A Semantics-Based Graph for the Bib-1 Access Points of the Z39.50 Protocol

Michalis Sfakakis & Sarantos Kapidakis

Laboratory on Digital Library and Electronic Publishing
Archive and Library Sciences Department
Ionian University, Corfu, Greece
{sfakakis, sarantos}@ionio.gr

10th European Conference on Digital Libraries
17-22 September 2006, Alicante, Spain

Necessity of the Bib-1 Semantics-Based Graph

2

- Clarification and better understanding of the:
  - exact semantics of every Access Point
  - Access Point inter-relationships

Graph Usage

3

- When Z39.50 query transformations are involved either for:
  - Query optimization
  - Access Points replacement due to unsupported Access Points, or in heterogeneous information retrieval environments
- As a guidance to the end user for posing more precise queries

Z39.50 Query mechanism

4

- The query mechanism specifies
  - Predefined abstract Access Points combined with specific attributes (Attribute Sets)
  - Query languages (query types)
- Recognition (not necessarily implementation) of the Attribute Set Bib-1 is required by the general conformance requirements of the protocol
  - The *use* attribute in the Bib-1 specifies an Access Point (e.g. Title, Author, etc.)

Access Points Semantics

5

- The semantics of the Access Points are defined in the "*Attribute Set BIB-1 (Z39.50-1995): Semantics*" document
  - Which represents consensus among the members of the Z39.50 Implementors Group (ZIG)
  - Maintained as an official document of the Z39.50 Maintenance Agency
  - Defines the semantics of the Access Points using the tag values of representative MARC bibliographic format fields
- An example
  - Access Point *Author-name-Personal* (or *use* attribute with value 1004) includes the data from the fields with MARC Tags {100, 400, 700, 800}

Method Description

6

- The procedure for the creation of the graph consists of three steps
- We create the graph of the Access Points, according to their subset relationships

Access Points subset relationship

7

- An Access Point is considered as a subset of an other one, if the set of the data fields used to create the first is a subset of the set of the data fields used to create the second
  - An example:
    - *Author-name* = {100, 110, 111, 400, 410, 411, 700, 710, 711, 800, 810, 811}
    - *Author-name-personal*={100, 400, 700, 800}
    - The Access Point *Author-name- personal* is considered being a subset of the *Author-name*

Graph Specification

8

- We represent the relationships between the Access Points with a directed graph *G*
  - Vertices represent Access Points
  - Edges represent subset relationships
- <*i, j*> is an edge of the graph if and only if Access Point *i* is a subset of the Access Point *j*
- The Access Points *Author-name* and the *Author-name-personal* will be represented by two vertices of the graph and their subset relationship from the edge <*Author-name-personal, Author-name*>

Graph Construction - Illustration

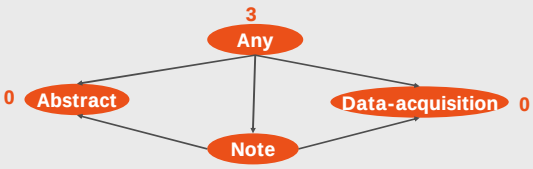
9

- Let's consider that the Bib-1 Attribute Set consists only of the next four Access Points:
  - *Any*. From its definition, can be thought as the superset of all the supported Access Points
  - *Abstract*={520}
  - *Data-acquisition*={541-subfield-d}
  - *Note*={500, 501, …, 520, …, 535, 536, …, 541, …, 586}
- We can see that:
  - All Access Points are subsets to *Any*
  - Access Points *Abstract* and *Data-acquisition* are subsets of the *Note* Access Point

Construction of the Graph: Step 1

10

- Vertices represent the Access Points
- Edges represent the subset relationships

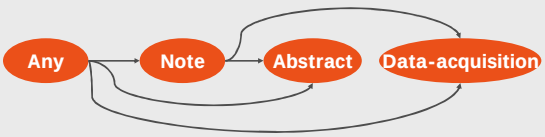


- Graph G

Construction of the Graph: Step 2

11

- Application of the topological sorting algorithm
  - The ordering is feasible due to the transitive and irreflexive properties of the proper subset relation



- Graph G after the topological sorting

Construction of the Graph: Step 3

12

- Remove edges implied by transitivity (i.e. delete the derivate subset relationships)
  - Number the vertices of the graph in the previous step from left to right
  - For each vertex, keep the incoming edge from the highest numbered vertex only

- Graph G0, the minimal subset of G

A practical Use of the Graph - I

13

- Consider two Z39.50 sources where:
  - The first supports the Access Point *Author-name* (i.e. can answer queries using this Access Point)
  - The second supports the Access Point *Author-name-personal*
  - Assume that both sources support the same value combinations for the remaining attribute types
- Obviously, all request to the first one for selecting data using the Access Point *Author-name-personal* will fail

A practical Use of the Graph - II

14

- A smart client could substitute the Access Point *Author-name-personal* with the Access Point *Author-name* into the queries, taking into account their subset relationship
  - The client could avoid the query failure
  - Unavoidably, the precision of the resulting query will be less than the precision of the original one

Future Extensions

15

- To create the semantics-based graph of the Access Points according to their intersection relationship (common fields)
- To check if the graph with the subset relationship is possible to combined with the graph with the intersection relationship and how.
- To apply our methods using the Bib-2 Attribute Set
  - The Bib-2 is designed for searching databases containing descriptions of a wide range of bibliographic materials