

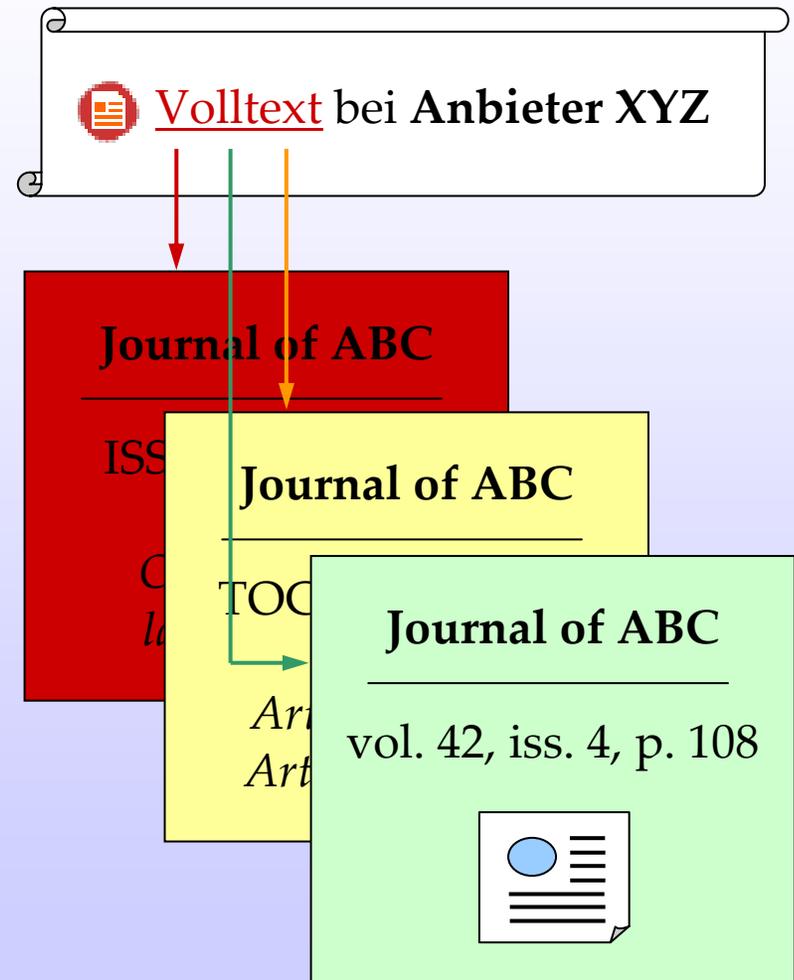
DFG-Projekt Generischer Link-Resolver

Ziele, Methoden und Ergebnisse

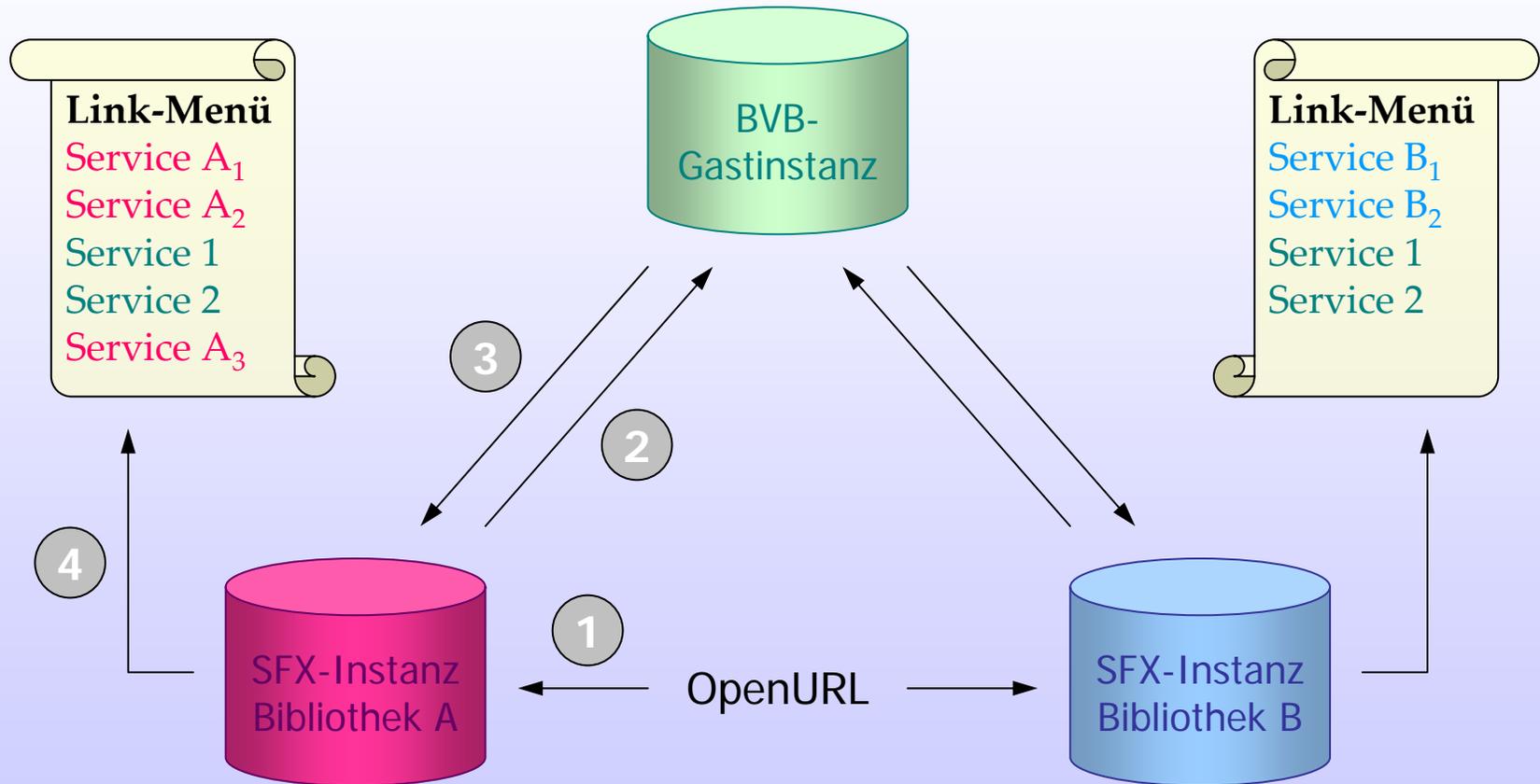
*Matthias Groß · Dr. Mathias Kratzer · Dr. Rolf Siepmann
Bibliotheksverbund Bayern (BVB)*

