

# Recommender Systems for (Scientific) Libraries.



Version: February 19, 2008

## **Andreas Geyer-Schulz**

Schroff-Chair

Information Services and Electronic Markets

Faculty of Business Engineering

Building 20.20 Computing Center, Level 1, Room 169

Universität Karlsruhe (TH)

D-76128 Karlsruhe

Phone: +49-721-608-8401 Fax: +49-721-608-8403

E-Mail: [Andreas.Geyer-Schulz@em.uni-karlsruhe.de](mailto:Andreas.Geyer-Schulz@em.uni-karlsruhe.de)

Contact arranged by E-Mail

<http://www.em.uni-karlsruhe.de/>

# Contents

<b>1</b>	<b>Menu:</b>	<b>5</b>
<b>2</b>	<b>Motivation.</b>	<b>6</b>
<b>3</b>	<b>Project goals.</b>	<b>7</b>
3.1	Example: Detailed View of Documents (UBKA) . . . . .	8
3.1.1	Example:List of Recommendations (UBKA) . . . . .	9
3.1.2	Example:List of Recommendations (UBKA) . . . . .	10
3.2	User Interface: Library OPAC Search (KVK) . . . . .	11
3.2.1	User Interface: Hit List (KVK) . . . . .	12
3.2.2	User Interface: Document Inspection Page (KVK) . . . . .	13
3.2.3	User Interface: List of Recommendations (KVK) . . . . .	14
<b>4</b>	<b>Scientific roots and theories combined.</b>	<b>15</b>

4.1	Paul A. Samuelson: Revealed Preference Theory (Nobel Prize in Economics 1970)	17
4.2	A. Michael Spence and Joseph Stiglitz: Market Signalling and Screening by Self-Selection (Nobel Prize in Economics 2001)	20
4.3	Modelling as a “Gestaltproblem”: Causal models versus noise models	21
4.4	Ideas from ensemble methods.	22
4.5	Ehrenbergs Repeat-Buying Theory.	23
<b>5</b>	<b>Statistical models.</b>	<b>24</b>
5.1	The LSD Model: A Mixture of Poisson Processes	25
5.1.1	Stochastic Processes	26
5.1.2	Generating Recommendations	27
5.2	Multinomial models: Small samples	28
5.2.1	Distributions of Co-Inspections (KVK Data)	29
5.2.2	Data Setup	30

5.2.3	General Concept . . . . .	31
5.2.4	An Ideal Decision Maker . . . . .	32
5.2.5	POSICI: Probability of Single Item Co-Inspections . . . . .	33
5.2.6	Inspection Probabilities for Growing $k$ and $n$ in POSICI . . . . .	34
5.2.7	Interpretation of POSICI . . . . .	35
5.2.8	POMICI: Probability of multiple items co-inspections . . . . .	36
5.2.9	Partition Probabilities for Growing $k$ and $n$ in POMICI . . . . .	38
5.2.10	Interpretation of POMICI . . . . .	39
<b>6</b>	<b>Other applications (prototypes).</b>	<b>40</b>
6.1	Restricted random walk clustering for indexing . . . . .	41
6.2	Adaptive recommender systems . . . . .	42
6.3	Exploration of the information space . . . . .	45
<b>7</b>	<b>System architecture and scalability.</b>	<b>46</b>
7.1	An Agency of Software Agents for Distributed Recommender Services . . . . .	47

7.2	Recommender Services for Legacy Library Systems I . . . . .	48
7.3	Recommender Services for Legacy Library Systems II . . . . .	49
<b>8</b>	<b>Service integration into web OPACS of scientific libraries.</b>	<b>50</b>
8.1	Integration Variant I: apache-logs, cgi-scripts . . . . .	51
8.2	Integration Variants II and III: Web 2.0 Technology . . . . .	52
<b>9</b>	<b>Development of a service concept and a service organization.</b>	<b>53</b>
<b>10</b>	<b>Development of a global service infrastructure for (scientific) li- braries.</b>	<b>54</b>
10.1	Reference Customers . . . . .	55

# 1 Menu:

- Project goals and motivation.
- Scientific roots and theories combined.
- Statistical models.
- Other applications (prototypes).
- System architecture and scalability.
- Service integration into web OPACS of scientific libraries.
- Development of a service concept and a service organization.
- Development of a global service infrastructure for (scientific) libraries.

(Context: Recommender services.)

## 2 Motivation.

- Scientists and students are more and more incapable of efficiently finding relevant literature in conventional database oriented catalog systems.
  - Problems: Supply complexity, evaluation of the quality, information overload
  - Typical literature research path: Asking peers
- More sophisticated access paths to prevent information overload are needed.
- Many universities are teaching a growing number of students with a more or less fixed staff size.
  - Students, university teachers and researchers could free some of their valuable time that is needed now to recommend each other e. g. standard literature of their fields which could be done easily by behavior-based expert advice services.

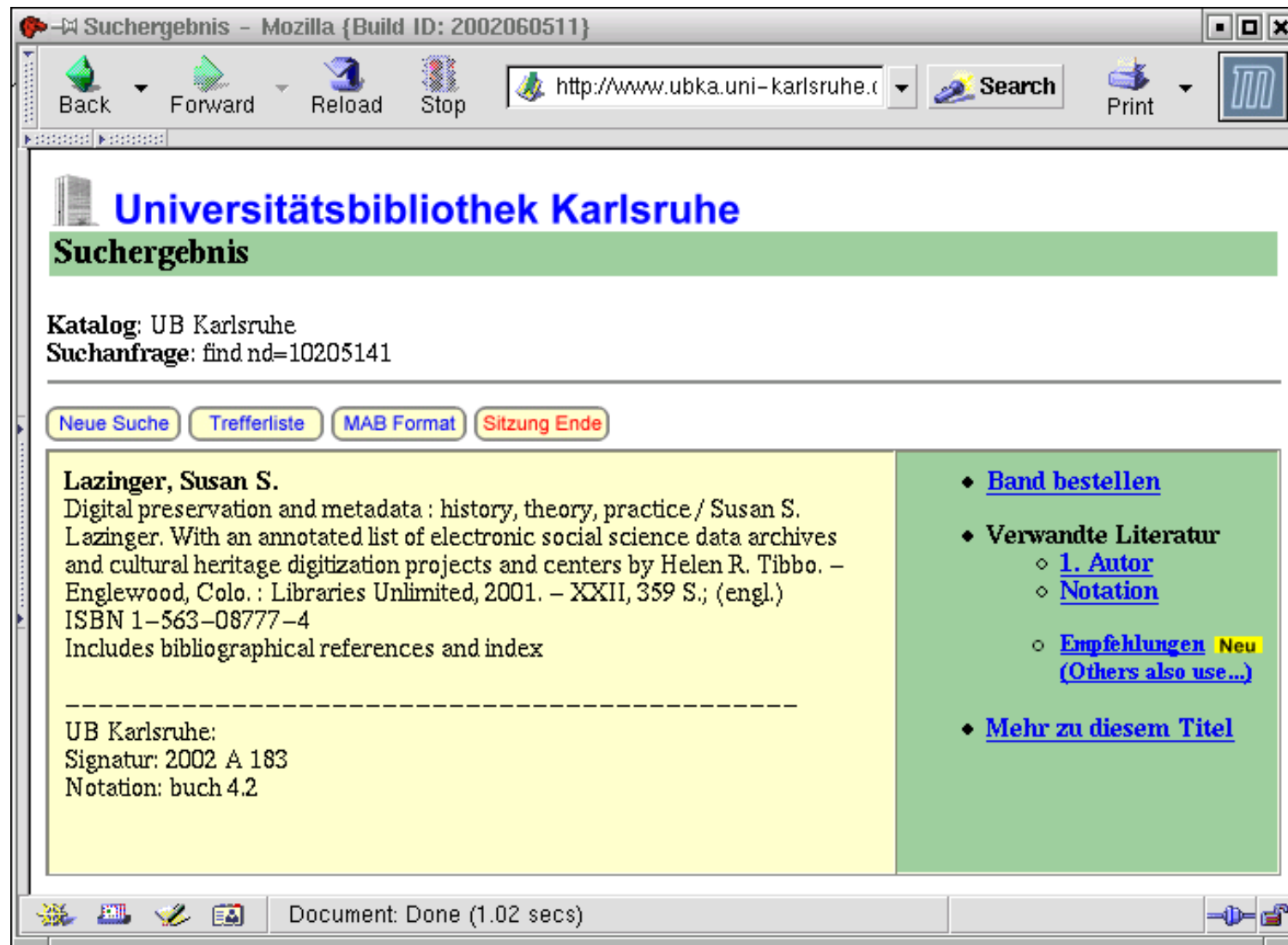
### 3 Project goals.

- Recommender systems aggregate knowledge from many peer groups to the level of expert advice services.
- Recommender systems bear the potential to significantly reduce transaction costs for literature searches by means of their aggregation capabilities.

**Goal:** To convert library OPACs into **customer oriented service portals supporting the interaction of the customers.**



## 3.1 Example: Detailed View of Documents (UBKA)



The screenshot shows a Mozilla browser window titled 'Suchergebnis - Mozilla {Build ID: 2002060511}'. The address bar contains 'http://www.ubka.uni-karlsruhe.c'. The browser interface includes navigation buttons (Back, Forward, Reload, Stop), a search button, and a print button. The main content area displays the 'Universitätsbibliothek Karlsruhe' logo and the heading 'Suchergebnis'. Below this, it shows 'Katalog: UB Karlsruhe' and 'Suchanfrage: find nd=10205141'. There are four buttons: 'Neue Suche', 'Trefferliste', 'MAB Format', and 'Sitzung Ende'. The search result for 'Lazinger, Susan S.' is displayed in a yellow box, including the title 'Digital preservation and metadata : history, theory, practice / Susan S. Lazinger', author information, publication details (Englewood, Colo. : Libraries Unlimited, 2001. - XXII, 359 S.; (engl.) ISBN 1-563-08777-4), and a note that it includes bibliographical references and index. To the right of the main text is a green box with a list of links: 'Band bestellen', 'Verwandte Literatur' (with sub-links for '1. Autor', 'Notation', and 'Empfehlungen Neu (Others also use...)'), and 'Mehr zu diesem Titel'. The status bar at the bottom indicates 'Document: Done (1.02 secs)'.

