TILBURG ◆ UNIVERSITY

# Is Integrated Search the answer of libraries to Google?

Thomas Place
Innovation manager Academic Services
Library and It Services
Tilburg University (NL)

Presentation for X Workshop Rebiun sobre Proyectos Digitales
Valencia, Friday, 8 October 2010

TILBURG ◆ UNIVERSITY

Overview

- A Personal History

- The Integrated Search project of Tilburg University

- Users

- Local versus hosted integrated search solutions

- Selecting a hosted solution

A Personal History
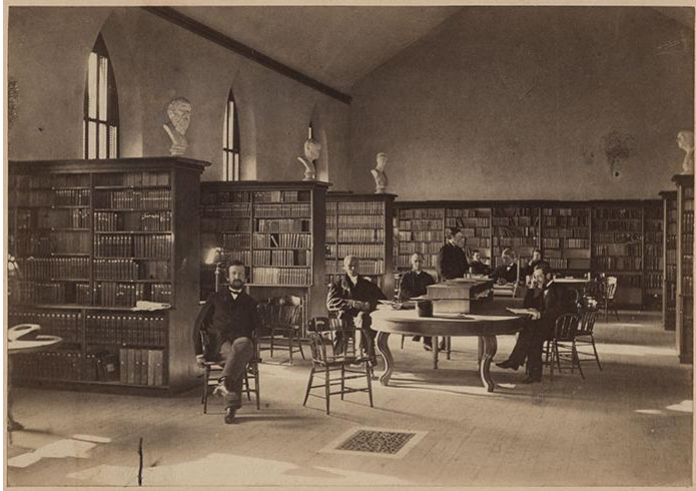
History: search @ Tilburg University

- 1988   Searching only in the library

- 1992   Searching moves to the desktop

- 1995-1997   Homogeneous search interface

- 2001   Metasearch plus dynamic linking: iPort

- 2009   Integrated search: Get It!

- 201?   Searching completely in the cloud

1988 Searching only in the library

- Psychological Abstracts: print index (appeared during1927 – 2006)
- Social Sciences Citation Index: print index (appeared during1973/4 - ????)
- OPAC terminals – Online in the library → Public
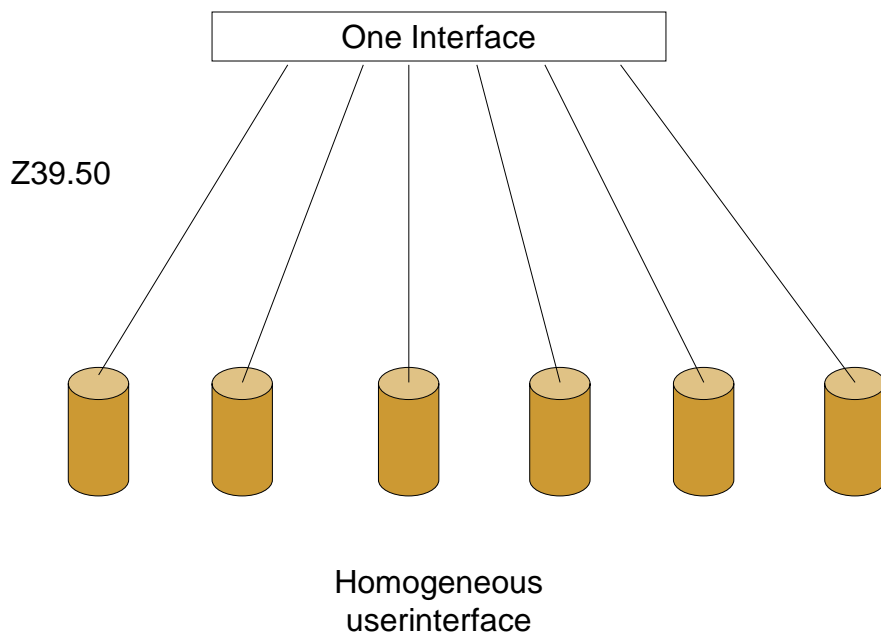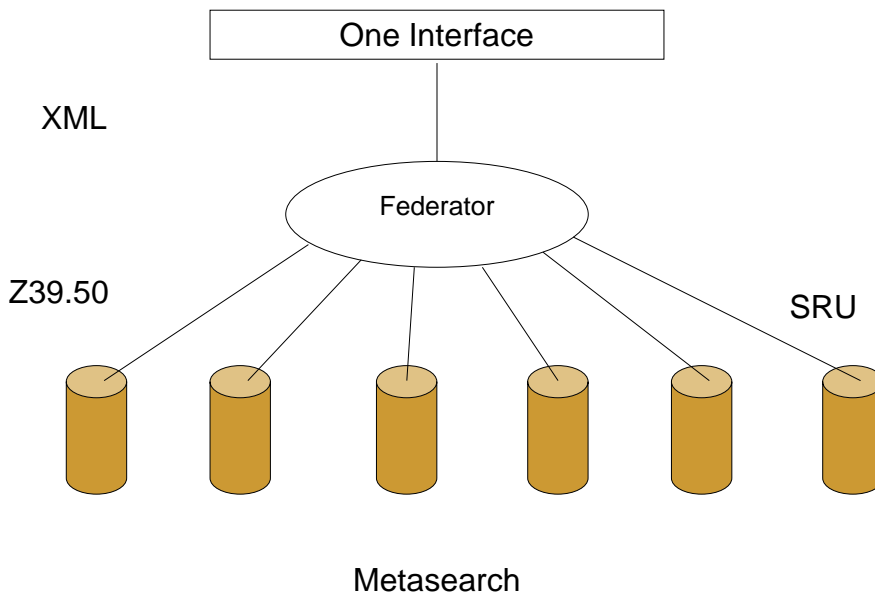- Stand alone PC in the library with CD-ROMs: PsycLit, SSCI