- 
1. Wozu soll das gut sein?
 2. Wie soll das gehen?
 3. Warum das alles andere als trivial ist ...
 4. Erste Schritte auf vielen neuen Wegen

- Der **Linking-Level** ist die Dokumentenebene, auf die ein Volltextlink den Benutzer führt.
- **Ideal:** so tief wie die vorhandenen Metadaten es erlauben
- bei **freien** Angeboten häufig nur bis auf die Journal-Homepage ☹



- Erschließung einer tief reichenden **inbound linking syntax** (ILS) und Implementierung eines entsprechenden **Link-Konstruktors** ist mit einigem Aufwand verbunden.
- **Problem 1:** geringe Effizienz bei vielen freien Zeitschriften (~1 Titel pro Server!) motiviert nicht zur Tiefenerschließung
- **Ziel 1:** Verbesserung der Integration freier Volltextangebote ins Open Linking



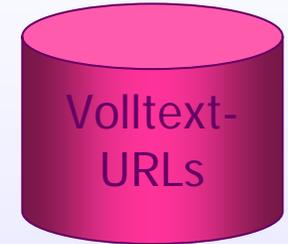
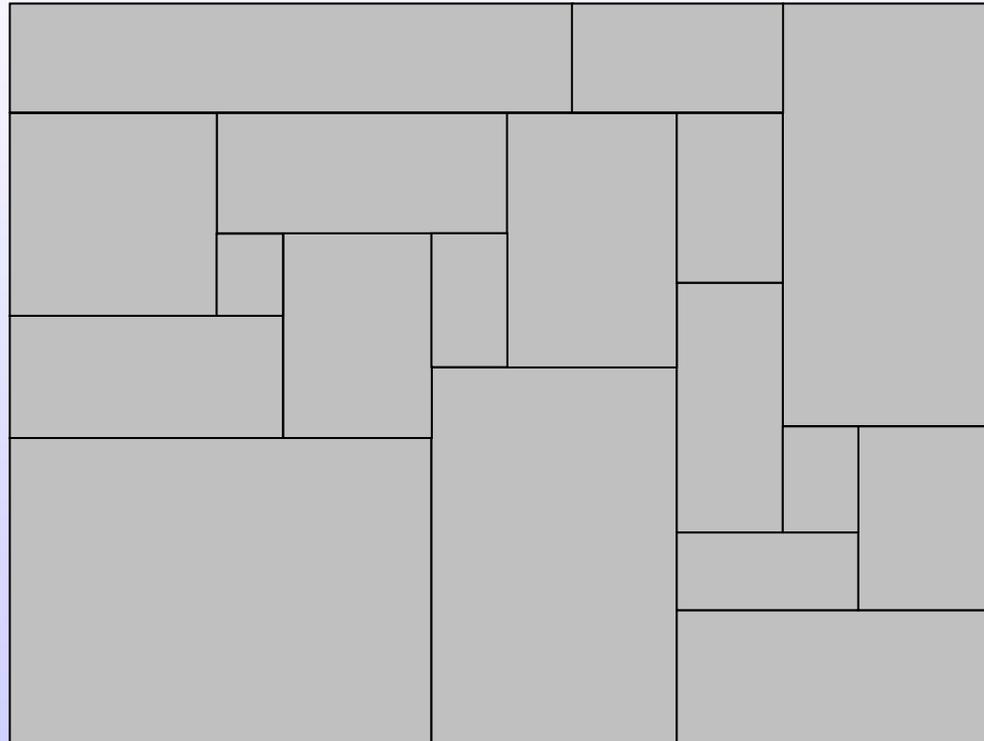
- Warum operiert eine solche „Gastinstanz“ eigentlich nur auf regionaler Ebene?
- **Ziel 2:** generischer Link-Resolver in dem Sinne, dass **jeder institutionelle** Link-Resolver freie Volltext-Services von dort beziehen kann
- **Problem 2:** Ohne produktübergreifende Schnittstellen ist nur ein geringes Maß an Akzeptanz zu erwarten.