**Universitätsbibliothek Karlsruhe**  
**Suchergebnis**

Katalog: UB Karlsruhe  
Suchanfrage: find nd=10205141

[Neue Suche](#) [Trefferliste](#) [MAB Format](#) [Sitzung Ende](#)

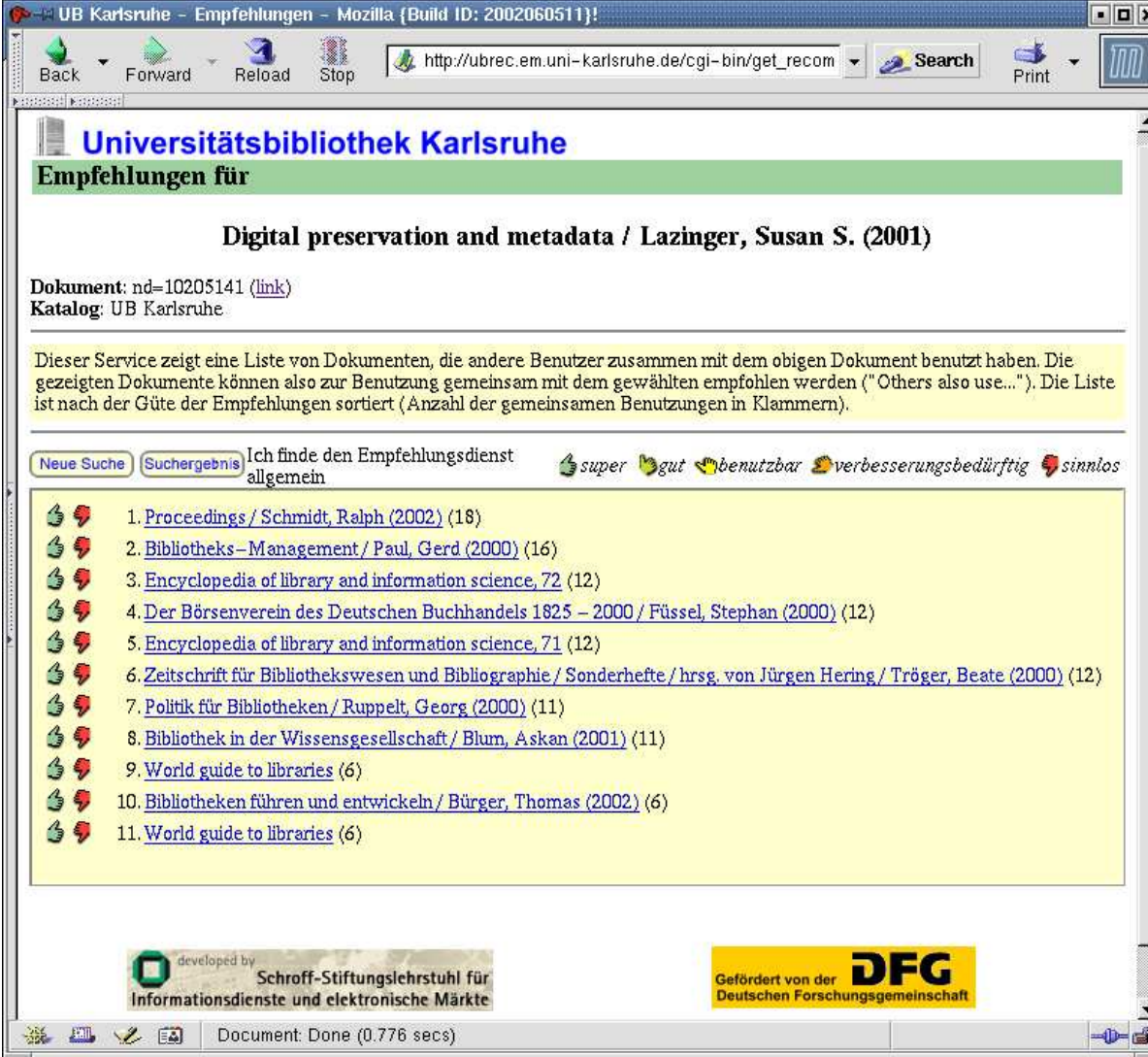
**Lazinger, Susan S.**  
Digital preservation and metadata : history, theory, practice / Susan S. Lazinger. With an annotated list of electronic social science data archives and cultural heritage digitization projects and centers by Helen R. Tibbo. - Englewood, Colo. : Libraries Unlimited, 2001. - XXII, 359 S.; (engl.) ISBN 1-563-08777-4  
Includes bibliographical references and index

UB Karlsruhe:  
Signatur: 2002 A 183  
Notation: buch 4.2

- ◆ [Band bestellen](#)
- ◆ [Verwandte Literatur](#)
  - [1. Autor](#)
  - [Notation](#)
  - [Empfehlungen Neu \(Others also use...\)](#)
- ◆ [Mehr zu diesem Titel](#)

Document: Done (1.02 secs)

### 3.1.1 Example: List of Recommendations (UBKA)



The screenshot shows a Mozilla browser window titled "UB Karlsruhe - Empfehlungen - Mozilla {Build ID: 2002060511}!". The address bar contains the URL "http://ubrec.em.uni-karlsruhe.de/cgi-bin/get\_recom". The page content is as follows:

**Universitätsbibliothek Karlsruhe**  
**Empfehlungen für**  
**Digital preservation and metadata / Lazinger, Susan S. (2001)**

Dokument: [nd=10205141](#) (link)  
Katalog: UB Karlsruhe

Dieser Service zeigt eine Liste von Dokumenten, die andere Benutzer zusammen mit dem obigen Dokument benutzt haben. Die gezeigten Dokumente können also zur Benutzung gemeinsam mit dem gewählten empfohlen werden ("Others also use..."). Die Liste ist nach der Güte der Empfehlungen sortiert (Anzahl der gemeinsamen Benutzungen in Klammern).

Neue Suche | Suchergebnis | Ich finde den Empfehlungsdienst allgemein | super | gut | benutzbar | verbesserungsbedürftig | sinnlos

1. [Proceedings / Schmidt, Ralph \(2002\)](#) (18)
2. [Bibliotheks-Management / Paul, Gerd \(2000\)](#) (16)
3. [Encyclopedia of library and information science, 72](#) (12)
4. [Der Börsenverein des Deutschen Buchhandels 1825 – 2000 / Füssel, Stephan \(2000\)](#) (12)
5. [Encyclopedia of library and information science, 71](#) (12)
6. [Zeitschrift für Bibliothekswesen und Bibliographie / Sonderhefte / hrsg. von Jürgen Hering / Tröger, Beate \(2000\)](#) (12)
7. [Politik für Bibliotheken / Ruppelt, Georg \(2000\)](#) (11)
8. [Bibliothek in der Wissensgesellschaft / Blum, Askan \(2001\)](#) (11)
9. [World guide to libraries](#) (6)
10. [Bibliotheken führen und entwickeln / Bürger, Thomas \(2002\)](#) (6)
11. [World guide to libraries](#) (6)

developed by Schroff-Stiftungslehrstuhl für Informationsdienste und elektronische Märkte

Gefördert von der DFG Deutschen Forschungsgemeinschaft

Document: Done (0.776 secs)

### 3.1.2 Example: List of Recommendations (UBKA)

UB Karlsruhe: Kataloge: Uni-Katalog : Empfehlungen

http://recommender.ubka.uni-karlsruhe.de/reckvk/recommendations/observeAndRecommend.php?searchsystem=UBKA

Universität Karlsruhe (TH) | Universitätsbibliothek  
Forschungsuniversität • gegründet 1825

Home | Mein Konto | Kataloge | Digitale Bibliothek | Lieferdienste | Fachgebiete | Informationen

**Uni-Katalog: Empfehlungen zu Ihrem Suchergebnis**

Dokument: nd=256011834  
Katalog: UB Karlsruhe

Neue Suche | Suchergebnis

**Ihr Suchergebnis**

Mac OS X internals / Singh, Amit , 2007



**Empfohlene Dokumente zu Ihrem Suchergebnis**

**Der links stehende Titel wurde auch zusammen mit folgenden Titeln aufgerufen:**  
(Anzahl der gemeinsamen Benutzungen in Klammern).

1. Einsteigen in Xcode / Anderson, Fritz , 2006, (15) + -
2. Apple Mac OS X 10.4 Tiger / Bechtel, Uthelm , 2006, (12) + -
3. AppleScript / Neuburg, Matt , 2006, (11) + -
4. Objective-C und Cocoa - Programmieren unter Mac OS X / Rodewig, Klaus M. , 2005, (10) + -
5. Advanced Mac OS X Programming / Dalrymple, Mark; Hillegass, Aaron , 2005, (9) + -
6. A practical guide to Unix for Mac OS X users / Sobell, Mark G.; Seebach, Peter , 2006, (6) + -
7. UNIX für Mac OS X-Anwender / Surendorf, Kai , 2006, (5) + -
8. Cocoa programming for Mac OS X / Hillegass, Aaron , 2004, (5) + -
9. AppleScript / Goldstein, Adam , 2005, (4) + -
10. Beginning Mac OS X programming / Trent, Michael; McCormack, Drew , 2005, (4) + -
11. UNIX für Mac OS X-Anwender / Surendorf, Kai , 2007, (4) + -
12. Learning Unix for Mac OS X Tiger / Taylor, Dave , 2005, (3) + -
13. AppleScript / Gumpinger, Marc , 1996, (3) + -
14. Java für Mac OS X / Much, Thomas , 2005, (3) + -

**BibTip**  
Ein Service der UB Karlsruhe

**Einschätzung der Empfehlungen:** Schätzen Sie die Empfehlungen mittels + und - ein und helfen Sie uns so, die Empfehlungsgüte weiter zu verbessern.

**Infos zum DFG-Projekt:**  
Zu den Projektseiten ...

**Umfrage:** Ich finde den Empfehlungsdienst allgemein:

- 👍 sehr gut
- 👎 gut
- 👎 mittel
- 👎 schlecht
- 👎 sehr schlecht

developed by

Schroff-Stiftungslehrstuhl für  
Informationsdienste und elektronische Märkte

Gefördert von der **DFG**  
Deutschen Forschungsgemeinschaft

– Seitenanfang –

## 3.2 User Interface: Library OPAC Search (KVK)

The screenshot shows the KVK Karlsruhe Virtueller Katalog search interface. The browser window title is "UB Karlsruhe: KVK Karlsruher Virtueller Katalog : Deutsch - Mozilla Firefox". The address bar shows the URL "http://www.ubka.uni-karlsruhe.de/kvk.html". The page header includes the University of Karlsruhe (TH) logo and the text "Universitätsbibliothek" and "Karlsruher Virtueller Katalog KVK". Below the header, there are navigation links for "KVK", "KVK Fernleihe", "KVK Kataloge", "Hilfe", and "Über den KVK".

The main search area is titled "KVK Deutsch" and contains a "Freitext" search box with a "NEU" label. Below this are several search criteria fields: "Titel" (with the value "Phönizier"), "Jahr", "Autor", "ISBN", "Körperschaft", "ISSN", "Schlagwort", and "Verlag". There are "Suchen" and "Löschen" buttons at the bottom of the search area, and a "Katalogauswahl löschen" button on the right.

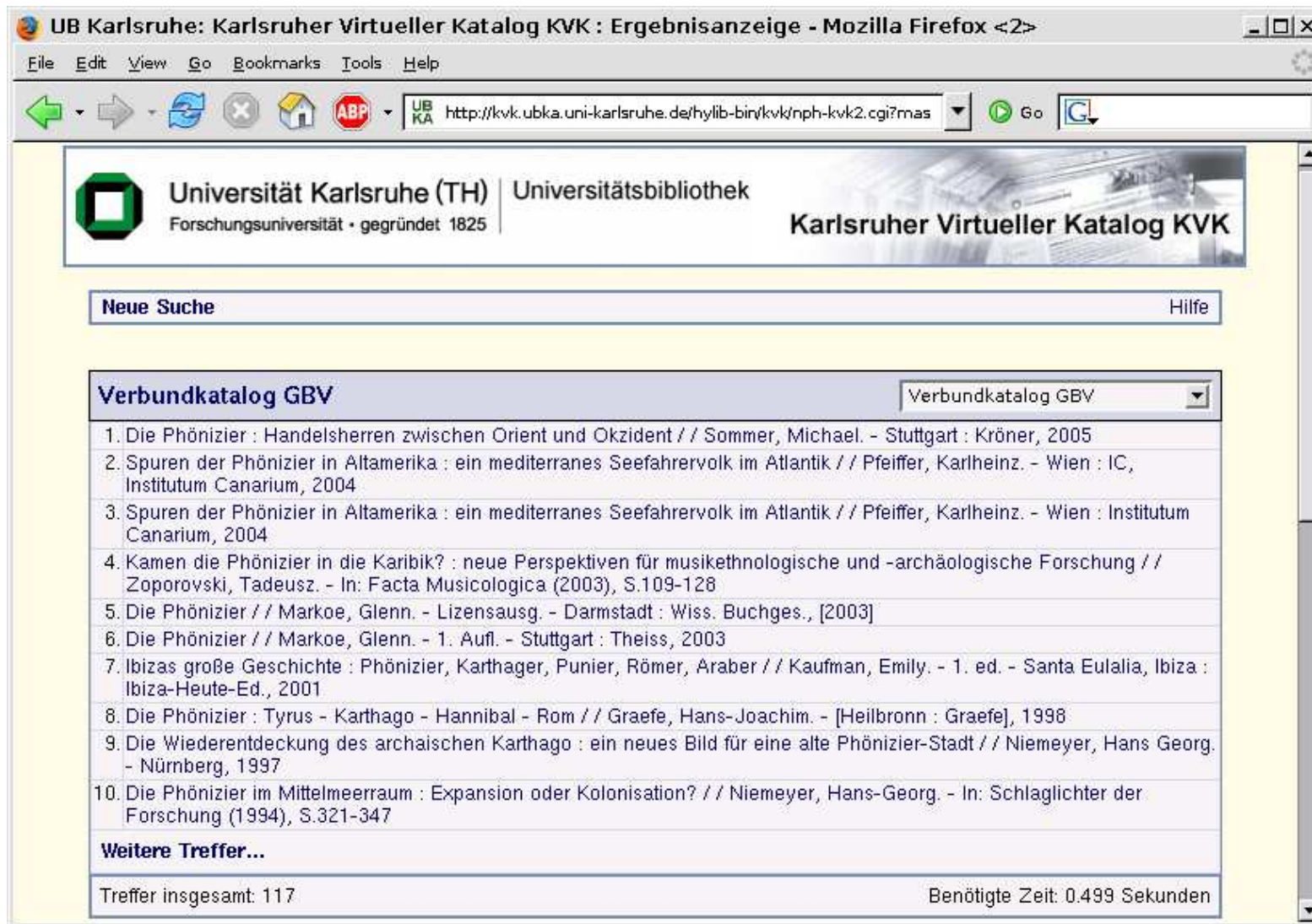
On the right side, there are settings for "Optionen bei Suche und Ergebnisanzeige", including "Einstellungen" (with "Sichern" and "Laden" buttons), "Volltitelanzeige" (with a checked "Neues Fenster" option), and a "Timeout" set to "120 Sekunden".

