Enhanced Federated Search for Digital Object Repositories

Christian Kohlschütter
Dierk Höppner
Maria Nejdl

iSearch IT Solutions GmbH

TIB | Technische Informationsbibliothek
Deutsche Zentrale Fachbibliothek für Technik sowie Architektur, Chemie, Informatik, Mathematik und Physik.

Motivation

- Specialized DORs
  There is no "Super-DOR" for everything
- Different implementations
  There is no "eSciDSfeDORprints..." software
- Need interoperability between DORs
  Search *is* key! Keep each DOR *visible*!

Federated Search

- Enable users to perform searches in multiple DORs simultaneously

- Enable search between heterogeneous/ incompatible DORs

"Federated Search"

- Naive Result Merging
  Per-source, Round-Robin, fifo, Random Order, Sort
- No relevance ranking
  Global information is missing
  Statistics are incompatible
  Re-ranking is too costly
- No sophisticated query options
  Not all systems support all features

Federated Search

- Enable users to perform searches in multiple DORs simultaneously
- Enable search between heterogeneous/ incompatible DORs

Enhanced Federated Search

- Enable users to perform searches in multiple DORs simultaneously
  with all the features of a single DOR system.
- Enable search between heterogeneous/ incompatible DORs
  without changing the underlying workflow.

GetInfo
Fachinformationen für Technik und Naturwissenschaften

TIB Technische Informationsbibliothek
Deutsche Zentrale Fachbibliothek für Technik sowie Architektur, Chemie, Informatik, Mathematik und Physik.

Fachsuche Technik

In ausgewählten fachspezifischen Datenbanken suchen.

Technik laser

Suche starten

Erweiterte Suche
Datenbankauswahl

Kurztitelanzeige

Ihre Suche nach **laser** ergab **140708** Treffer.

1 - 20

1 Sicherheit bei der Anwendung der Femtosekunden-Laser-Technologie im Pulsdauerbereich von 100 fs bis 30 ps : Femtosekunden-Technologie, im Rahmen des Förderungskonzeptes "Laser 2000" ... ; EUREKA-Projekt ; Laufzeit:01.02.01-30.04.04 ; Abschlussbericht = Safety for applications of femtosecond-laser-technology
Bunte, Jens; Püster, Thomas; Burmester, Tomas; | **TIBKAT** | 2004
Zur Detailanzeige Zur Merkliste hinzufügen Zur Bestellung

2 Sicherheit bei der Anwendung der Femtosekunden-Laser-Technologie im Pulsdauerbereich von 100 fs bis 30 ps : Femtosekunden-Technologie, im Rahmen des Förderungskonzeptes "Laser 2000" ... ; EUREKA-Projekt ; Laufzeit:01.02.01-30.04.04 ; Abschlussbericht = Safety for applications of femtosecond-laser-technology
Bunte, Jens; Püster, Thomas; Burmester, Tomas; | **TIBKAT** | 2004
Zur Detailanzeige Zur Merkliste hinzufügen Zum Volltext

3 Verbundvorhaben PRIMUS: Präzise Materialbearbeitung mit Ultrakurzpuls-Strahlquellen - im Rahmen des Förderkonzeptes LASER 2000: "Femtosekunden-Technologie" des Bundesministeriums für Bildung und Forschung : Schlussbericht für das Teilvorhaben Diodengepumpte Ultrakurzpulslaser für die präzise Bearbeitung technischer Materialien der Firma Trumpf Laser GmbH + Co. KG
Sutter, Dirk; | **TIBKAT** | 2004
Zur Detailanzeige Zur Merkliste hinzufügen Zur Bestellung

4 Effect of a laser spike on the beat-wave excitation of a plasma wave
Chauhan, P K; Sharma, R P; Purohit, G; | **TIBSCHOLAR** | 2008
Zur Detailanzeige Zur Merkliste hinzufügen Zum Volltext

5 Aufbau eines skalierbaren, fasergeseedeten Optisch Parametrischen

Treffererschließung

Autoren
- unknown (4343)
- Jacques, S. L. (573)
- Geiger, M. (478)
- mehr anzeigen

Dokumentformat
- Print (90335)
- Online-Ressource (31312)
- DOI (18793)
- mehr anzeigen

Dokumenttyp
- Beitrag (84921)
- Aufsatz (49566)
- Buch (2011)
- mehr anzeigen

Erscheinungsjahr
- 2002 (9722)
- 1998 (6543)
- 1994 (6486)
- mehr anzeigen

Sprache
- Englisch (134996)
- Deutsch (2810)
- Japanisch (986)

Features

- Consistent Relevance Ranking (→ Precision)
- "100% Recall" (Coverage)
- Faceted Browsing (Drill-down)
- Support for Custom Query Types
- Connectivity to other Federations
- Extensible Protocols