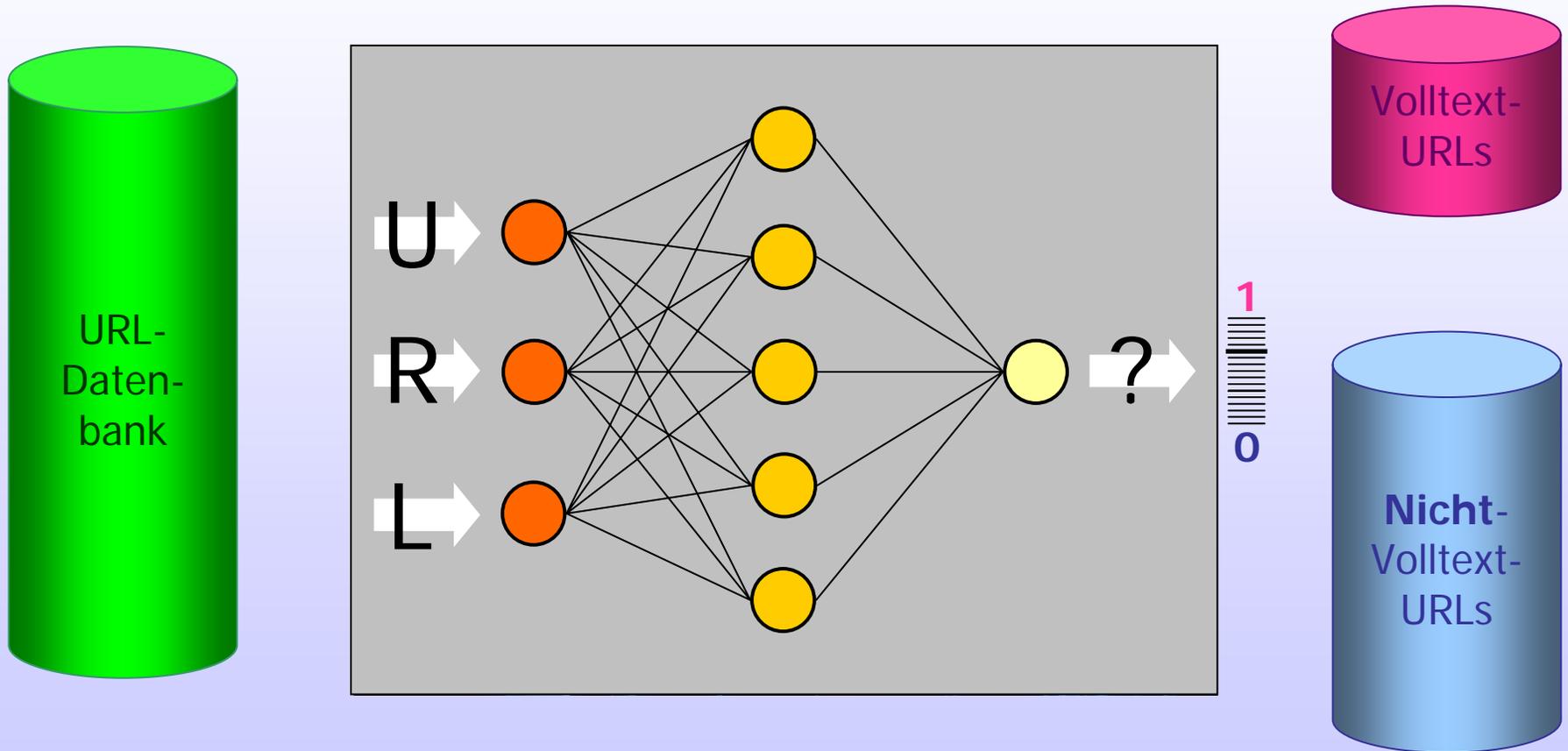
- 
1. Wozu soll das gut sein?
 2. Wie soll das gehen?
 3. Warum das alles andere als trivial ist ...
 4. Erste Schritte auf vielen neuen Wegen

```
● http://eku.comu.edu.tr/...  
● http://www.iam.fmph.uniba.sk/...  
● http://www.shd.org.yu/...  
● http://www.revistaorbis.org.ve/...  
● http://epaa.asu.edu/...  
● http://antalya.uab.es/...  
● http://www.denbridgepress.com/...  
● http://www.egms.de/...  
● http://profetic.org/...  
● http://www.jsnm.org/...  
● http://jrp.icaap.org/...  
● http://jcom.sissa.it/...  
● http://edoc.hu-berlin.de/...  
:  
:
```



pro Journal nach Editierdistanz:





Base-URL

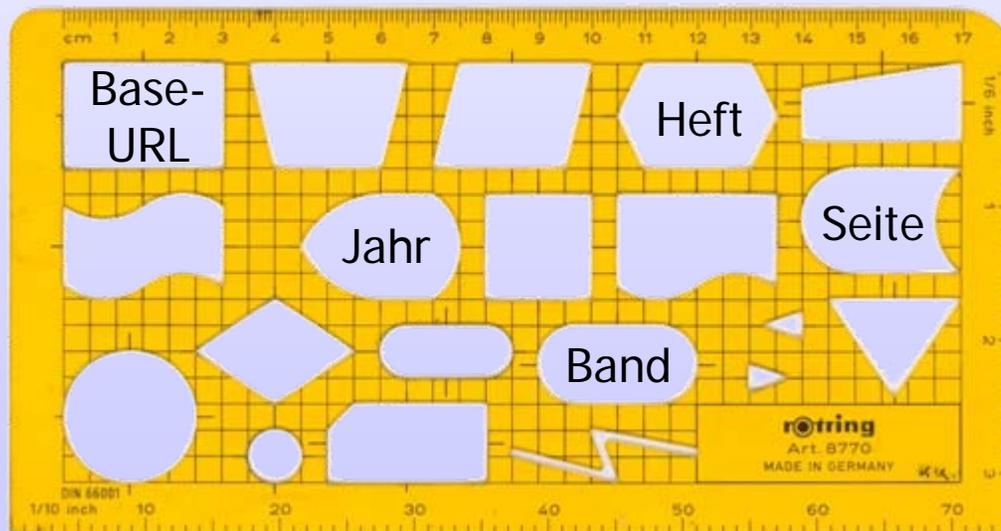
```
http://www.oajournal.net/art?yr=2007&vol=20&iss=9&pg=412
http://www.oajournal.net/art?yr=2005&vol=18&iss=4&pg=238
http://www.oajournal.net/art?yr=2003&vol=16&iss=1&pg=42
http://www.oajournal.net/art?yr=2002&vol=15&iss=7&pg=349
...
```

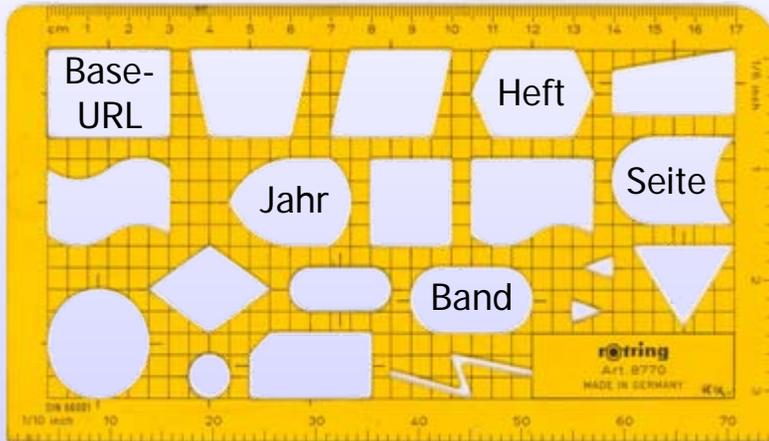
Jahr

Band

Heft

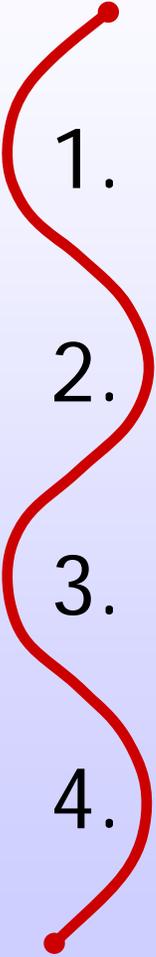
Seite





```
#!/usr/bin/perl
package Link::Constructor::OAJ;
use base qw(Link::Constructor);
use URI;
use strict;
use warnings;

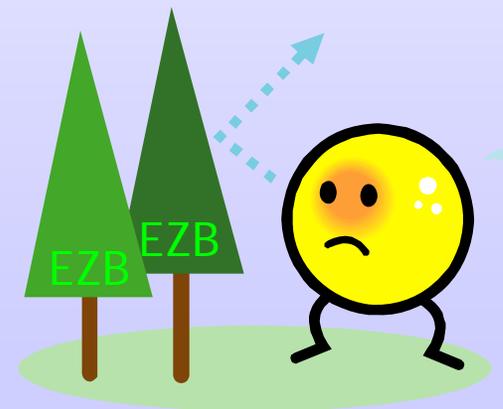
sub linkToFullTxt {
    my ($this) = @_;
    my $base = $this->{'base'};
    my $year = $this->{'year'};
    my $vol  = $this->{'volume'};
    my $iss  = $this->{'issue'};
    my $page = $this->{'spage'};
    my %qry  = (...);
    :
    return URI->new($base,\%qry);
}
```

- 
1. Wozu soll das gut sein?
 2. Wie soll das gehen?
 3. Warum das alles andere als trivial ist ...
 4. Erste Schritte auf vielen neuen Wegen

- Volltexte mitunter nicht auf derselben Domain wie die Journal-Homepage
- Volltexte mitunter nur über interaktive Such-Applets zugänglich, keine Hyperlinks
- Nicht alles was  ist, ist „frei“ genug:



Wir müssen leider
draußen bleiben!