At the bottom, there are four columns of checkboxes for selecting libraries and bookstores:

- Deutschland:** SWB, BVB, HBZ, HEBIS, HEBIS-Retro (NEU), KOBV, GBV (checked), DNB, StaBi Berlin, TIB Hannover, ÖVK.
- Österreich:** Österr. BV, Österr. Landesbibl., Österr. NB 1501 - 1929, Österr. NB 1930 - 1991, Österr. NB 1992 -.
- Schweiz:** Helveticat NB Bern, IDS Basel/Bern, IDS Zürich Universität, NEBIS / ZB Zürich, Westschweizer BV RERO.
- Weltweit:** Australische NB, Dänische NB, Finnische NB, Französische NB, Französischer VK, Britischer VK, British Library, Italienischer VK, Italienischer ZS-VK, Kanada CISTI Kat., Kanadischer VK.
- Buchhandel:** abebooks.de, Amazon.de, Booklooker.de, KNV, Libri.de, ZVAB.



## 3.2.1 User Interface: Hit List (KVK)



The screenshot shows a Mozilla Firefox browser window displaying the search results for the Karlsruhe Virtual Catalog (KVK). The browser's address bar shows the URL: `http://kvk.ubka.uni-karlsruhe.de/hylib-bin/kvk/nph-kvk2.cgi?mas`. The page header includes the logo of the University of Karlsruhe (TH) and the text "Universitätsbibliothek" and "Karlsruher Virtueller Katalog KVK". Below the header, there is a search bar with the text "Neue Suche" and a "Hilfe" link. The main content area is titled "Verbundkatalog GBV" and displays a list of 10 search results. At the bottom of the results, it indicates "Treffer insgesamt: 117" and "Benötigte Zeit: 0.499 Sekunden".

UB Karlsruhe: Karlsruher Virtueller Katalog KVK : Ergebnisanzeige - Mozilla Firefox <2>

File Edit View Go Bookmarks Tools Help

http://kvk.ubka.uni-karlsruhe.de/hylib-bin/kvk/nph-kvk2.cgi?mas

Universität Karlsruhe (TH) | Universitätsbibliothek  
Forschungsuniversität • gegründet 1825

Karlsruher Virtueller Katalog KVK

Neue Suche Hilfe

Verbundkatalog GBV Verbundkatalog GBV

1. Die Phönizier : Handelsherren zwischen Orient und Okzident // Sommer, Michael. - Stuttgart : Kröner, 2005
2. Spuren der Phönizier in Altamerika : ein mediterranes Seefahrervolk im Atlantik // Pfeiffer, Karlheinz. - Wien : IC, Institutum Canarium, 2004
3. Spuren der Phönizier in Altamerika : ein mediterranes Seefahrervolk im Atlantik // Pfeiffer, Karlheinz. - Wien : Institutum Canarium, 2004
4. Kamen die Phönizier in die Karibik? : neue Perspektiven für musikethnologische und -archäologische Forschung // Zoporovski, Tadeusz. - In: Facta Musicologica (2003), S.109-128
5. Die Phönizier // Markoe, Glenn. - Lizenzausg. - Darmstadt : Wiss. Buchges., [2003]
6. Die Phönizier // Markoe, Glenn. - 1. Aufl. - Stuttgart : Theiss, 2003
7. Ibizas große Geschichte : Phönizier, Karthager, Punier, Römer, Araber // Kaufman, Emily. - 1. ed. - Santa Eulalia, Ibiza : Ibiza-Heute-Ed., 2001
8. Die Phönizier : Tyrus - Karthago - Hannibal - Rom // Graefe, Hans-Joachim. - [Heilbronn : Graefe], 1998
9. Die Wiederentdeckung des archaischen Karthago : ein neues Bild für eine alte Phönizier-Stadt // Niemeyer, Hans Georg. - Nürnberg, 1997
10. Die Phönizier im Mittelmeerraum : Expansion oder Kolonisation? // Niemeyer, Hans-Georg. - In: Schlaglichter der Forschung (1994), S.321-347

Weitere Treffer...

Treffer insgesamt: 117 Benötigte Zeit: 0.499 Sekunden

## 3.2.2 User Interface: Document Inspection Page (KVK)

GBV Search and Order - results/titledata - Mozilla Firefox

File Edit View Go Bookmarks Tools Help

http://gso.gbv.de/DB=2.1/CHARSET=UTF-8/IMPLAND=Y/LNG Go

Suchen Suchergebnis Erweiterte Suche Zwischenablage Benutzer-Info Hilfe © 1998-2007 OCLC PICA

GBV Gesamtkatalog

suchen [und] Alle Wörter [ALL] sortiert nach Erscheinungsjahr

Benutzergruppe: 8164 | IP: 129.13.72.153

Suchgeschichte Kurzliste **Titeldaten** Nachweisinformationen

1 von 1

■ Ihre Aktion suchen [und] ISB 3520454017

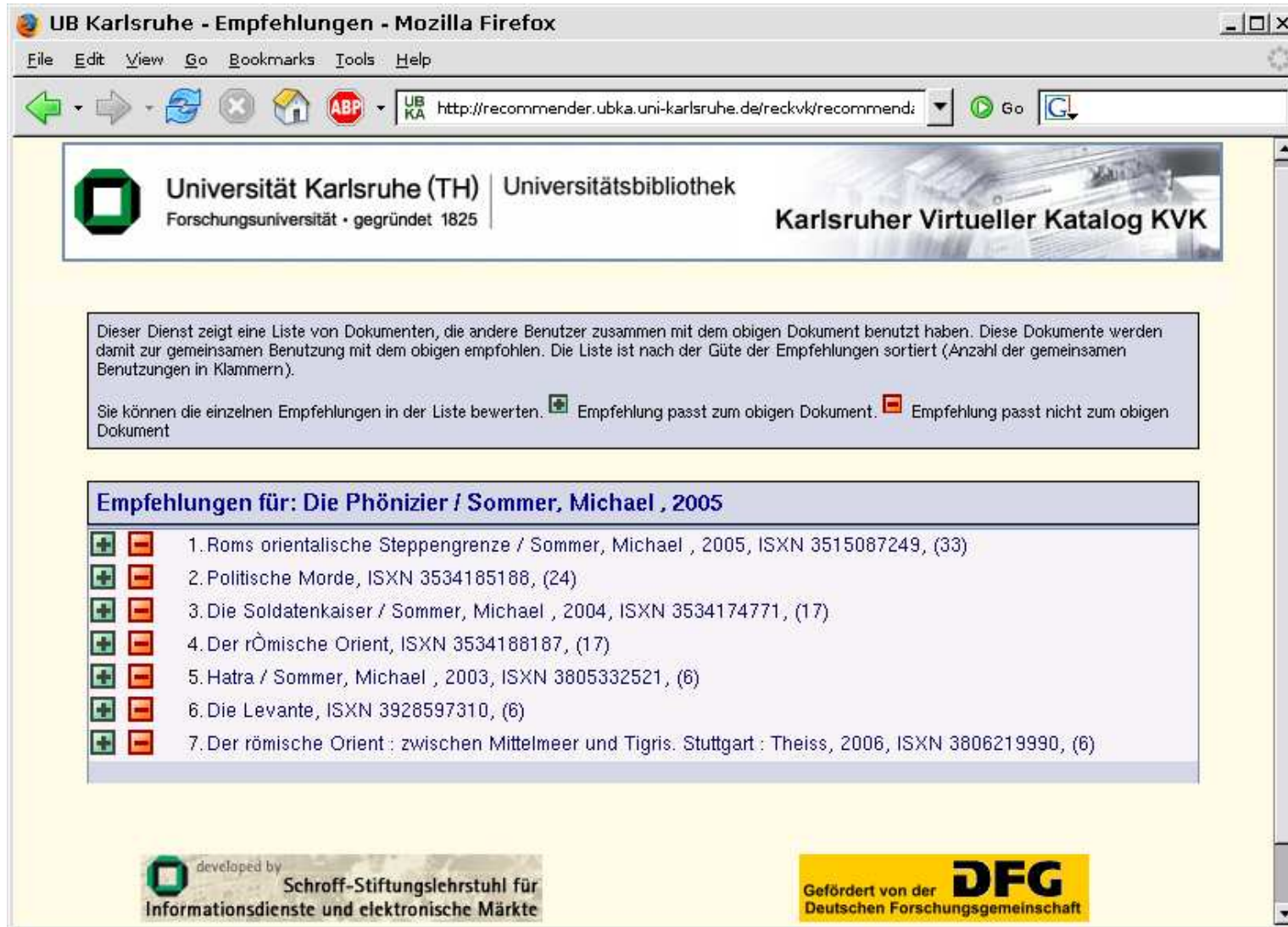
**Titel:** Die [Phönizier](#) : [Handelsherren zwischen Orient und Okzident](#) / [Michael Sommer](#)  
**Verfasser:** [Sommer, Michael](#) \*1970-  
**Erschienen:** [Stuttgart](#) : [Kröner](#), 2005  
**Umfang:** VI, 277 S. : Ill., Kt. ; 18 cm  
**Schriftenreihe:** Kröners Taschenausgabe ; 454  
**Anmerkung:** Literaturverz. S. [249] - 264  
**ISBN:** **3-520-45401-7**

**Schlagwörter:** \*[Phöniker](#) / [Geschichte](#)  
 \*[Phoenicians](#)

**Sachgebiete:** [15.26](#) ; [Alter Orient, Nordafrika](#) <[Alte Geschichte](#)>  
**Mehr zum Thema:** [DS81](#) Klassifikation der Library of Congress  
**Elektron. Referenz:** [http://deposit.ddb.de/cgi-bin/dokserv?id=2633441&prov=M&dok\\_var=1&dok\\_ext=htm](http://deposit.ddb.de/cgi-bin/dokserv?id=2633441&prov=M&dok_var=1&dok_ext=htm)

1 von 1 [zugehörige Publikationen](#) **KVK** [Empfehlungen](#) gehe zu 1 von 1

### 3.2.3 User Interface: List of Recommendations (KVK)



The screenshot shows a Mozilla Firefox browser window titled "UB Karlsruhe - Empfehlungen - Mozilla Firefox". The address bar contains the URL "http://recommender.ubka.uni-karlsruhe.de/reckvk/recommend:". The page header includes the logo of the University of Karlsruhe (TH) and the text "Universitätsbibliothek" and "Karlsruher Virtueller Katalog KVK".