1992 searching moves to the desktop

- New library building
- New library system: LBS3 of Pica (now OCLC)
  - building of database from union catalogue took weeks: transfer by tapes
  - updates: online
  - OPAC: accessible via the Internet (telnet)
- Tilburg the first Dutch university with a Campus Wide Information System (1991) with entry points for the local bibliographical databases:
  - Catalogue
  - Excerpta Informatica
  - Online Contents: journal articles
  - Student theses
  - Attent: reports in economics
  - Brabant database

  and for external databases on the internet.
- CD-ROMs available via campus network

1995-1997 homogeneous search interface

- All local databases (Trip) have Z39.50 interface; exception: the catalogue
- Z39.50 MS Windows client (Kwik)
- Soon replaced by a Web application (Trix)
  - Homogeneous access to internal and external Z39.50 databases via a Web browser (Netscape)
  - Each database was, however, searched separately like in 1988 with the print indexes.
  - Users didn't understand that Catalogue is for books and journals, Online Contents is for articles, etc. Default selection is the first database in the list

One Interface

Z39.50

Homogeneous
userinterface

One Interface

XML

Federator

Z39.50

SRU

Metasearch

TILBURG ◆ UNIVERSITY

2001 metasearch plus dynamic linking

- European project Decomate II ➔ **iPORT**

- **iPORT** commercialization by OCLC PICA,
    software development by Tilburg University
      not at the market anymore, other products are …

- First Dutch implementation of metasearch; still running.

- Database lists, homogeneous userinterface for SRU/Z39.50 databases, metasearch, de-duplication, dynamic linking to fulltext + OpenURL resolving, book shelves, current awareness services

- Local databases only available via user interface of **iPORT**

- User interface conforms to house style

Problems with metasearching

- the performance is sometimes disappointing (no Google-like performance)
- the presentation of the information is not optimal (merging, sorting)
- users find it difficult to select the right databases for a federated search (as a solution they select all databases which has a negative effect on the performance and increases the noise in the search results).
- users don't know how to formulate the best queries for the databases they have chosen (in many cases this is also not possible because a query that is optimal for one database is not the optimal query for another database in which the user also wants to search – indexes differ over db's).

