

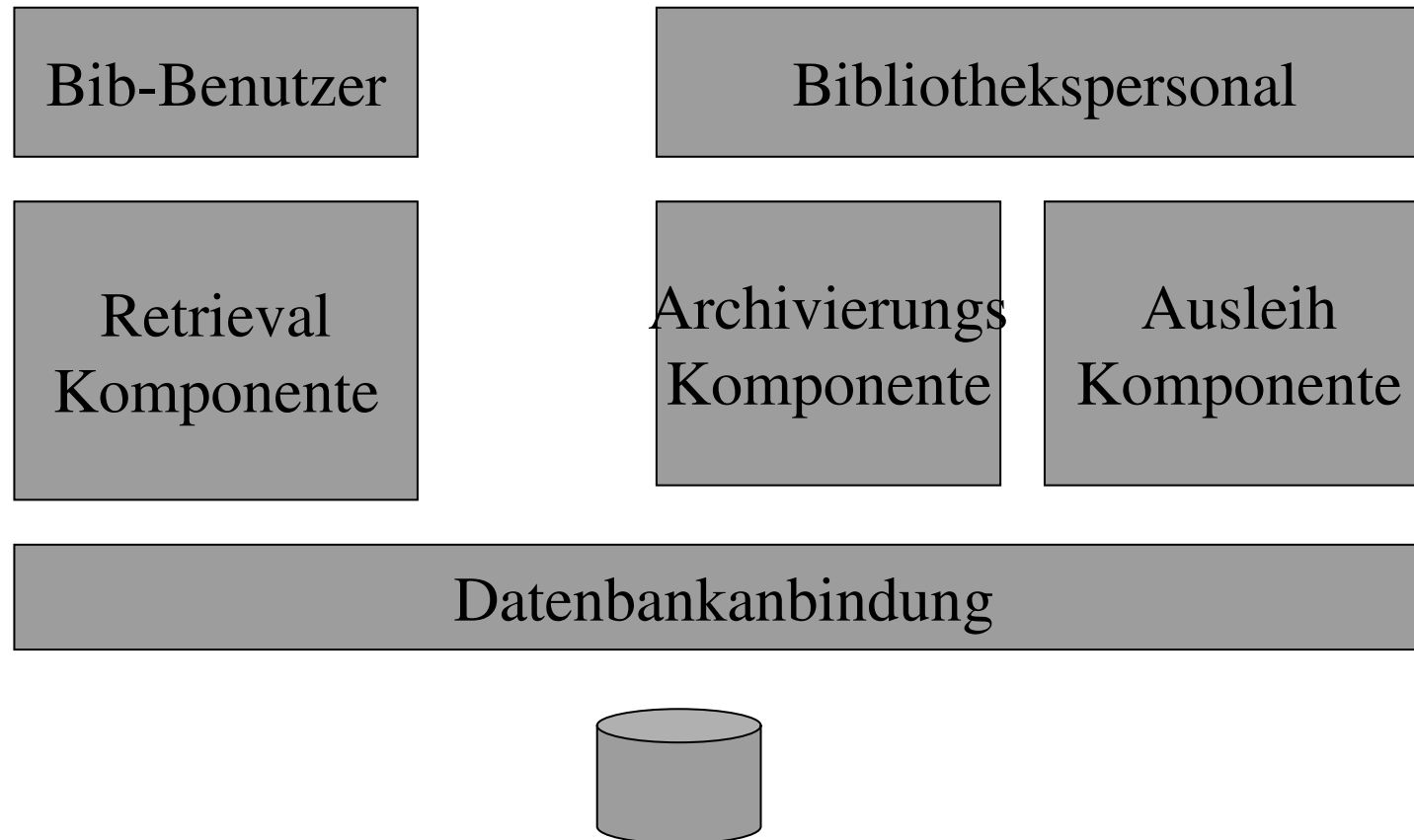


Übung Datenbanksysteme
Entwurf eines
Bibliothekssystems

9.2.2004



Architektur des Systems



Eine oder mehrere DBS?

- Probleme:
- Konsistenz zwischen Ausleihe/Recherche/Archivierung
- Erfasste Bücher automatisch ausleihbar?
- Probleme bei verteilten Archivierungsteams?