- Fehlende semantische Information bei teilweise oder völlig pfadbasierter ILS:

<http://www.bsp.de/2007/mai/4-12.pdf>

- Rein ID-basierte ILS (z.B. quasi-zufällige Dokumentnummern in CMSen), die sich auch aus noch so guten bibliographischen Metadaten nicht rekonstruieren lässt:

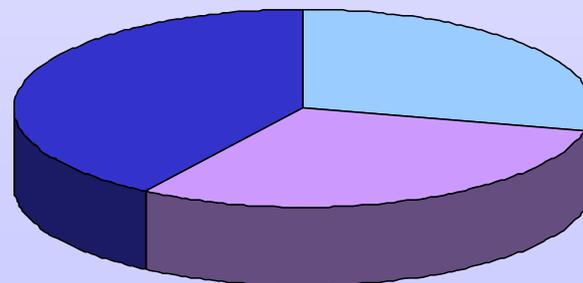
http://clever.org/get_doc?id=4711

- Keine praktikabel evaluierbaren Kriterien für eine **regel-basierte** Klassifikation von URL-Clustern: „1.000 Cluster sind nur wenig übersichtlicher als 100.000 URLs!“
- Keine hinreichend scharfen Trainingssets für **das eine universelle** neuronale Netz (ohne DOAJ gar keine!)

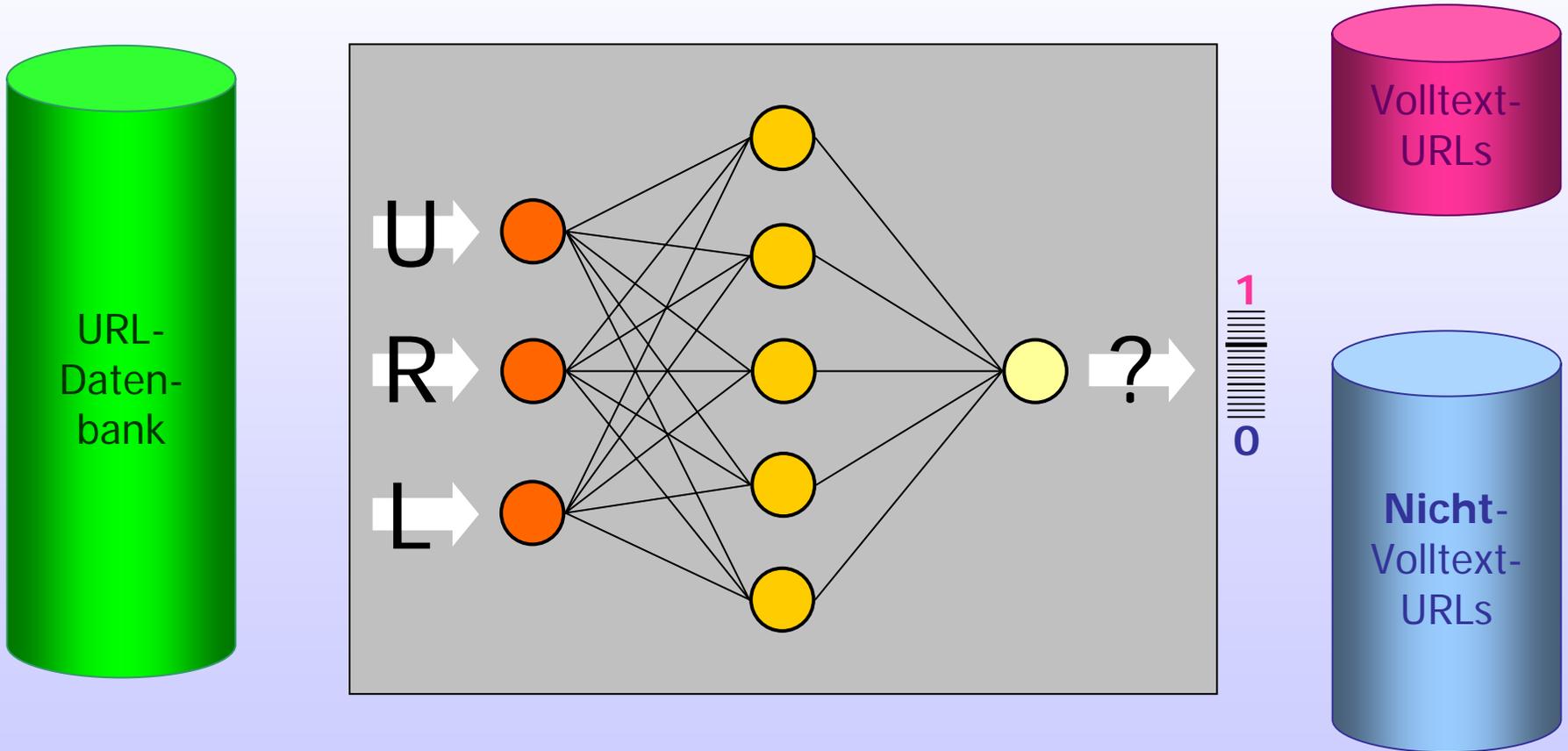
- 
1. Wozu soll das gut sein?
 2. Wie soll das gehen?
 3. Warum das alles andere als trivial ist ...
 4. Erste Schritte auf vielen neuen Wegen