One Interface

Z39.50

Homogeneous
userinterface

One Interface

XML

Federator

Z39.50   SRU

Metasearch

One Interface

XML   SRU

OAI-PMH
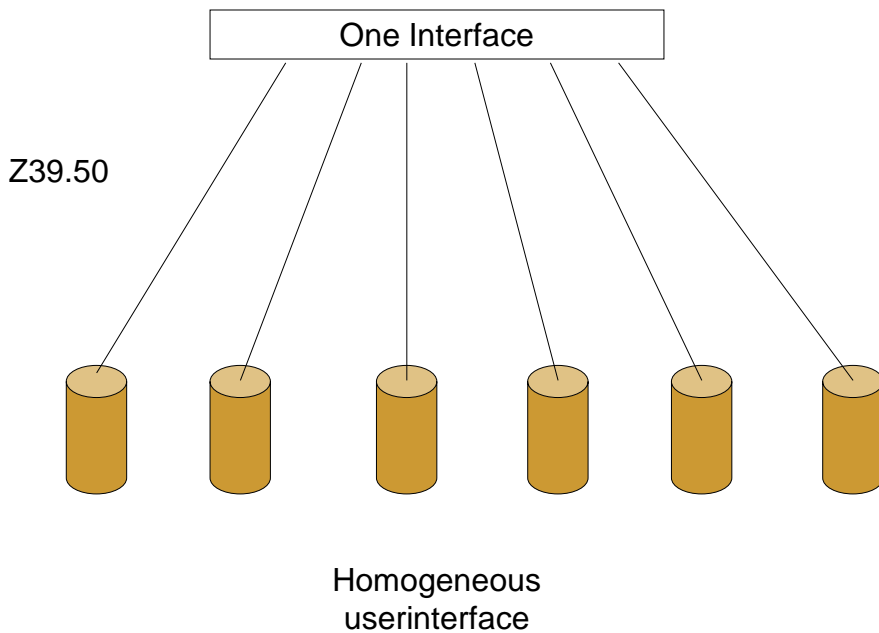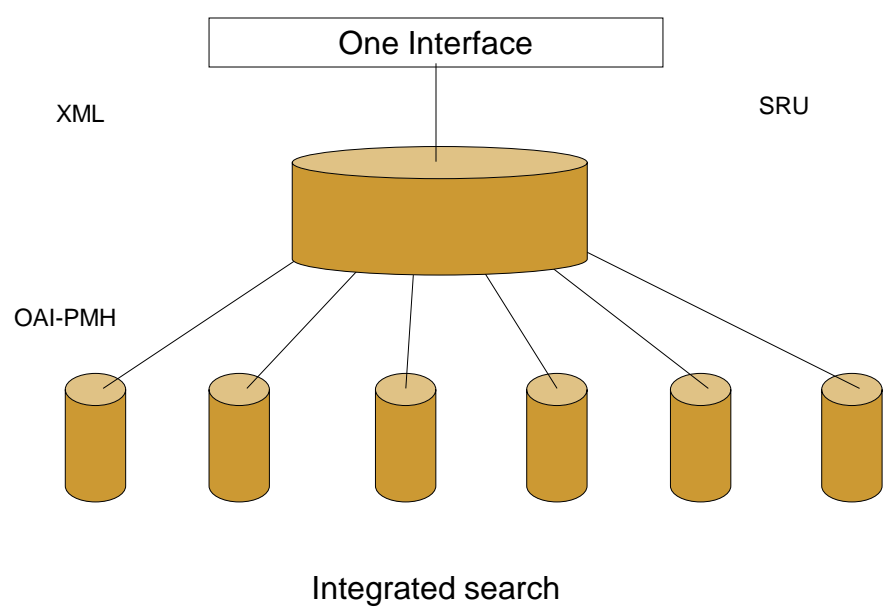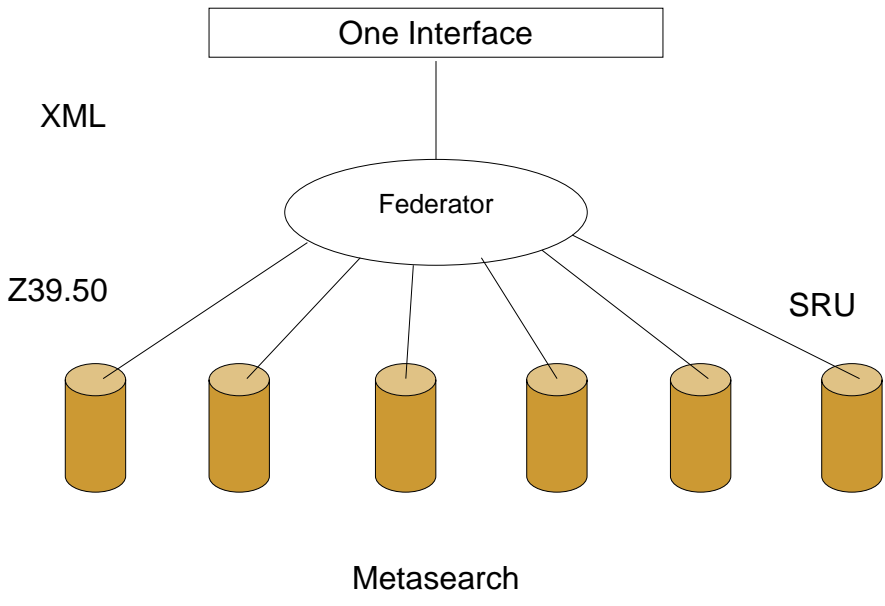
Integrated search