SRU/SRW

- Search and Retrieve via URL
- Widely used standard by Library of Congress
- Queries as Strings (HTTP GET or SOAP)
  http://z3950.loc.gov:7090/voyager?version=1.1&operation=searchRetrieve&query=dinosaur&recordSchema=dc&maximumRecords=10
- Response as XML

SRU/SRW

Mozilla Firefox

http://z3950.loc.gov:7090/voyager?version=1.1&

– <zs:searchRetrieveResponse>
<zs:version>1.1</zs:version>
<zs:numberOfRecords>1968</zs:numberOfRecords>
– <zs:records>
– <zs:record>
<zs:recordSchema>info:srw/schema/1/dc-v1.1</zs:recordSchema>
<zs:recordPacking>xml</zs:recordPacking>
– <zs:recordData>
– <srw_dc:dc xsi:schemaLocation="info:srw/schema/1/dc-schema http://www.loc.gov/standards/sru/resources/dc-schema.xsd">
<title>3-D dinosaur adventure [computer file].</title>
<creator>Knowledge Adventure, Inc.</creator>
<type>software, multimedia</type>
<publisher>Glendale, CA : Knowledge Adventure,</publisher>
<date>c1995.</date>
<language>eng</language>
– <description>
Employs a dinosaur theme-park setting to introduce users to Triassic, Jurassic, and Cretaceous periods. Features hypertext dinosaur encyclopedia covering 150 million years of paleontology. Includes animated video simulations, three-dimensional dinosaur museum, narration, games, activities, and color illustrations.
</description>
<description>Ages 5 to 10.</description>

Fertig

SRU/SRW

- Scan: Term Vocabulary statistics
- Explain: Describe server's capabilities
- Diagnostics: Error Handling
- CQL: String-based query language

  fish
  dc.title any fish sortBy dc.date/sort.ascending
  dc.title any/rel.algorithm=cori fish
  dc.title any fish prox/unit=word/distance>3 dc.title any squirrel

SRX/FS

- Built upon SRU/SRW
- Special operation to exchange statistics
- Faceted Search support
- Additional Query Types
- XML-based Query Language

XCQL/FS

- Resurrect XCQL

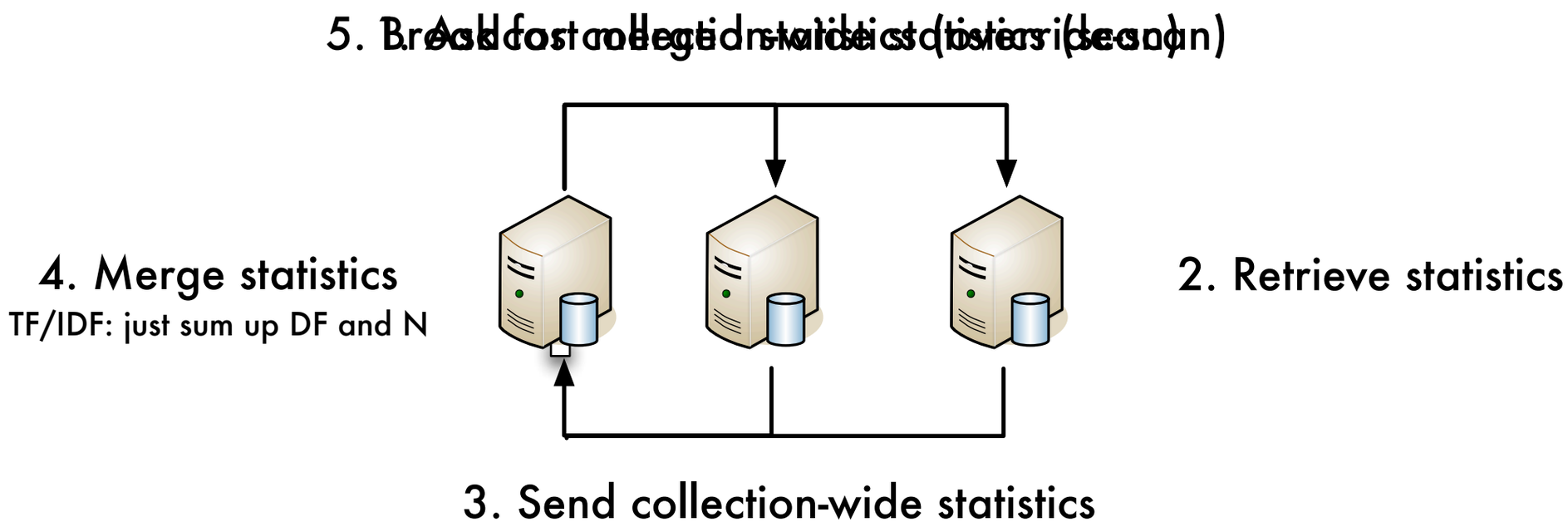
```
<searchClause>
    <index>dc.title</index>
    <relation><value>=/string/unmasked</value></relation>
    <term>fish</term>
</searchClause>
```