The main content area contains a text box explaining the recommendation service: "Dieser Dienst zeigt eine Liste von Dokumenten, die andere Benutzer zusammen mit dem obigen Dokument benutzt haben. Diese Dokumente werden damit zur gemeinsamen Benutzung mit dem obigen empfohlen. Die Liste ist nach der Güte der Empfehlungen sortiert (Anzahl der gemeinsamen Benutzungen in Klammern). Sie können die einzelnen Empfehlungen in der Liste bewerten.  Empfehlung passt zum obigen Dokument.  Empfehlung passt nicht zum obigen Dokument".

Below this is a section titled "Empfehlungen für: Die Phönizier / Sommer, Michael , 2005". It lists seven recommended documents, each with a green plus icon and a red minus icon for evaluation:

1. Roms orientalische Steppengrenze / Sommer, Michael , 2005, ISXN 3515087249, (33)
2. Politische Morde, ISXN 3534185188, (24)
3. Die Soldatenkaiser / Sommer, Michael , 2004, ISXN 3534174771, (17)
4. Der römische Orient, ISXN 3534188187, (17)
5. Hatra / Sommer, Michael , 2003, ISXN 3805332521, (6)
6. Die Levante, ISXN 3928597310, (6)
7. Der römische Orient : zwischen Mittelmeer und Tigris. Stuttgart : Theiss, 2006, ISXN 3806219990, (6)

At the bottom, there are logos for "developed by Schroff-Stiftungslehrstuhl für Informationsdienste und elektronische Märkte" and "Gefördert von der Deutschen Forschungsgemeinschaft DFG".

## **4 Scientific roots and theories combined.**





- Paul A. Samuelson: Revealed Preference Theory (Nobel Prize in Economics 1970)
- A. Michael Spence and Joseph Stiglitz: Market Signalling and Screening by Self-Selection (Nobel Prize in Economics 2001)
- Modelling as a “Gestaltproblem”. Causal models versus noise models.
- Ideas from ensemble methods:  
e.g. Albert Einstein’s work about the molecular Brownian motion. (Nobel prize in physics 1921).
- Ehrenbergs repeat-buying theory.

# 4.1 Paul A. Samuelson: Revealed Preference Theory (Nobel Prize in Economics 1970)









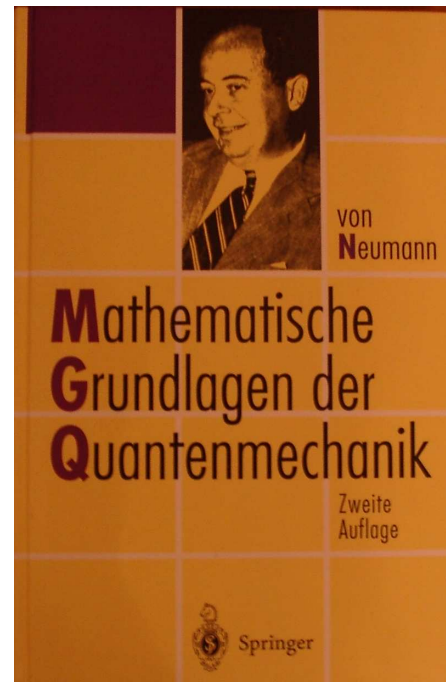


**Purchases reveal preferences.**

## 4.2 A. Michael Spence and Joseph Stiglitz: Market Signalling and Screening by Self-Selection (Nobel Prize in Economics 2001)

### Signalling (M. Spence):

- The choice of John von Neumann's *Quantenmechanik* (quantum mechanics) signals experience and expert knowledge.
- Incentive compatibility: voluntary and truthful.



### Screening (J. Stiglitz):

- A professor offers a menu of course books each of which requires a certain level of experience and expert knowledge.
- Selection of a book by a student reveals the student's level.
- Incentive compatibility: voluntary and truthful.

**Self-selection processes build homogenous clusters of readers.**

## 4.3 Modelling as a “Gestaltproblem”: Causal models versus noise models



Economics and management science versus natural sciences.

## 4.4 Ideas from ensemble methods.

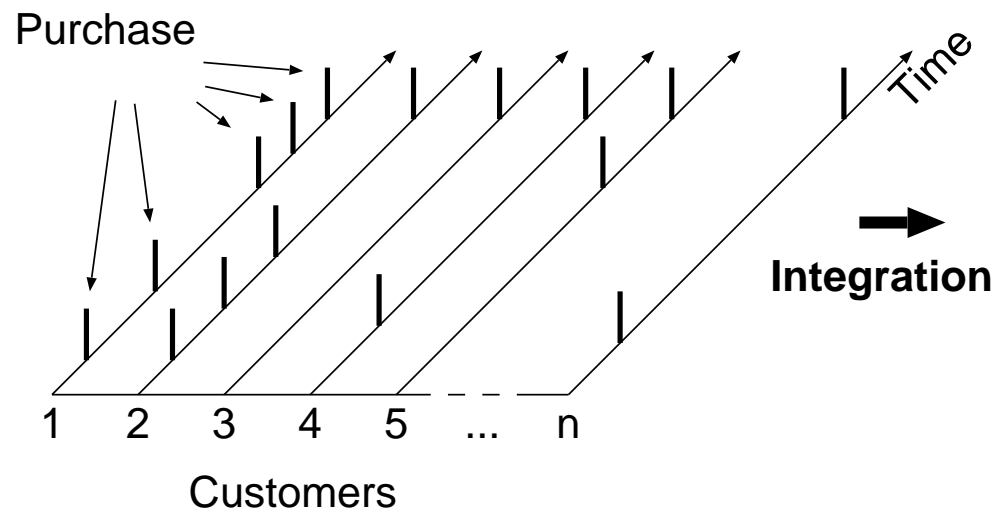
Ensemble methods are methods which measure the behavior and the interactions of a large set of particles only by observing of the states of the aggregate (e.g. pressure, temperature and volume of a gas).

### **Examples:**

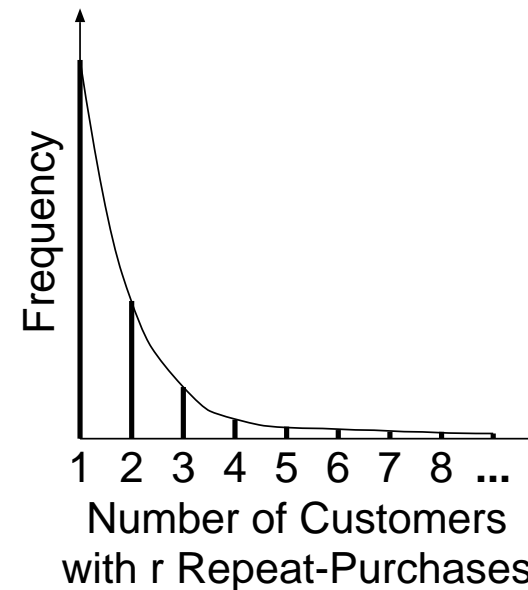
- Ludwig Boltzmann's Thermodynamics.
- Albert Einstein's work about the molecular Brownian motion.  
(Nobel prize in physics 1921).
- Ilya Prigogine: Nonequilibrium Statistical Mechanics.  
(Nobel prize in chemistry 1977).

## 4.5 Ehrenbergs Repeat-Buying Theory.

**Independent Stochastic Purchase Processes  
for Product  $x$  or Productclass  $x$**



**Frequency Distribution of  
Repeat Purchases  
of Period  $t$**



**Independent Poisson processes act as models of noise and filters:  
Outliers are recommendations.**



## 5 Statistical models.

## 5.1 The LSD Model: A Mixture of Poisson Processes

The logarithmic series distribution describes, how many consumers have bought a specific product 1, 2, 3, ... times in the observation period (without taking non-buyers into account):

$$P(r \text{ purchases}) = \frac{-q^r}{r \ln(1 - q)}, \quad r \geq 1 \quad (1)$$

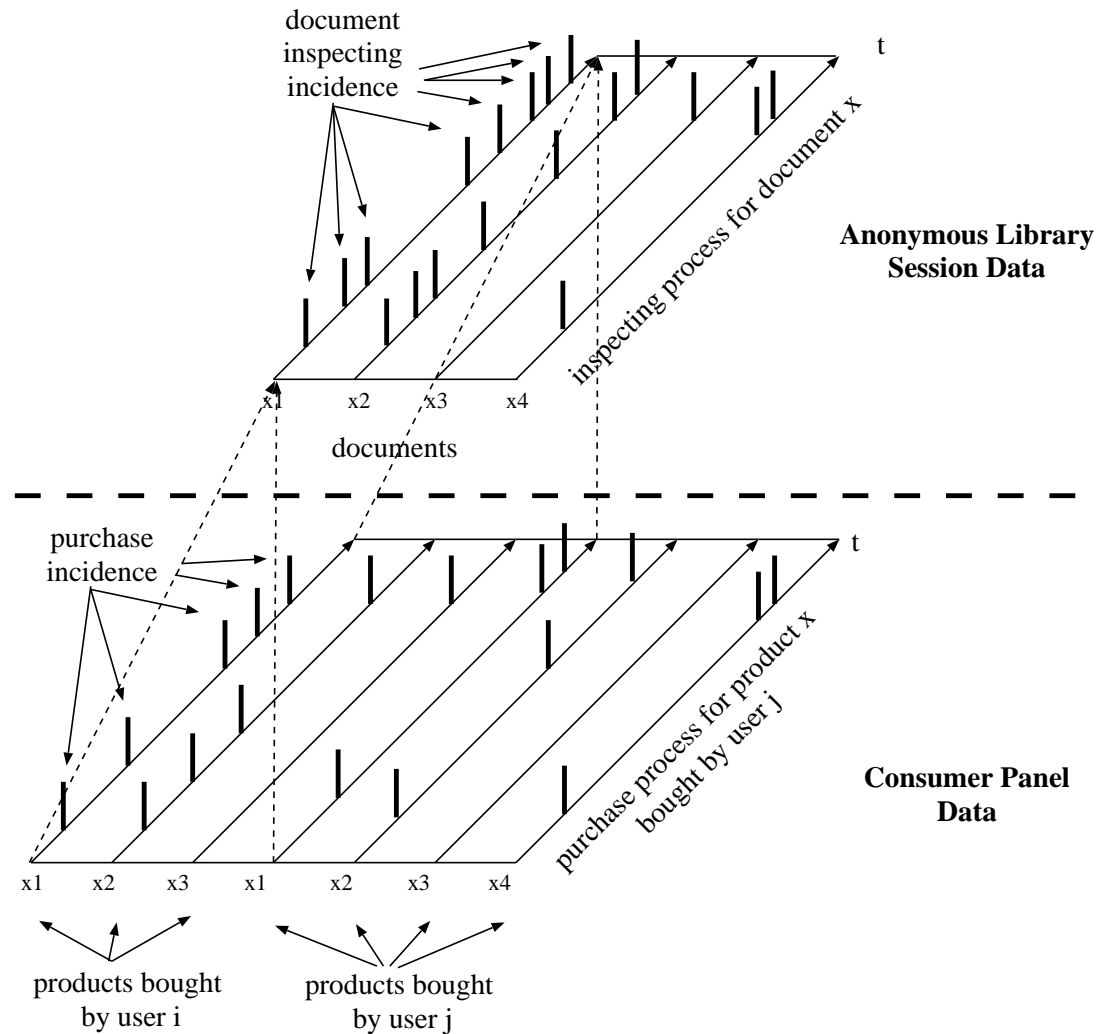
Mean Purchase Frequency:

$$w = \frac{-q}{(1 - q) \ln(1 - q)} \quad (2)$$

Variance:

$$\sigma^2 = \frac{-q \frac{1+q}{\ln(1-q)}}{(1 - q)^2 \ln(1 - q)} \quad (3)$$

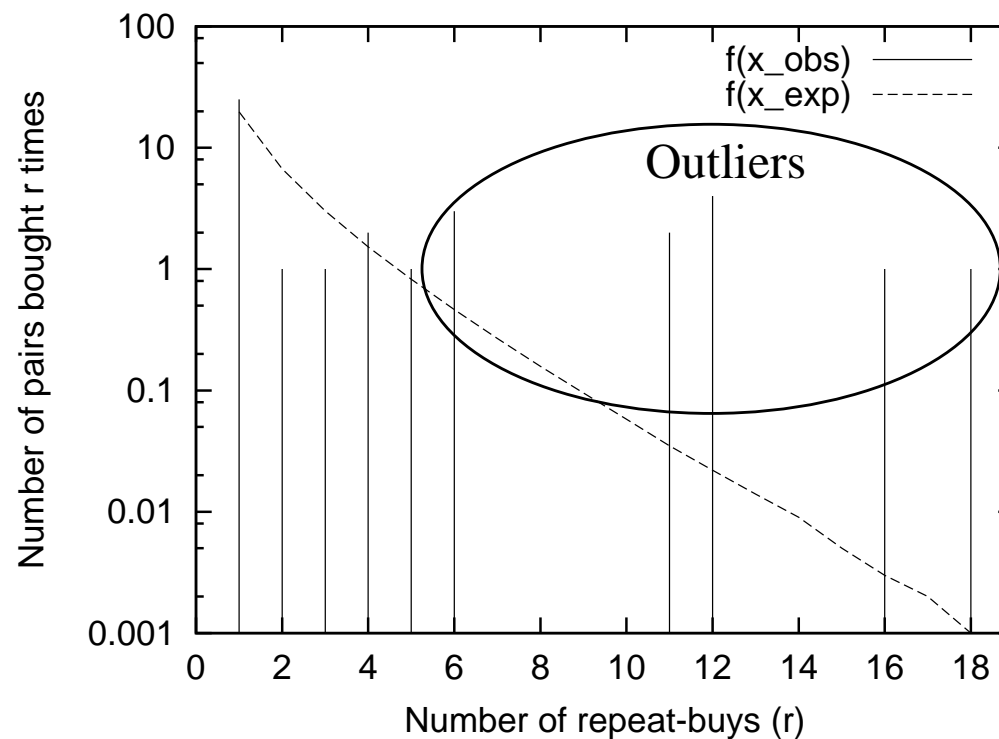
# 5.1.1 Stochastic Processes



From Consumer Panels to Anonymous Library Users

## 5.1.2 Generating Recommendations

- Products which are bought together more frequently than expected by the stochastic model
- violate independence assumptions of the model



Observed Distribution Function

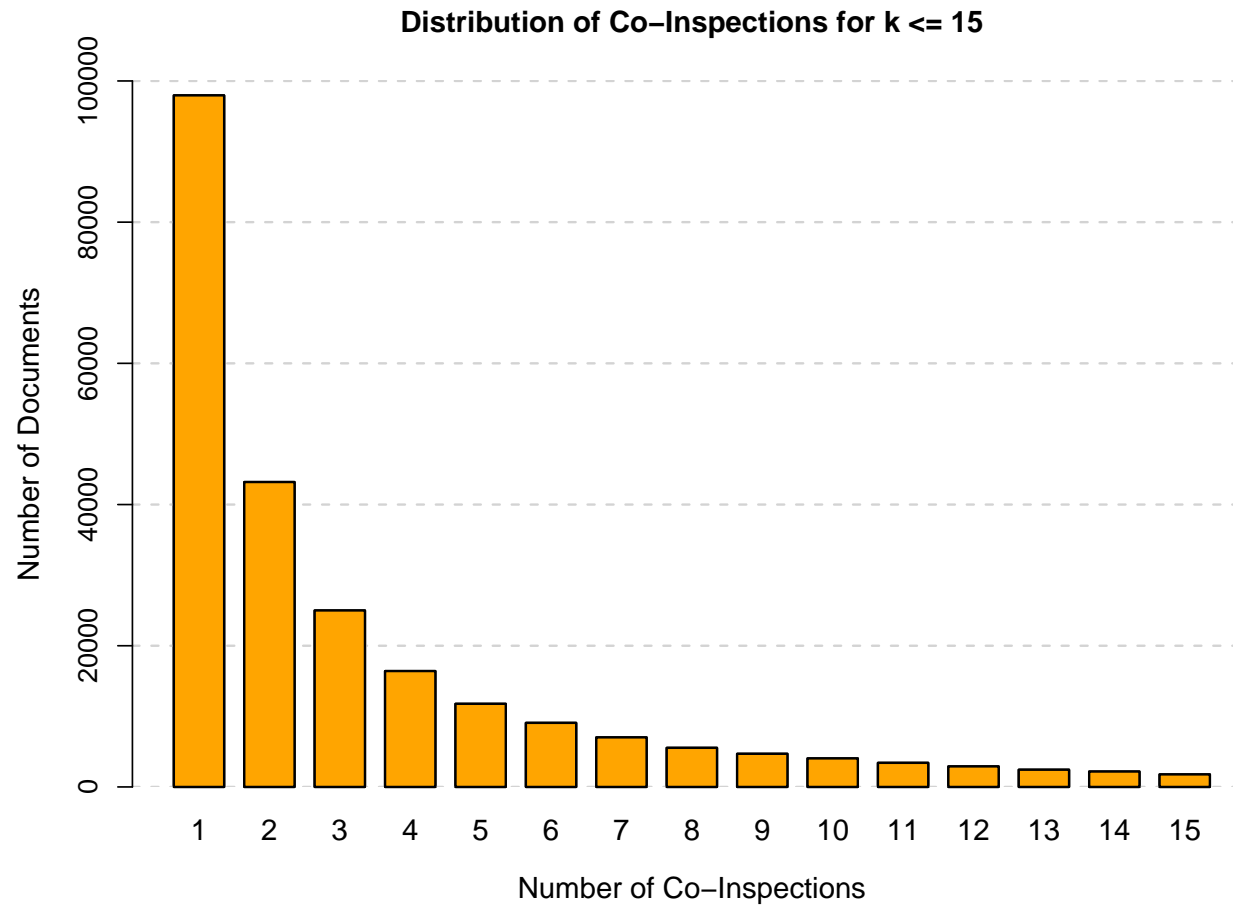
## 5.2 Multinomial models: Small samples

Status of 2007-02-19

	UBKA	KVK
Number of total documents in catalog	1,000,000	> 10,000,000
Number of total co-inspected documents	527,363	255,248
Average market baskets size	4.9	2.9
Av. aggregated co-inspections per document	117.4	5.4

- When starting to monitor new catalogs (stores) no information about the us age of the documents (products) in the catalog is available.
- With smaller market baskets and a greater number of total documents it takes much more time for many co-inspections per document to occur.
- Due to sample size constraints methods using statistical tests on distributions (like LSD) are only reliably applicable with many co-inspections.
- Special small sample statistics are needed to compute recommendations out of samples of few co-inspections.

## 5.2.1 Distributions of Co-Inspections (KVK Data)



## 5.2.2 Data Setup

- The number of total documents  $n + 1$  in the catalog is finite but unknown.
- Recommendations are computed separately for each document  $D$ .
- Each user session (market basket) contains all documents that the user inspected within that session, multiple inspections of the same document are counted as one.
- All user sessions are aggregated.
- The set  $C(D)$  contains all documents, that at least one user has inspected together with  $D$ .
- The number of co-inspections with  $D$  of all elements of  $C(D)$  is known, this histogram is called  $H(D)$ .
- $H(D)$  can be interpreted as an integer partition with the number of co-inspections of each co-inspected document as the summands. (For partition theory see Euler, MacMahon, Hardy, Ramanujan, Rademacher.)

## 5.2.3 General Concept

**Assumption:** All documents in the catalog have the same probability of being co-inspected.

- In real systems generally this assumption does not hold.
- Especially when starting to observe new catalogs no information about the underlying distribution of the inspection processes of documents is known.
- We do not develop a causal model of a decision maker, we construct **an ideal decision maker without preferences.**



## 5.2.4 An Ideal Decision Maker

- Is a prototype for a class of homogenous decision-makers
- Makes choices randomly. But from what choice set?
- The conceptual model of the choice set:  
Total set  $\supseteq$  awareness set  $\supseteq$  consideration set  $\supseteq$  choice set (Kotler 1980)
- Evidence for the model:
  - Awareness set from 3 – 11 products (Narayana and Markin (1975))
  - Interaction of emotional and rational brain activity (Bechara et al. 1997): Emotional activity reduces size of choice set of normal persons, whereas for patients with prefrontal damage this could not be observed.
  - In a brand choice experiment successful branding reduces the choice set to 1 (the winning brand) and the regions of the brain responsible for emotions show high activity when measured by fMRI (Deppe et al. 2005).

Consequence: We consider a sequence of increasing event spaces.

## 5.2.5 POSICI: Probability of Single Item Co-Inspections

**Question:** What is the probability  $p_j$  that at least one other document has been co-inspected exactly  $j$ -times with document  $D$ ?

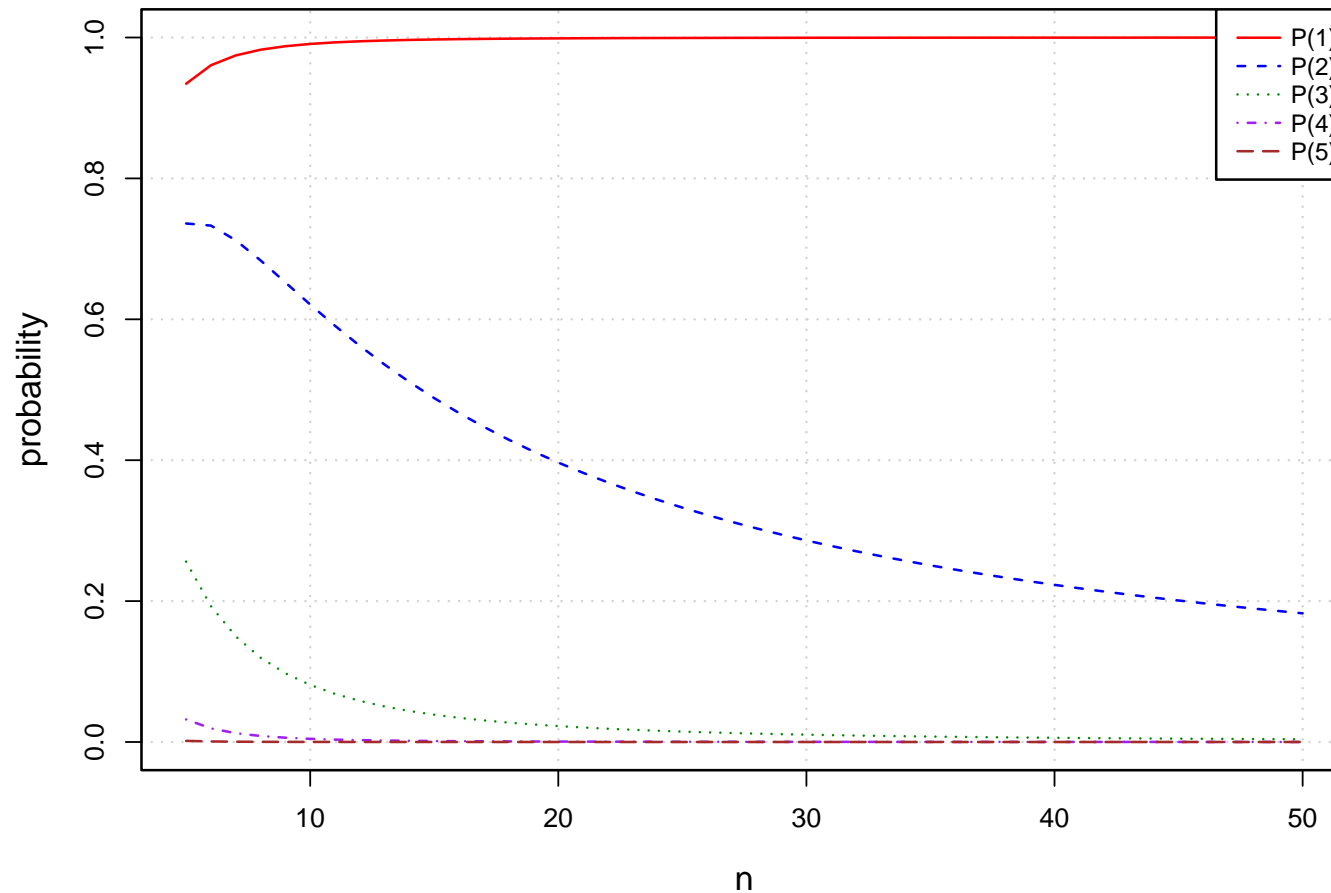
- Let  $n+1$  be the number of total documents,  $k$  the number of non-aggregated co-inspections (multiple inspections in different sessions are counted separately)
- Let  $(N_1, \dots, N_n)$  be the vector of the number of times document  $i$  ( $1 \leq i \leq n$ ) was co-inspected with  $D$
- Then  $(N_1, \dots, N_n) \sim \mathcal{M}(k; p_1, \dots, p_n)$ ,  $p_i = \frac{1}{n}$ ,  $1 \leq i \leq n$
- Let  $A_i = \{N_i = j\}$