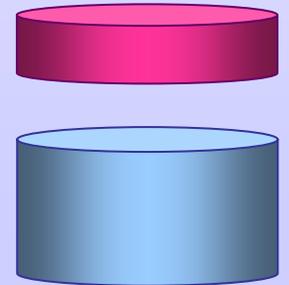
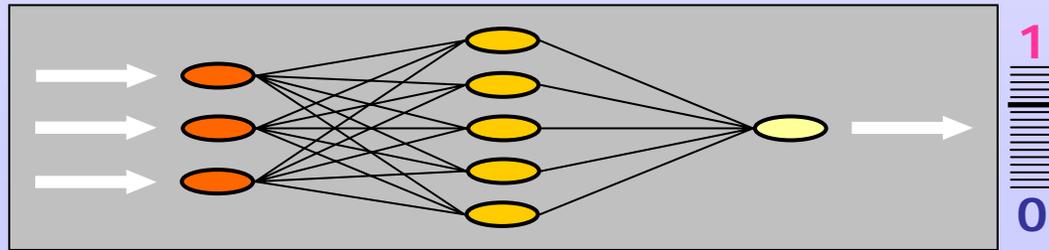
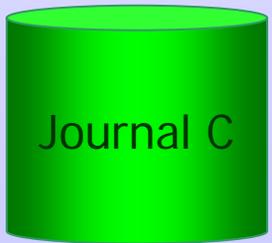
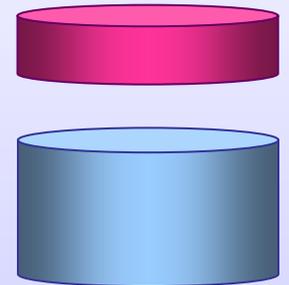
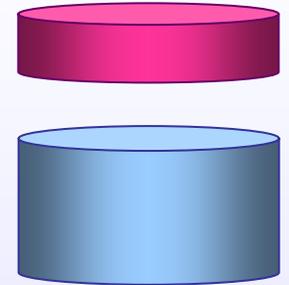
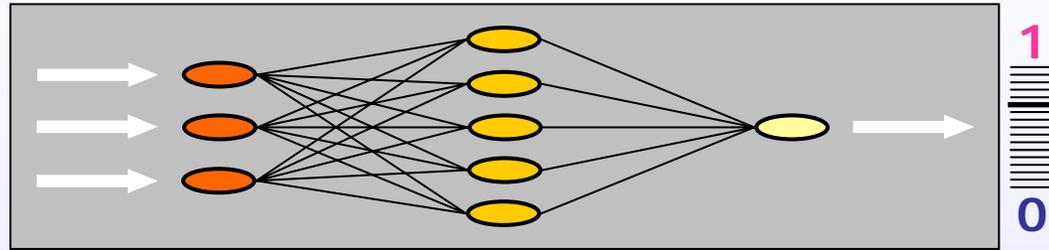
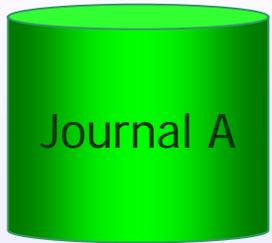


- insgesamt 15.628.185 URLs
- von 275 verschiedenen Journals
- Minimum: 12 URLs / Jnl
- Maximum: 1.399.193 URLs / Jnl
- Durchschnitt: 56.830 URLs / Jnl

