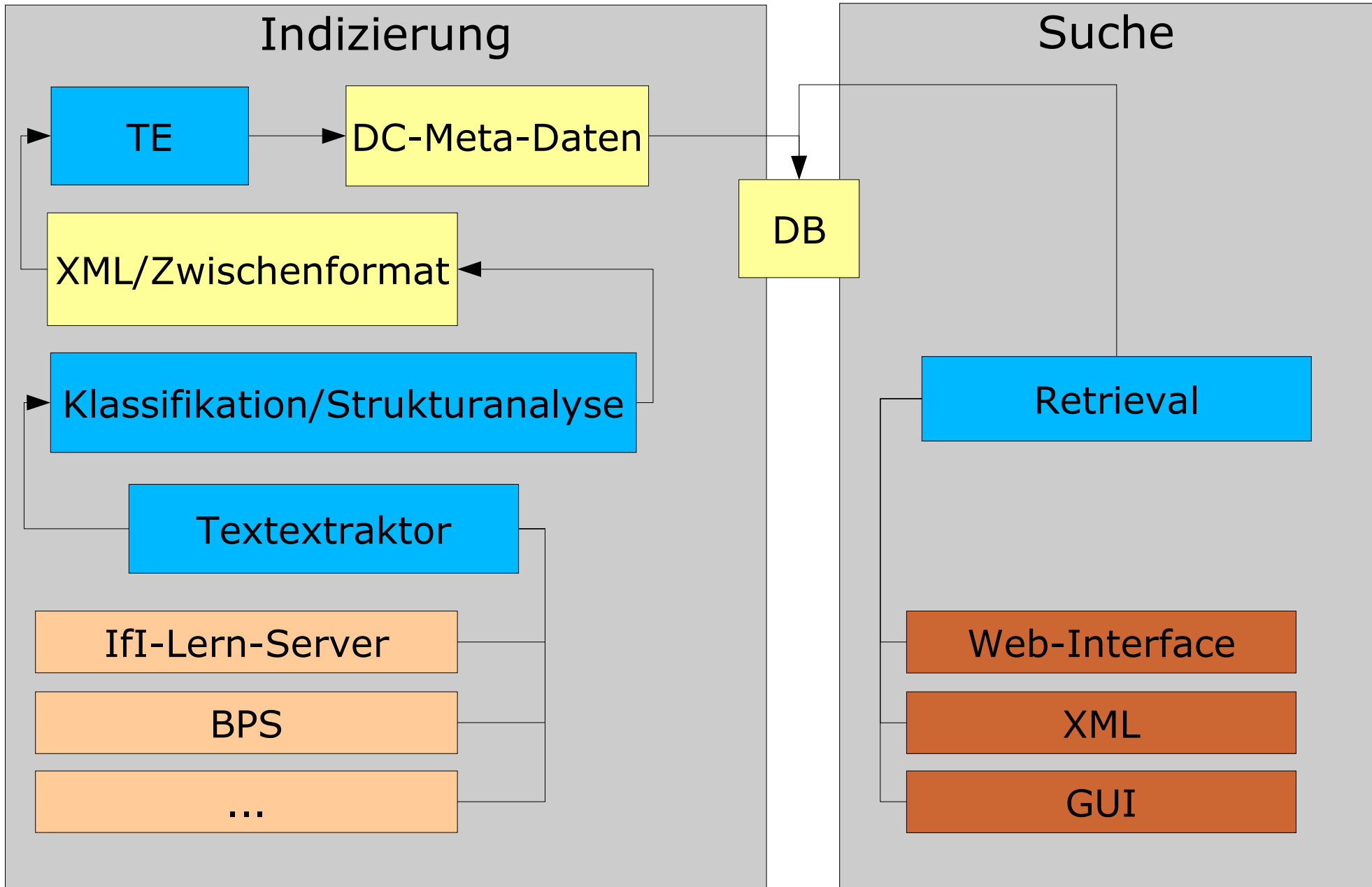
- Algorithmen & Datenstrukturen *Lerneinheiten*
 - Sortieralgorithmen *Dokumente*
 - Quicksort *Folien/Seiten*
 - Heapsort
 - Bubblesort
 - Kompressionsalgorithmen
 - ...
- Computerlinguistik
 - ...