Archivierungsprozeß

- Institutionen erfassen nur einen kleinen Teil der Literatur (fast ausschließlich Bücher)
- Katalogisierungsprozeß ist manuell und teuer (vor ca. 10 Jahren bereits DM 80 pro Buch)
- Einzelne Artikel in Journalen und Magazinen werden nicht einzeln erfaßt, dadurch auch nicht recherchierbar.

Neuer Ansatz

- Nicht herkömmlichen Prozeß übernehmen, sondern:
- Scannen
- OCR
- Volltext-Datenbanken
- Digitale Dokumente
- Web-Interface

Katalogisierungsprozeß

- Dokumente
 - Strukturfelder (Autor, Titel, Verlag, Ausgabe, etc.)
 - Volltexte (gespeichert und indexiert)
 - Nach Möglichkeit gescannte Seiten der Dokumente oder vollständige Dokumente in Postscript oder PDF

Katalogisierungsprozeß (2)

- Maschinelles Austauschformat Bibliotheken MAB
- Autoren*
- Titel
- Untertitel
- Verleger
- Jahr
- Ort
- Zeitschrift
- Band*
- Heft
- erste Seite
- letzte Seite
- Fingerprint ...
- Ca. 950 einzelne Felder (!)

Scannen


- Darf Material gescannt werden?
- Kompressionstechniken für Bilder?
- 1x komprimiert bei Erfassung
- X-fach dekomprimiert beim Anfragen
- PNG, JBIG, PDF erzeugen
- Geschwindigkeit: Dekomprimieren nicht länger als 2 Sekunden

OCR

Warum OCR?

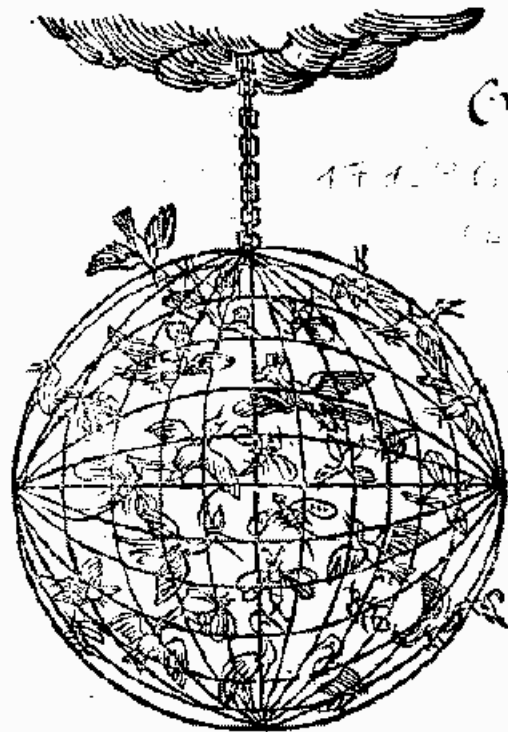
Titelseite, Inhaltsverzeichnis und erste Inhaltseiten
enthalten Schlagwörter

Aber:

- Text im allgemeinen fehlerhaft
 - Deshalb Verwendung eines Wörterbuchs
 - Stammreduktion der Wörter
 - Abgleich des gescannten Textes mit Wörterbuch
 - Wenige Korrekturvorschläge
- 

Quotlibetisches
Welt=Kestig:

Darinnen das gegenwertige Weltgetummel/
Schümmel und Getrümel/ Wüthen und Toben/
Tren/ Wirren und Sinceriren, Liegen/ Trogen und
Kriegen/ gleich als in einem Spiegel zusehen.
Kidentem dicere verum, nil verat, & prodest.