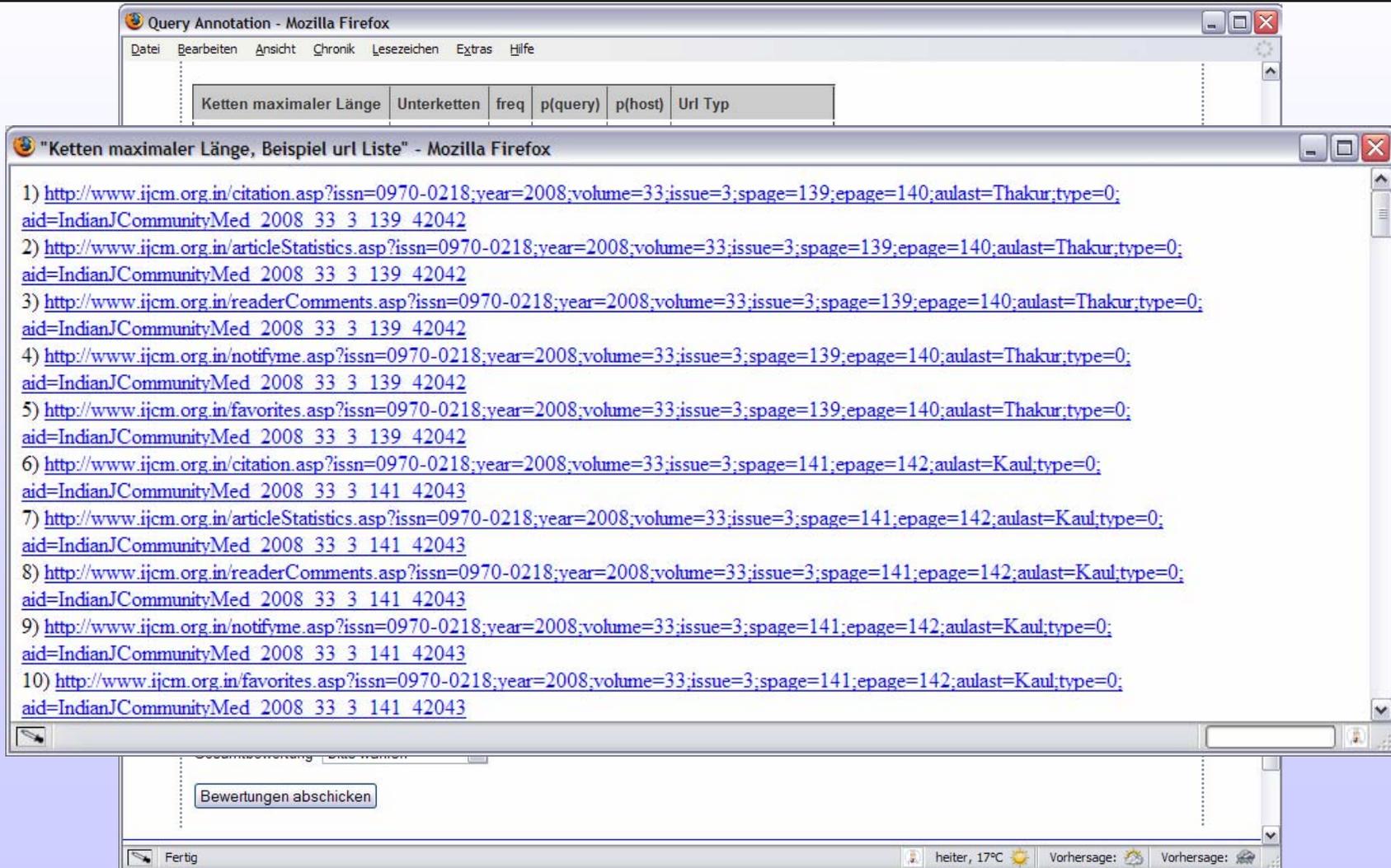


- #URLs < 1001
- 1000 < #URLs < 10001
- 10000 < #URLs





- **Query**: Kette von Parameter-Wert-Paaren in URLs, eingeleitet durch das '?'
- **formal standardisiert** (in RFC 3986)
- trägt **semantische** Informationen, sofern hinreichend sprechende Parameternamen benutzt werden
- Webfrontend zur strukturierten Analyse von Query-URLs pro Hostname: „Wie viele Parameter braucht ein Volltext-URL?“



- „**user generated knowledge base**“
- Besondere Motivation für user = vendor?
- Webformular zur interaktiven Registrierung einer **metadatenbasierten** ILS:
 - 4 möglichst einfache Arbeitsschritte
 - Neuregistrierungen warten auf Freigabe
 - objektorientierte Datenstruktur implementiert URL-Schablone und Link-Konstruktor in einem

Register Your Inbound Linking Syntax With the Generic Link Resolver - Mozilla Firefox

Datei Bearbeiten Ansicht Chronik Lesezeichen Extras Hilfe

Register Your Inbound Linking Synta...

of issues below.
Otherwise you may leave these fields empty and proceed to Step 3.

From: year volume issue

To: year volume issue (leave empty if inbound linking syntax is still valid for the most recent and coming issues)

Step 3 – Example Full Text URL

Please enter now an example full text URL which is "typical" for the journal specified in Step 1:

Step 4 – What Is Where?