Hierarchischer Index: Indizes

- ◆ Drei-Ebenen-Hierarchie
- ◆ Drei Invertierte Listen
 - ◆ mit Gewichtung

Term	Dokument-ID	Prüfgröße	Frequenz
...			
Eigenschaft	85	787	87
Psychologie	85	1501	129
Intelligenz	85	1650	144
Differentiell	85	2079	109
...			

Architektur



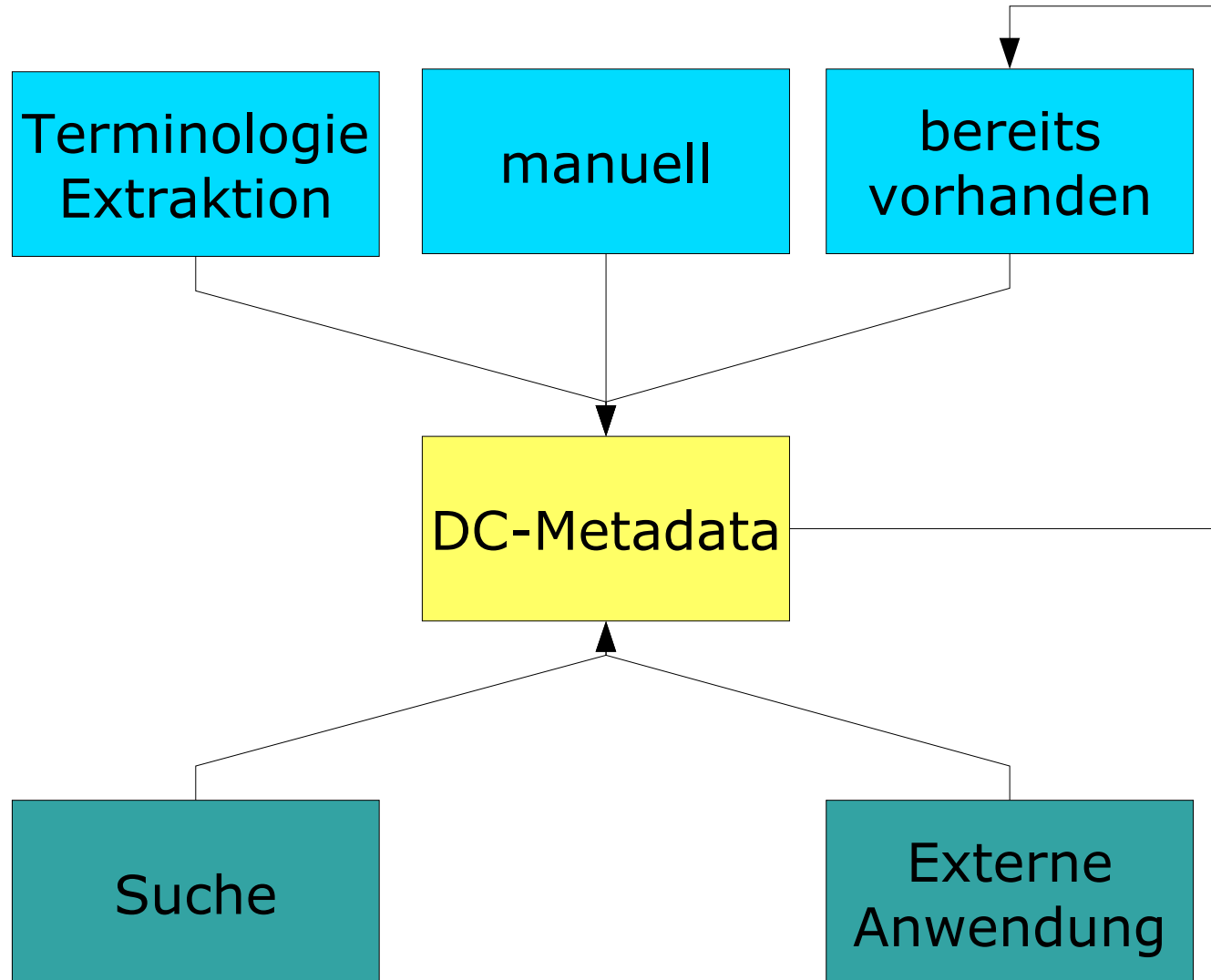
Terminologie Extraktion

- ◆ Grundlage: Liste von 6Mio Wörtern (Grundformen) mit Frequenzangaben bzgl. eines 542Mio Wörter Korpus
- ◆ Differenzanalyse