- Extension and Simplification

```
<fs:booleanQuery coord="true" minOptMatch="1" maxOptMatch="*">
    <fs:clause occur="should">
        <fs:termQuery index="dc.title" term="fish" />
    </fs:clause>
    <fs:clause occur="should">
        <fs:termQuery index="dc.title" term="squirrel" />
    </fs:clause>
</fs:booleanQuery>
```

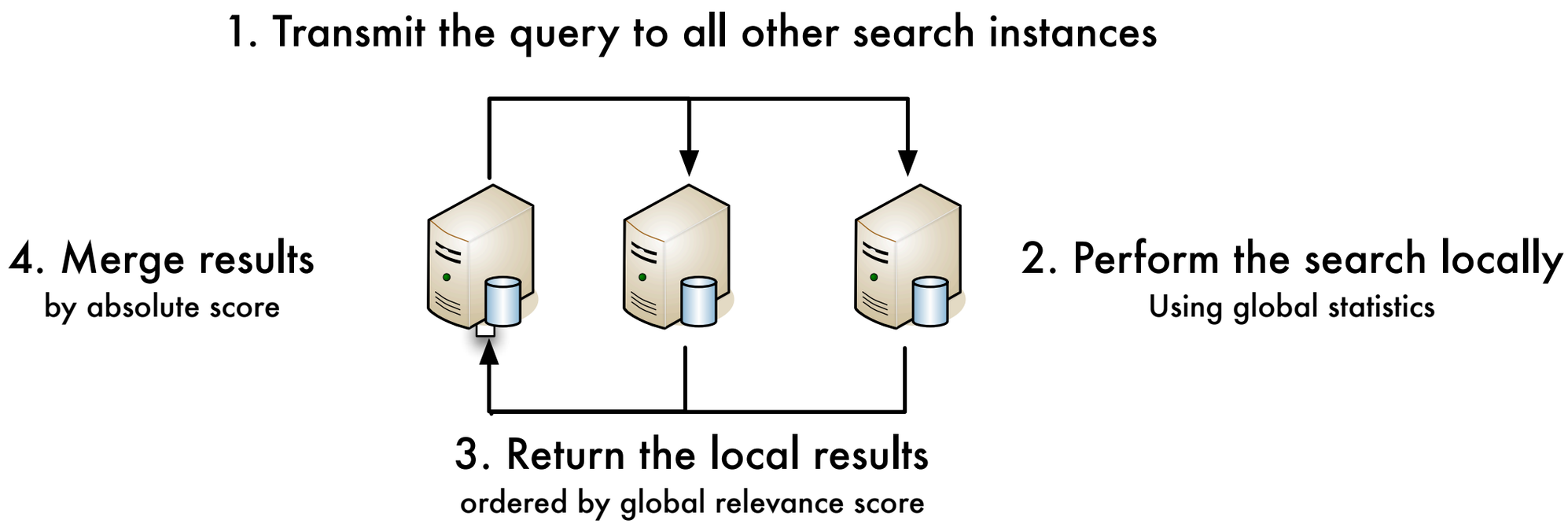
Relevance Ranking

- Exchange of Statistics through "override-scan"