**Answer:** (inclusion-exclusion principle)

$$p_j = P\left(\bigcup_{i=1}^n A_i\right) = \sum_{\nu=1}^n (-1)^{\nu-1} \sum_{1 \leq i_1 < \dots < i_\nu \leq n} P(A_{i_1} \cap \dots \cap A_{i_\nu})$$

## 5.2.6 Inspection Probabilities for Growing $k$ and $n$ in POSICI

Co-inspection probabilities for  $k = 5$  ( $n = 5$  to 50)



## 5.2.7 Interpretation of POSICI

Example for  $k = 9$ :

- Observed partitions:
  - (1)  $5+2+2$
  - (2)  $5+2+1+1$
- The 5-times co-inspected document has the same probability to be a recommendation in both cases.

POSICI is built on the theory, that co-inspections other than  $j$ -times add more noise than information about the incentive to co-inspect the current document  $j$ -times. (The form of the histogram does not reveal additional information)

## 5.2.8 POMICI: Probability of multiple items co-inspections

**Question:** What is the probability  $p_{part}$  that the partition corresponding to the complete histogram  $H(D)$  of all co-inspections with  $D$  occurs?

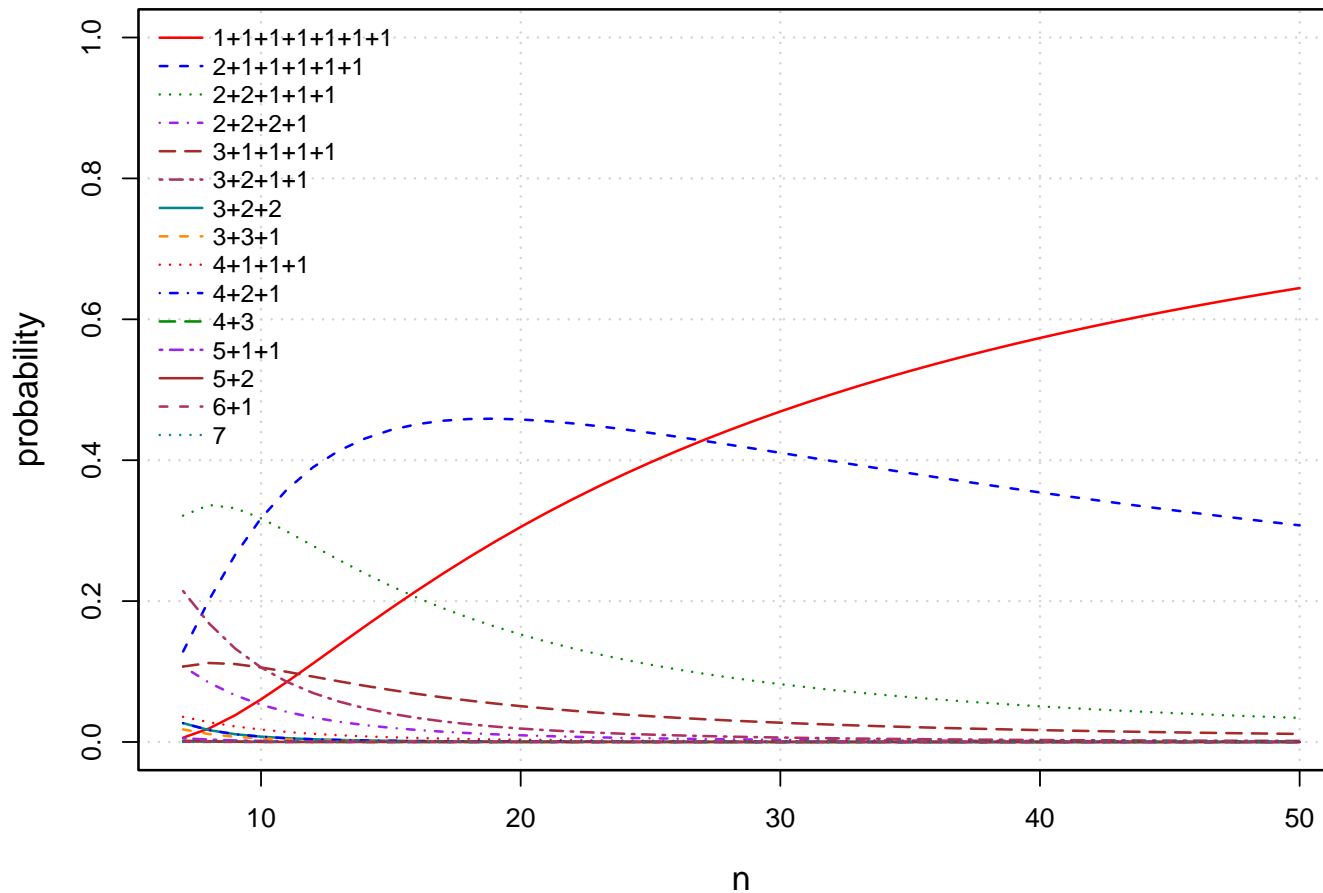
- Let  $n+1$  be the number of total documents,  $k$  the number of non-aggregated co-inspections (multiple inspections in different sessions are counted separately)
- Let  $X$  be the set of words of length  $k$  from an alphabet of  $n$  letters, and  $l_i$  the number of letters (i. e. documents), that occur exactly  $i$ -times in  $x \in X$  (i. e. in  $H(D)$ )
- First examine the actions of the group  $G = S_n \times S_k$  on the set  $X$ , and then the actions of the stabilizer subgroup  $G_x$  on the set  $S_n$  for the identity  $id \in S_n$

**Answer:** (applying two times the orbit-stabilizer theorem together with Lagrange's theorem from group theory and some counting arguments)

$$p_{part} = \frac{|G_x|}{|X|} = \frac{|G|}{|G_x|} = \frac{|G|}{|G_{xid}| |G_{xid}|} = \frac{n! k!}{(n - \sum_{i=1}^k l_i)! \prod_{j=1}^k l_j! (j!)^{l_j}}$$

## 5.2.9 Partition Probabilities for Growing $k$ and $n$ in POMICI

Partition probabilities for  $k = 7$  ( $n = 7$  to  $50$ )



## 5.2.10 Interpretation of POMICI

Example for  $k = 9$ :

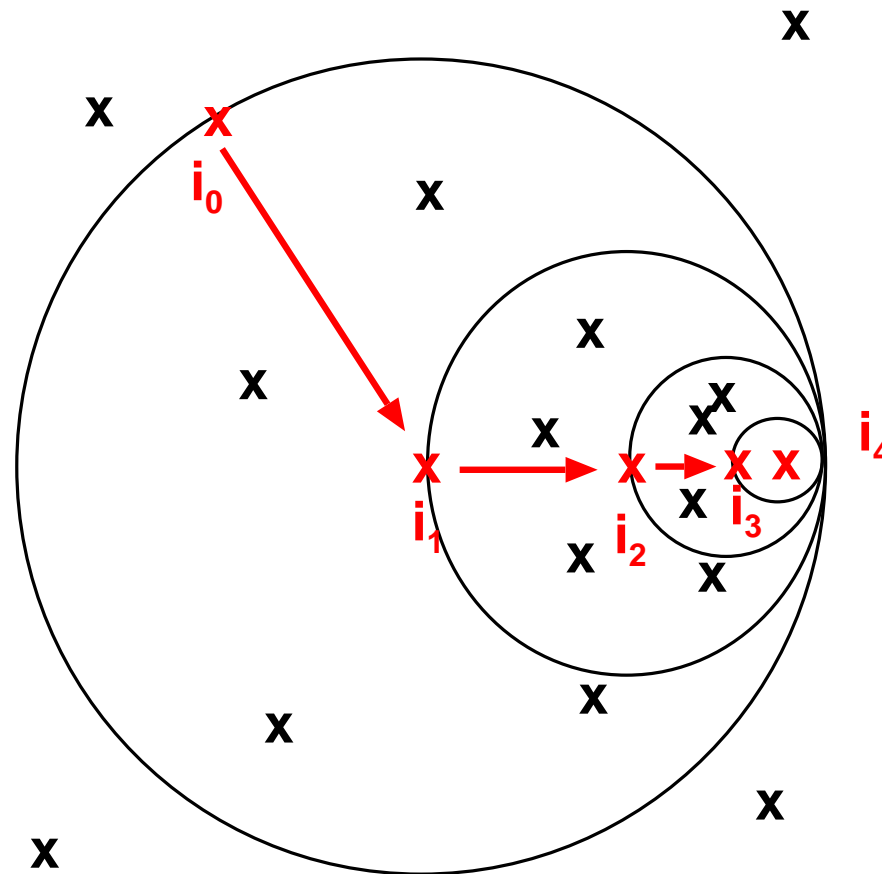
- Observed partitions:
  - (1) 5+2+2
  - (2) 5+2+1+1
- The 5-times co-inspected document has **not** the same probability to be a recommendation in both cases.
- Partition (2) is more likely, so here the 5-times co-inspected document has a very slightly higher chance of being recommended.

POMICI is built on the theory, that the distribution of co-inspections other than  $j$ -times reveals more information than noise about the incentive to co-inspect the current document  $j$ -times.