2009 Integrated search

- Page with databases is no longer the start, but the search box.
  - No database selection just search   demo of Get It!

- Technical solution: > MERESCO < of CQ2   SEEK YOU TOO B.V.
  - Open Source

- We worked together with the TU Delft who implements also Meresco: Discover

- Meresco infrastructure is also used for special services, e.g., Economists Online

What were the goals of the Tilburg project?

- To implement the one and only search engine of Tilburg University
- Searching scientific information (library) AND non-scientific information (website, learning material)
- QuIery leads (in the future) to
  - Relevant documents and web pages (Get It!)
  - Experts (expert finding system developed by master student)
  - Specialised databases
- Finding of documents: no longer clicking to full record display; most important information is directly presented in result list
- Informing the user about the search results: facets, clusters
- Added value: add-ons / mash-ups, integration in the workflow

Project organisation

- Project manager / developer
  - Reporting to MT of Library and It Services

- User interface group
  - Tasks
    - Functional requirements
    - Usability study
  - Members
    - information/subject specialists (2)
    - service desk librarian
    - digital library manager

Project organization (2)

- Technical group
  - Database specialist
  - User interface specialist (html, css, javascript)
  - Java programmer ('place' locator)
  - Developer (Python, javascript, Ajax)
  - Developer / system administrator

- Overlap with technical group responsible for the development of the subject portal Economists Online

Project organization (3)

- Regular contact with
  - Team University Web site
  - IT department of Library and It services
  - Academic Services of Library and It services
  - Library and It Service desk

- External contacts
  - Seek You Too – Meresco
  - Technical University of Delft
  - National working group Innovation Digital Library
    - Architecture
  - Search & Find 2.0: Informal group of experts from ULB (Brussels), U Gent, TU Delft, Tilburg U, Royal Library (NL), CQ2

What is now (October 2010) in the search engine?

Catalogue	691.000
Online Contents (local)	2.368.000
Elsevier	2.200.000
Institutional Repository	56.000
Student theses	16.000
RePEc	900.000
EO repositories	78.000
Total	**6.309.000**

What will be added?