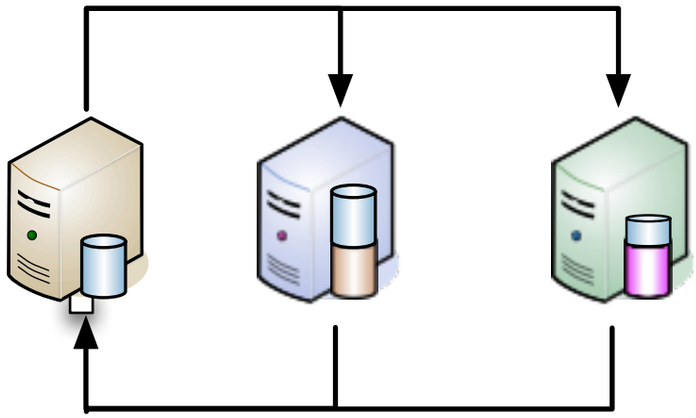
Relevance Ranking

- Ranking is trivial now



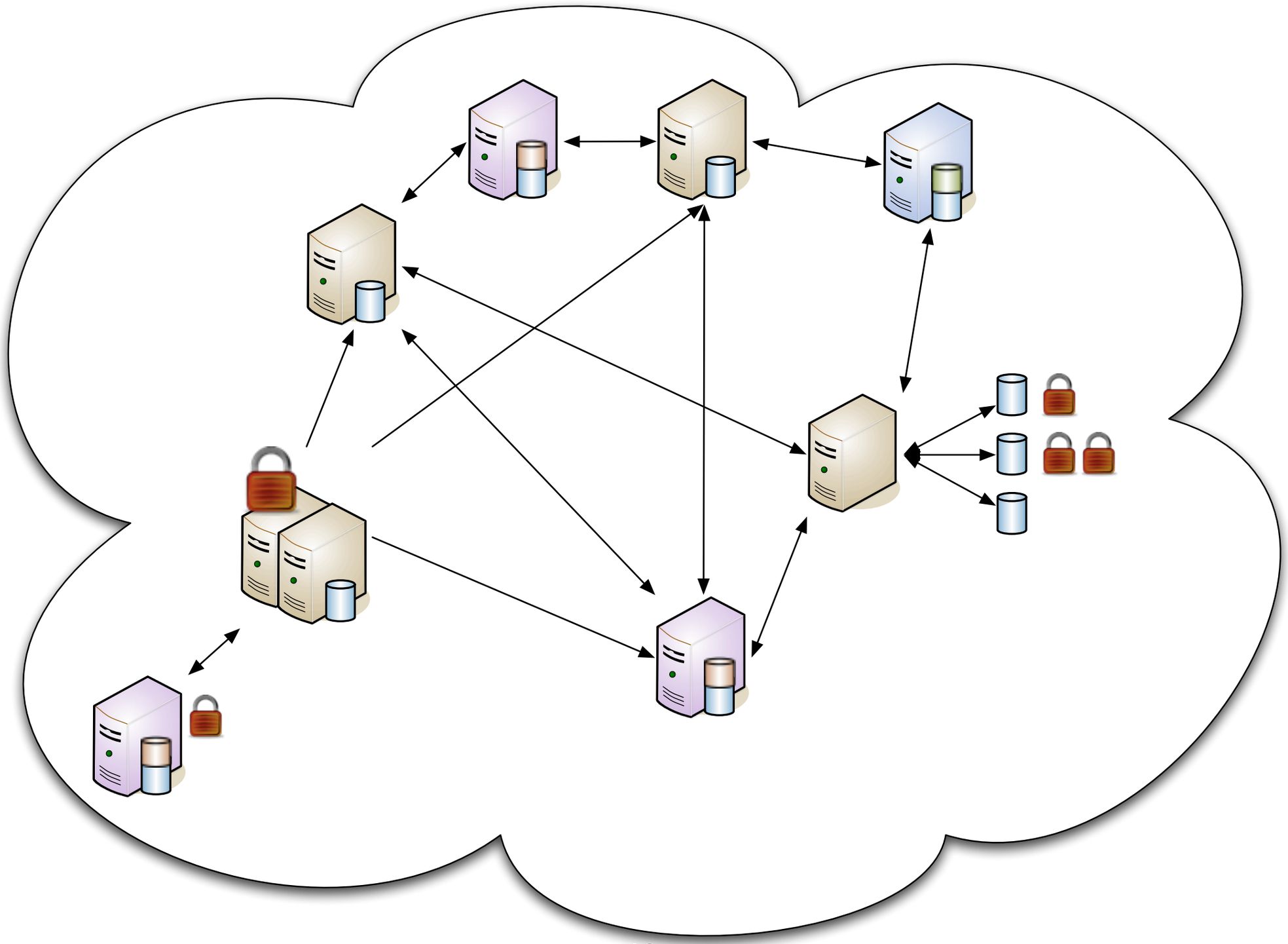
Relevance Ranking

- How to deal with heterogeneity?
- Wrap the existing systems by Plug-ins [ICADL06]



Sergey Chernov, Christian Kohlschütter, Wolfgang Nejdl: A Plugin Architecture Enabling Federated Search for Digital Libraries. ICADL 2006: 202-211

The Big Picture



Implementation

- API is engine-independent
- Reference Implementation based upon Lucene
- Very efficient Faceted Browsing functionality
- Lucene-based Plugin to support other vendors
- Custom Lucene Searcher class (easy integration)
- Interface to vascoda Federation

Evaluation

- Federation between TIB and FIZ Technik
- 3 M bibliographic objects
- Homogeneous Relevance Ranking
- Search times ~50-100ms
- Drill-down times ~100ms-3secs.

Challenges

- Integration of SRX/FS into DOR systems
  Fedora GSearch, Summa, eSciDoc, aDORe, ...
- Integration of DORs into the TIB Federation
- Protocol Standardization
- Additional Features (Query Types, Security, ...)
- Open Source License

Christian Kohlschütter
kohlschuetter@isearch-it-solutions.de