## **6 Other applications (prototypes).**

## 6.1 Restricted random walk clustering for indexing



## 6.2 Adaptive recommender systems

Recommendations for

**Finding groups in data - an introduction to cluster analysis / Leonard Kaufman ; Peter J. Rousseeuw (1990)**

*Few precise hits*



*Many, but less precise hits*

- (75) Robust regression and outlier detection / Peter J. Rousseeuw ; Annick M. Leroy (1987)
- (66) Cluster analysis and data analysis / M. Jambu and M.-O. Lebeaux (1983)



UB Karlsruhe - Empfehlungen - Mozilla Firefox

Datei Bearbeiten Ansicht Gehe Lesezeichen Extras Hilfe

Universität Karlsruhe (TH)  
Universitätsbibliothek

Home | Mein Konto | Kataloge | Digitale Bibliothek | Liefersdienste | Fachgebiete | Infos

Recommendations for

**Automatische Klassifikation - theoretische und praktische Methoden zur Gruppierung und Strukturierung von Daten / von Hans Hermann Bock (1974)**

*Few precise hits*  *Many, but less precise hits*

- (83) Clustering algorithms / John A. Hartigan (1975)
- (83) Clusteranalyse - anwendungsorientierte Einführung / Johann Bacher (1996)
- (83) Mathematical classification and clustering / Boris G. Mirkin (1996)
- (80) Forecasting economic time series / Michael P. Clements and David F. Hendry (1998)
- (77) Data analysis - scientific modeling and practical application ; with 45 tables / Wolfgang Gaul ... (eds.) (2000)
- (70) Cluster analysis for applications / Michael R. Anderberg (1973)

developed by  
Schroff-Stiftungslehrstuhl für  
Informationsdienste und elektronische Märkte

Gefördert von der  
Deutschen Forschungsgemeinschaft

## Recommendations for

**Finding groups in data - an introduction to cluster analysis / Leonard Kaufman ; Peter J. Rousseeuw (1990)***Few precise hits**Many, but less precise hits*

- (75) Robust regression and outlier detection / Peter J. Rousseeuw ; Annick M. Leroy (1987)
- (66) Cluster analysis and data analysis / M. Jambu and M.-O. Lebeaux (1983)
- (50) Clusteranalyse - anwendungsorientierte Einführung / Johann Bacher (1996)
- (50) Empirical methods for artificial intelligence / Paul R. Cohen (1995)
- (50) Clustern mit Hintergrundwissen / Andreas Hotho (2004)
- (50) Bayeslösungen des Ausreißerproblems / Friedrich Gebhardt (1961)
- (50) Ausreisser bei ein- und mehrdimensionalen Wahrscheinlichkeitsverteilungen / Rudolf Mathar (1981)
- (33) Clustering algorithms / John A. Hartigan (1975)
- (33) Mathematical classification and clustering / Boris G. Mirkin (1996)
- (33) Data analysis - scientific modeling and practical application ; with 45 tables / Wolfgang Gaul ... (eds.) (2000)
- (33) Modern regression methods / Thomas P. Ryan (1997)
- (33) Untersuchung zur zeitlich-räumlichen Ähnlichkeit von phänologischen und klimatologischen Parametern in Westdeutschland u / von Xiaoqiu Chen (1994)
- (33) Cluster analysis / Brian Everitt (1974)
- (33) Social Science Research Council / Social Science Research Council ()
- (25) New approaches in classification and data analysis / E. Diday ... (eds.) (1994)
- (25) Clusteranalyse - Einführung in Methoden und Verfahren der automatischen Klassifikation ; mit zahlreichen Algorithmen, FO / Detlef Steinhausen ; Klaus Langer (1977)
- (25) Tests und Schätzungen in Ausreißermodellen / Ursula Gather (1984)
- (25) Concurrence probabilities for a locally slotted packet radio network by combinatorial methods / Rudolf Mathar (1990)
- (25) Classification and dissimilarity analysis / Bernard Van Cutsem (ed.) (1994)
- (25) Fallstudien Cluster-Analyse / Helmuth Späth (1977)
- (11) Mastering data mining - the art and science of customer relationship management / Michael J. A. Berry ; Gordon Linoff (2000)
- (11) Entwicklung von Kundenbeziehungen - theoretische und empirische Analysen unter dynamischen Aspekten / Dominik Georgi (2000)
- (11) Kundenwert - Grundlagen - innovative Konzepte - praktische Umsetzungen / Bernd Günter ... (Hrsg.) (2003)
- (11) Relationship Marketing - das Management von Kundenbeziehungen / von Manfred Bruhn (2001)
- (11) Customer-Lifetime-Value-Management - Kundenwert schaffen und erhöhen: Konzepte, Strategien, Praxisbeispiele / Markus Hofmann ... (Hrsg.) (2000)
- (11) Kundenwertmanagement - Konzept zur wertorientierten Analyse und Gestaltung von Kundenbeziehungen / Gunter Eberling. Mit einem Geleitw. von Günter Specht (2002)
- (11) Den Kundennutzen managen - so beschreiten sie den Weg zur Wertschöpfungskette / Harald Münzberg (1995)
- (8) Scheduling theory / Tanaev, Vjaceslav S. (1994)

## 6.3 Exploration of the information space

The screenshot shows a web browser window titled "Unikatalog Karlsruhe" with the URL <http://www.ism.uni-karlsruhe.de/RecoDiver.html>. The page header includes the logos for "Universität Karlsruhe (TH) Universitätsbibliothek" and "RecoDiver IISM". A navigation menu contains links for "Home", "Mein Konto", "Kataloge", "Digitale Bibliothek", "Liefersdienste", "Fachgebiete", and "Infos".

The main content area is titled "RecoDiver Applet" and features a network diagram of book covers. The diagram is a radial network with a central node and many peripheral nodes, all connected by lines. The nodes are represented by book covers, some of which are highlighted with green borders. A search bar at the top right of the applet is labeled "Search in titles:". On the left side, there is a vertical list of years from 1991 to 2007, with horizontal bars indicating the number of books for each year. A "clear selection" button is located at the top left of the applet.

Two book covers are highlighted with green borders and have their titles and authors displayed in a tooltip:

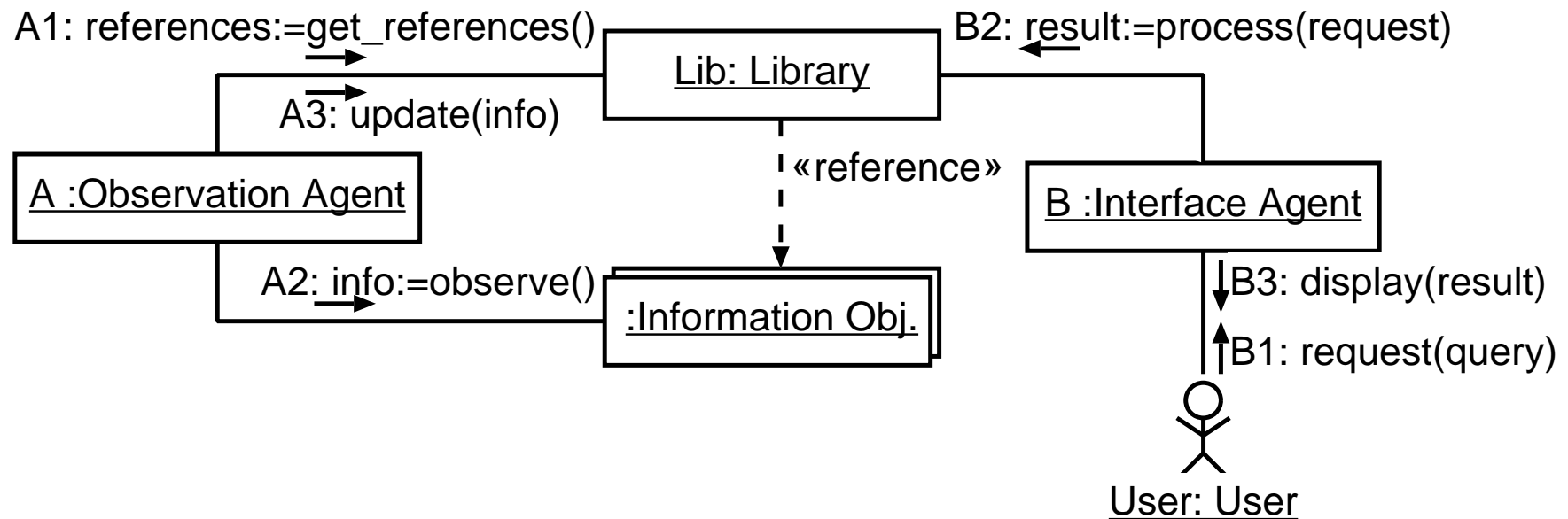
- AppleScript  
Neuburg, Matt  
2006, O'Reilly
- Mac OS X internals - Singh, Amit (Addison-Wesley)

On the right side of the applet, there is a detailed view of the "AppleScript" book cover, showing the title "AppleScript The Definitive Guide", the author "Matt Neuburg", the year "2006", and the publisher "O'Reilly". Below the cover, the text "AppleScript Neuburg, Matt 2006, O'Reilly" is displayed, followed by a green square icon and the word "available".



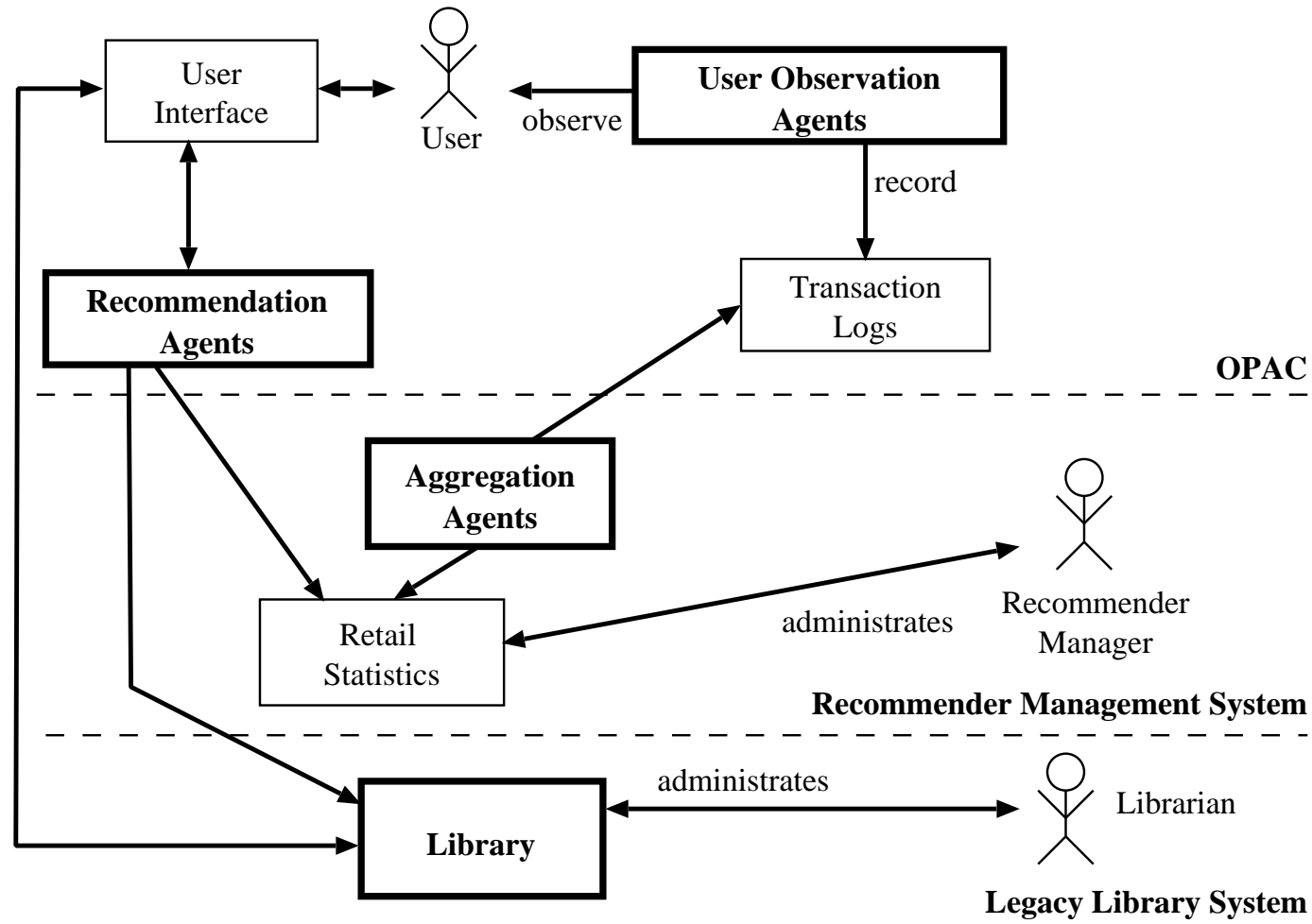
## **7 System architecture and scalability.**