- ◆ Ergebnis: Relevante Begriffe für jeweiligen Textabschnitt

Dublin Core Metadata

- ◆ Dublin Core Metadata Element Set V1.1
 - ◆ Title
 - ◆ Creator
 - ◆ Subject
 - ◆ Description
 - ◆ ...
- ◆ Kodierung in
 - ◆ HTML
 - ◆ XML/RDF

Meta-Daten



Schnittstellen

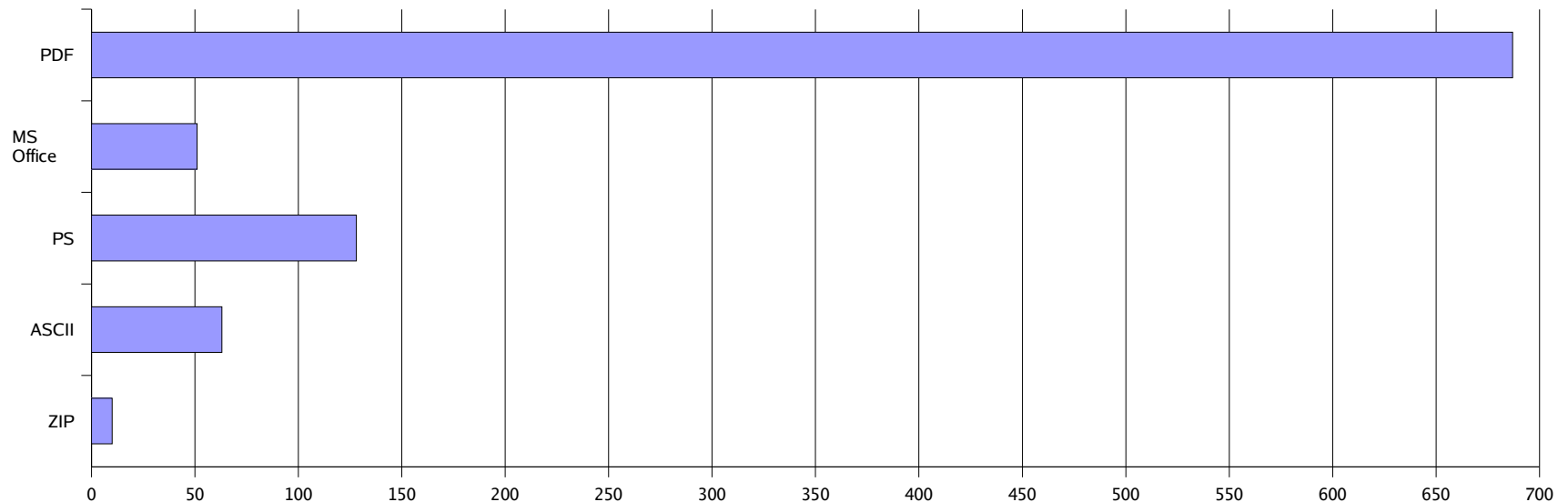
- ◆ implementiert:
 - ◆ Kommandozeile
 - ◆ Java-API
 - ◆ Corba
- ◆ geplant:
 - ◆ Web-Service (bzw. einfach HTTP/XML)
- ◆ möglich:
 - ◆ z. B. RMI