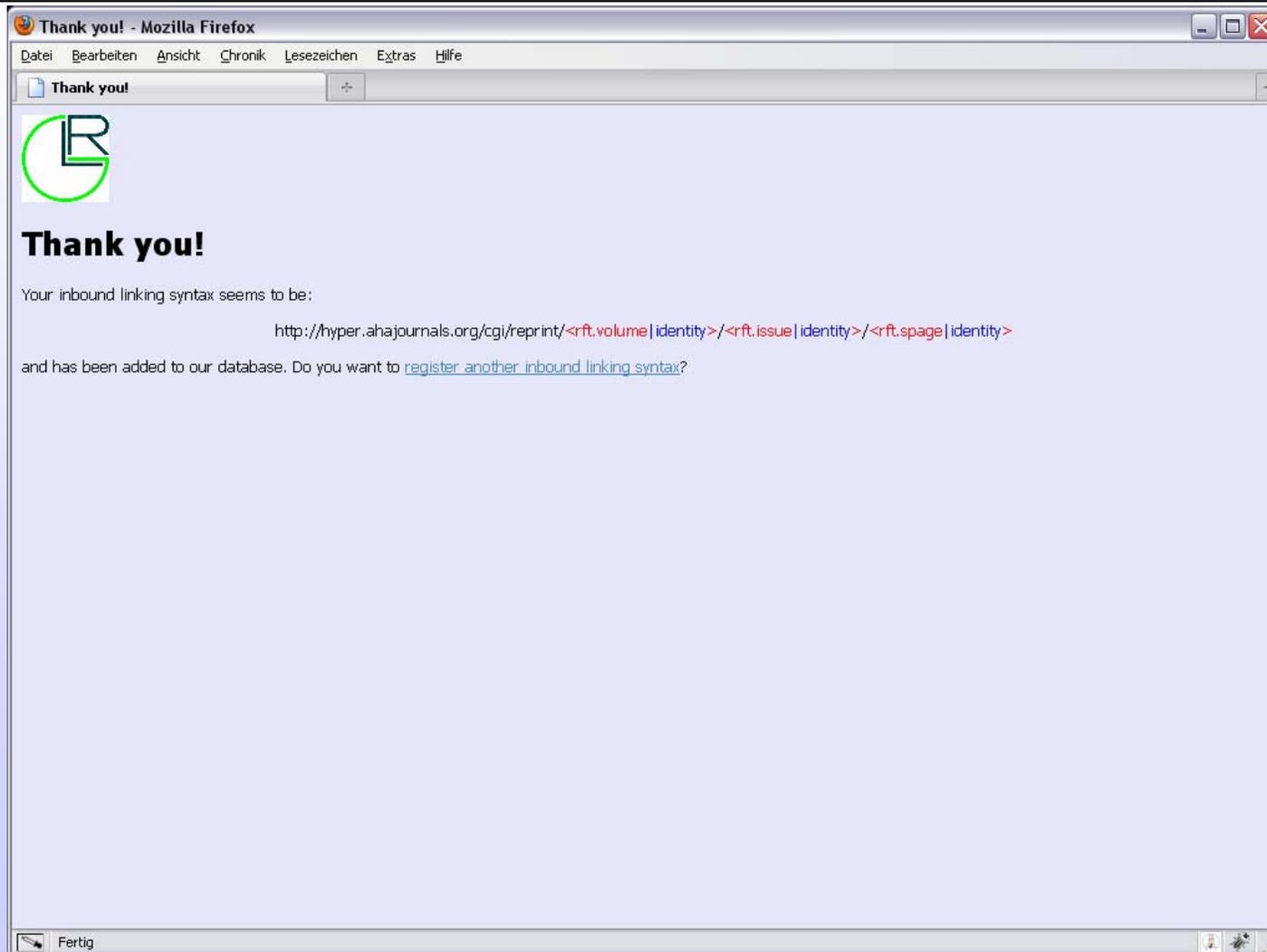
Please explain which pieces of bibliographic information are contained *from left to right* in the full text URL given in Step 3; select the type of information first and then retype the concrete value as it occurs in the URL:

Volume

Issue

Start Page

Fertig



gtr.bib-bvb.de / localhost / deep_linking / URL_TEMPLATE | phpMyAdmin 3.2.0.1 - Mozilla Firefox

Server: localhost ▶ Datenbank: deep_linking ▶ Tabelle: URL_TEMPLATE
"stores inbound linking syntax templates"

Anzeigen Struktur SQL Suche Einfügen Exportieren Importieren

Operationen Leeren Löschen

Feld	Typ	Funktion	Null	Wert
INTERNAL_ID	int(10) unsigned zerofill			0000000001
ISSN	varchar(9)			0362-4323
JOURNAL_TITLE	varchar(100)			Hypertension
JOURNAL_KEY	varchar(30)			hyper.ahajournals.org
JOURNAL_KEY_TYPE	enum	--		HOSTNAME
BASE_URL	varchar(50)			http://hyper.ahajournals.org
YAML_STREAM	text			--- base: http://hyper.ahajournals.org glue: - /cgi/reprint/ - / - / - ..
VALIDITY_PERIOD_START	varchar(14)		<input type="checkbox"/>	1979_0001_0001
VALIDITY_PERIOD_END	varchar(14)		<input type="checkbox"/>	1994_0024_0006
SAMPLE_URL	varchar(255)		<input type="checkbox"/>	http://hyper.ahajournals.org/cgi/reprint/
STATUS	enum	--		APPROVED
SOURCE_OF_RECORD	varchar(60)			share/cgi-bin/templ
RECORDED	datetime			
MODIFIED	timestamp			

Fertig

GLR Service Window - Mozilla Firefox

Datei Bearbeiten Ansicht Chronik Lesezeichen Extras Hilfe

GLR Service Window



GLR Services for this Record

Article: Defective protein phosphorylation associated with hypofunctions in stroke-prone spontaneously hypertensive rat platelets
Author: Tomita, T.
in: Hypertension [0362-4323], vol. 14 , (1989), iss. 3 , pg. 304 - 315

- ♦ [Read the electronic full text of this article](#)
- ♦ [Contact the administrator of the Generic Link Resolver](#)

http://hyper.ahajournals.org/cgi/reprint/14/3/304

- OpenURL-Schnittstelle (Version 0.1 & 1.0)
- damit auch als sog. „Target“ in SFX bzw. analog durch andere institutionelle Link-Resolver nutzbar
- dediziertere Schnittstellen (z.B. auf SOAP-Basis) bei Bedarf nachrüstbar, solange sie **produktübergreifend** einsetzbar sind
- Erfassung grüner Titel in der EZB künftig **mit Registrierung der** zugehörigen **ILS?**

- ... an die DFG für den Mut zur Förderung eines hochgradig experimentellen Projekts!

Deutsche
Forschungsgemeinschaft

DFG

- ... an das DOAJ für seinen Rettungsanker!

DOAJ DIRECTORY OF
OPEN ACCESS
JOURNALS

- ... und an **Sie** für Ihre Aufmerksamkeit!