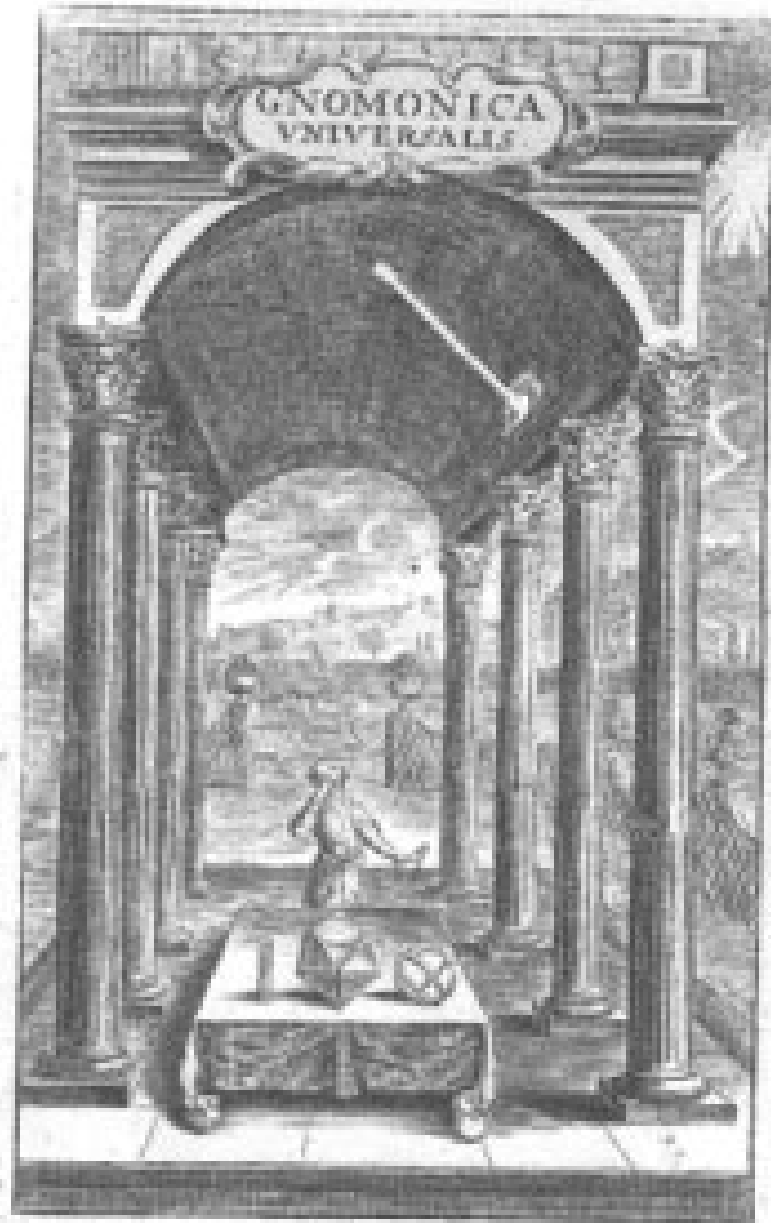
(v. p. 229. 7.)

1711

(12)

Interdum genus est peritiæ, vitare, quod doctis placeat.

Druckt im Großen Wunder Jahr.



Nicht technische Probleme

- Copyright der eingescannten Seiten
- Weiterverteilung der Volltexte
- Bibliothekare überflüssig?

Weitere mögliche Dienste

- Bestellservice für lokal nicht vorhandene Artikel/Bücher/etc.
elektronische Lieferung: Problem des Copyrights
- Statistische Auswertungen des Bestandes, Veröffentlichungshäufigkeit, Ausleihhäufigkeit

Weitere Dienste (2)

- Referenzen und Zitate betrachten.
- Entweder automatisch Erfassen oder manuelle Hinweise/Annotationen erlauben.
- Automatische Erfassung bei der Archivierung: <http://citeseer.nj.nec.com/cs>
Archivierung durch Harvesting
- Bei manuellen Hinweisen: Trennung von Dokumenten und Annotationen

Klausurrelevantes

- Wann kann nur „REFERENCES ... ON DELETE ...“ angewandt werden (1), wann „CONSTRAINT ... FOREIGN KEY ... REFERENCES ...“ (2)?

(2) Bei zusammengesetzten Schlüsseln, wenn Attribute unterschiedlich benannt sind.

Klausurrelevantes

- Unterschied:

(1) „... WHERE Kontonr IN ...“

(2) „... WHERE Kontonr = ...“

Zu (1) „IN“ entspricht „Element von ...“, also Menge auf der rechten Seite (kann auch einelementig sein)

Zu (2) prüft einen Wert

Klausurrelevantes

- Vorgegebene Tabelle: Woran erkennt man funktionale Abhängigkeiten? Sind es nur zufällige Wertebelegungen?

Semantik der Attribute ist wichtig!

Belegung läßt nur Vermutungen zu (siehe z.B. Data-Mining)

Klausurrelevantes

- Kann man das „HAVING“ Prädikat auch in die WHERE-Klausel schreiben?
- Nein, weil sich HAVING auf Gruppen bezieht und eine Nachfilterung darstellt.
- Siehe auch Zusatzblatt zu Blatt 7 (jetzt ist der Zusatz wirklich dort, sorry...)