## 7.1 An Agency of Software Agents for Distributed Recommender Services



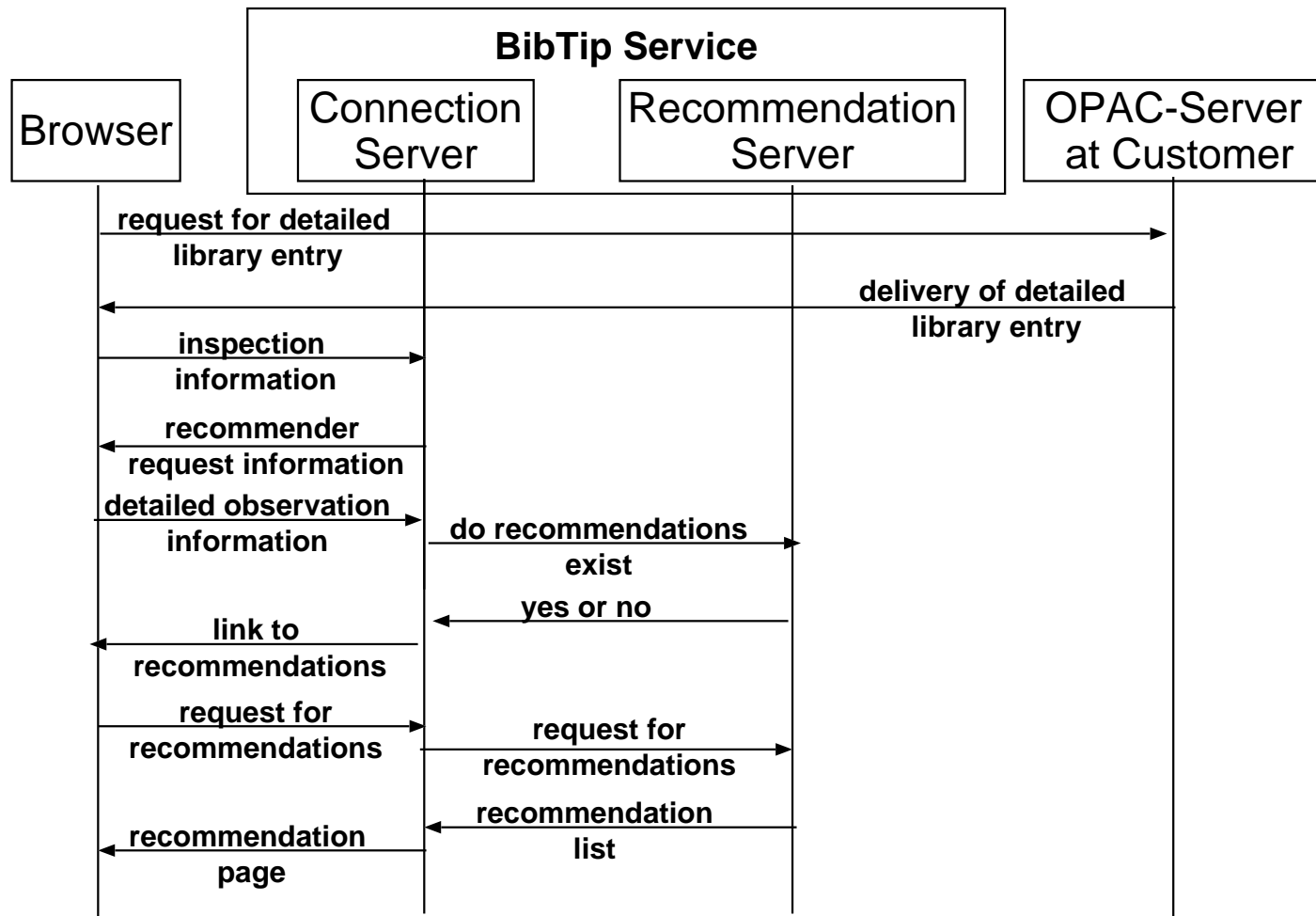
An Agency of Software Agents as Analysis Pattern

## 7.2 Recommender Services for Legacy Library Systems I



The Architecture of a Library System with Recommender Services

## 7.3 Recommender Services for Legacy Library Systems II



Sequence Diagram and Message Trace for a Detailed Document View

## **8 Service integration into web OPACS of scientific libraries.**

## 8.1 Integration Variant I: apache-logs, cgi-scripts

- User Observation Agent:
  - http-logs with link embedded session IDs
  - preprocessing:
    - \* extraction of GET requests and session splitting to take care of public access terminals
- Aggregation Agent:
  - information market: selecting information = purchase
  - computes market-baskets, estimates a logarithmic series distribution (LSD), and identifies and extracts outliers as recommendations
  - performs incremental updates periodically
- Recommendation Agent:
  - CGI-script on the recommendation server generates recommendation pages and is accessed via embedded links in the OPAC

Problems with maintenance and operations.



## 8.2 Integration Variants II and III: Web 2.0 Technology

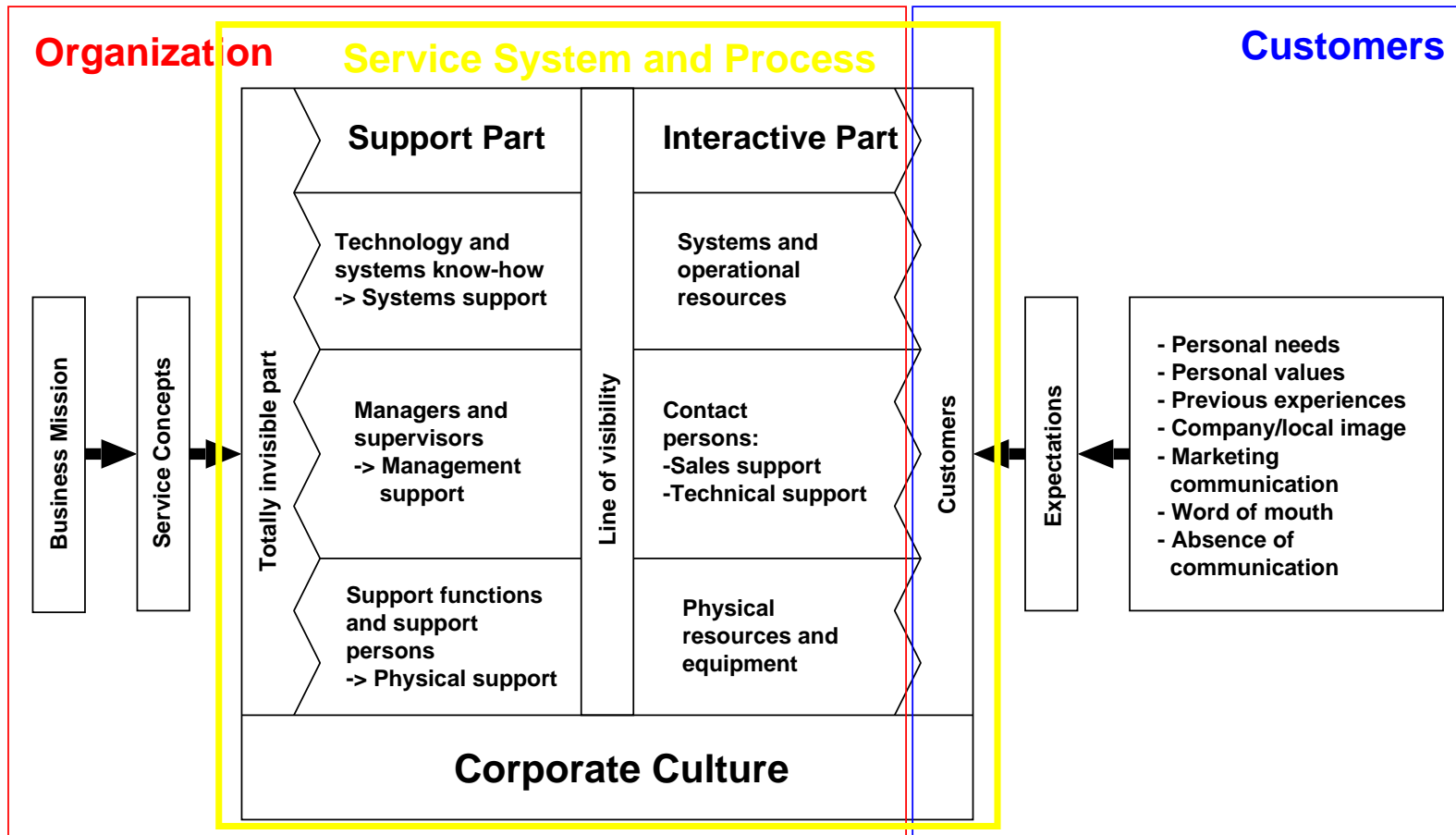
- As a web-service:
  - The recommender server is an (internal) web-service.  
(Interface: SOAP, XML, WSDL)
  - Connection server may provide a web-service interface.

**Advantage:** Customer has a full control on layout and layout change processes.

- With JavaScript:
  - A few lines of JavaScript are integrated at the customer's OPAC.
  - The script is dynamically rewritten and controls the process shown in the sequence diagram.

**Advantage:** Minimal integration overhead.

# 9 Development of a service concept and a service organization.



## **10 Development of a global service infrastructure for (scientific) libraries.**

- Partners/alliances.
- Development of infrastructure.
- Migration to a distributed system.

## 10.1 Reference Customers

- Universität Karlsruhe (TH) (OPAC und KVK), operational since 2001, 2005.
- University of Freiburg, observed.
- Badische Landesbibliothek (a State library of Baden-Württemberg), observed.
- Technische Universität Berlin, observed.
- HSB Esslingen, Reutlingen, Konstanz, in preparation.
- Universität Stuttgart, in preparation.
- Universität Hohenheim, in preparation.
- Deutsche National Bibliothek, in preparation.

# Thanks for your attention!

## Project Links:

<http://reckvk.em.uni-karlsruhe.de/>

<http://www.ubka.uni-karlsruhe.de/catalog/>

<http://www.bibtip.org>

## Contact:

Andreas.Geyer-Schulz@kit.edu

Andreas.Neumann@iism.uni-karlsruhe.de

## Acknowledgement.

We thank the DFG for the continued support and funding of the project Recommender Systems for Meta Library Catalogues.

## References

- Geyer-Schulz, A., Hahlser, M., Neumann, A., and Thede, A. (2003a). Recommenderdienste für wissenschaftliche Bibliotheken und Bibliotheksverbände. In Geyer-Schulz, A. and Taudes, A., editors, *Informationswirtschaft – ein Sektor mit Zukunft*, volume P-33 of *GI-Edition – Lecture Notes in Informatics (LNI)*, Bonn. Köllen.
- Geyer-Schulz, A., Hahsler, M., Neumann, A., and Thede, A. (2003b). Behavior-based recommender systems as value-added services for scientific libraries. In Bozdogan, H., editor, *Statistical Data Mining & Knowledge Discovery*. Chapman & Hall / CRC.
- Geyer-Schulz, A., Hahsler, M., Neumann, A., and Thede, A. (2003c). An integration strategy for distributed recommender services in legacy library systems. In Schader, M., Gaul, W., and Vichi, M., editors, *Between Data Science And Applied Data Analysis*, Studies in Classification, Data Analysis, and Knowledge Organization, Heidelberg-Berlin. Springer.
- Geyer-Schulz, A., Neumann, A., Heitmann, A., and Stroborn, K. (2003d).



Strategic positioning options for scientific libraries in markets of scientific and technical information – the economic impact of digitization. *Journal of Digital Information*, 4(2).

Geyer-Schulz, A., Neumann, A., and Thede, A. (2003e). An architecture for behavior-based library recommender systems. *Information Technology and Libraries*, 22(4):165–174.

Geyer-Schulz, A., Neumann, A., and Thede, A. (2003f). Others also use: A robust recommender system for scientific libraries. In Koch, T. and Sølvsberg, I. T., editors, *Research and Advanced Technology for Digital Libraries*, volume 2769 of *Lecture Notes in Computer Science*, pages 113 – 125, Berlin Heidelberg New York. Springer.

Neumann, A. W. (2007). Motivating and supporting user interaction with recommender systems. In Kovács, L., Fuhr, N., and Meghini, C., editors, *Research and Advanced Technology for Digital Libraries*, LNCS 4675, pages 428–439, Berlin Heidelberg New York. Springer.

Neumann, A. W. and Geyer-Schulz, A. (2007). Applying small sample test statistics for behavior-based recommendations. In *Proceedings of the 31st*

*Annual Conference of the German Classification Society.* Springer. To appear.