Brabant databases   includes an image database
JSTOR
Springer
Linkdatabase (e-holdings)
Fulltext (EO, IR, Theses)

Users (Delft)

- **Students** lack an overview of the domain in which they search. They are inexperienced searchers and **don't know the terminology** of the disciplines in which they search. The challenge for students is to find structure in the chaos of information.
- Students search without a clear plan. They **want to be able to revisit earlier search paths**. This is not well supported by present systems.
- When a student starts searching there is no clear idea of what (s)he is searching for. During the search process their information need becomes gradually more clear and they discover the relevant search terms.
- For students it is **difficult to verify the trustworthiness of the information** that they find during searching..
- The way students search is not very well organised. They change strategies and goals. They are very receptive for unexpected results (**serendipity**) which give them new leads for searching more information.
- Students don't know RSS

Metalib statistics of the University of Groningen

50% zero or false results:

- Misspellings and typos in search terms
- Picking databases at random
- Unable to understand QuickSearch, MetaSearch, Find Database
- Using the wrong search keys
- Using search keys wrongly
- Using Dutch search terms in English language databases
- Using non-specific terms, phrases that are too broad
- Lack of understanding of Boolean logic or database peculiarities

Metalib statistics

Local versus hosted solutions

- Last year, local solutions were hot:
    - Meresco,
    - Primo,
    - Summa,
    - VuFind

- This year, the floor is to hosted solutions also known as discovery services:
    - EBSCO Discovery Service
    - Primo / Primo Central (Ex Libris)
    - Summon (Serials Solutions)
    - Worldcat Local (OCLC)

Why a local solution?

- You already started with a local solution before there were hosted solutions.
- You want complete control
- You want flexibility
- You want to be able to innovate
- E.g., Tilburg uses local integrated search solution for subject portal for Economics and for integrating datasets.

Why not a local solution?

- You don't have the human resources (e.g., developers).
- You don't have the expertise.
- You don't have the money,
- Negotiating content takes a lot of time.
- Cleaning and converting metadata is a lot of work
- You don't have the time; you and your users cannot wait.

Why a hosted solution?

- It is there.
- It is relatively cheap – there is a lot of competition and there is the advantage of the scale
- It is easy
  - You can concentrate on your users in stead of spending your time on technical issues or on acquiring content.
  - Within a few months it is up and running for your users. [What happens if all libraries want it now?]

Which hosted solution to select?

- There is a lot of competition. The best solution today will be second best tomorrow and so on.

- The suppliers use different business models and have different backgrounds.
  - subscription agent, library automation system vendor, members organisation, databases host
  - a certain business model or background can be attractive because of your local situation.

Possible selection criteria

- Content
  - Coverage
    - Some publishers are reluctant to supply their metadata and full text
  - Quality of metadata
    - Article metadata from toc databases -> no subject metadata
    - How important are subject metadata if the majority of the searches are for known items?
- E-shelves / downloading / integration with ref. managers

More criteria

- Personalization – Library 2.0
  - RSS feeds
  - Tagging
  - Ratings
  - Reviews, annotations
- API
- Integration with local solution
- Usage statistics
- Fulfillment.
  - in one click to the full text
  - ***Availability info must be actionable***: on loan = reserve, in depot = request

More criteria

- Integration of (article/database) recommender service
  - E.g., bX
- Look and Feel can be customized
- Management of holdings in one place
  - If locally SFX, not same work in hosted solution.
- Search performance: google-like
- Multilingual search: searching with Spanish keywords in English content and vice versa.
- Access
  - No logon at the front door
    - No access hurdles for free material

More criteria

- Authentication/authorisation
  - Federation
  - Shibboleth
  - LDAP
- Provider
  - Quality
  - Reputation
  - Support
- Search
  - Indexes – same as those of the traditional catalogue?
  - Query language – Google like? Booleans? Searching on fields?
  - Search History

Criteria related to the search results

- Facets
  - Timeline
- Clustering
- FRBR
- Own material on top
- Full text indexing
- Relevance
  - What is relevant?
  - Boosting
  - Record size must be taken into account
    - full text items must not dominate the metadata-only items.
- Remember: most searches are for known items!

TILBURG ◆ UNIVERSITY

What do you think?

Do we have as libraries an answer to Google (Scholar)?