Indizierung des Lernservers

◆ Indiziert

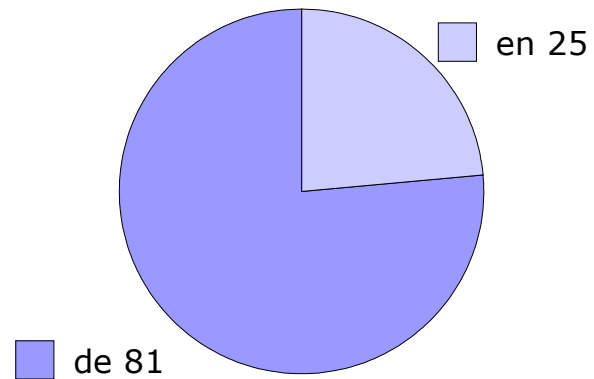
- ◆ Kollektionen 105
- ◆ Dokumente 634
- ◆ Seiten 7676

◆ Formate

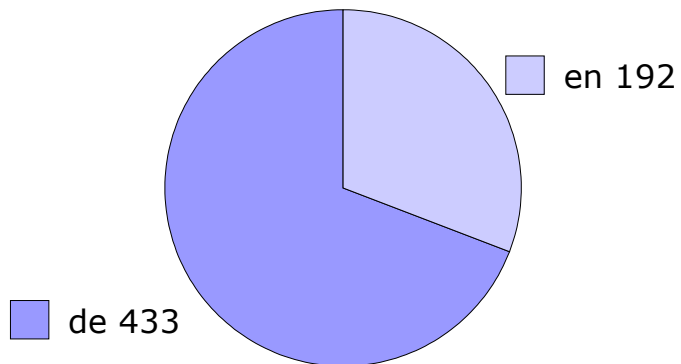


Lernserver: Sprachen

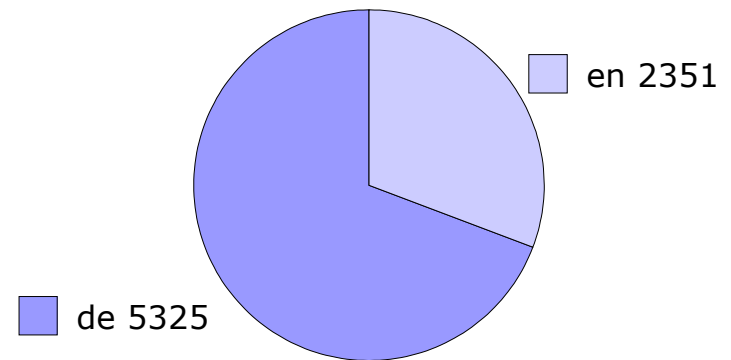
Kollektionen



Dokumente



Seite



Fazit

- ◆ Semantische Mark-Ups helfen die Suche in unstrukturierten Textdokumenten zu verbessern
- ◆ *Dublin Core* beinhaltet eine Menge einfacher Metatags
- ◆ Die Werte einzelner DC Metatags können automatisch berechnet werden
- ◆ Dadurch kann auf Seitenebene in unstrukturierten Textdokumenten gesucht und die Effektivität der Suche verbessert werden

Perspektiven

- ◆ Stabilisierung
- ◆ Performanz-Optimierung
- ◆ neue Terminologie Extraktion
- ◆ Schnittstellen
- ◆ weitere Dateiformate (PS, Office, ...)
- ◆ weitere Dokumentarten (z. B. Übungsaufgaben)
- ◆ Dokument-Struktur-Analyse (Kapitel, etc.)
- ◆ BPS-Integration
- ◆ Management-Funktionalität
- ◆ Evaluation