Value-Aware RoXSum: Effective Message Aggregation for XML-Aware Information Dissemination

MIRELLA M. MORO – UCR/UFRGS (Brazil)

Joint work with

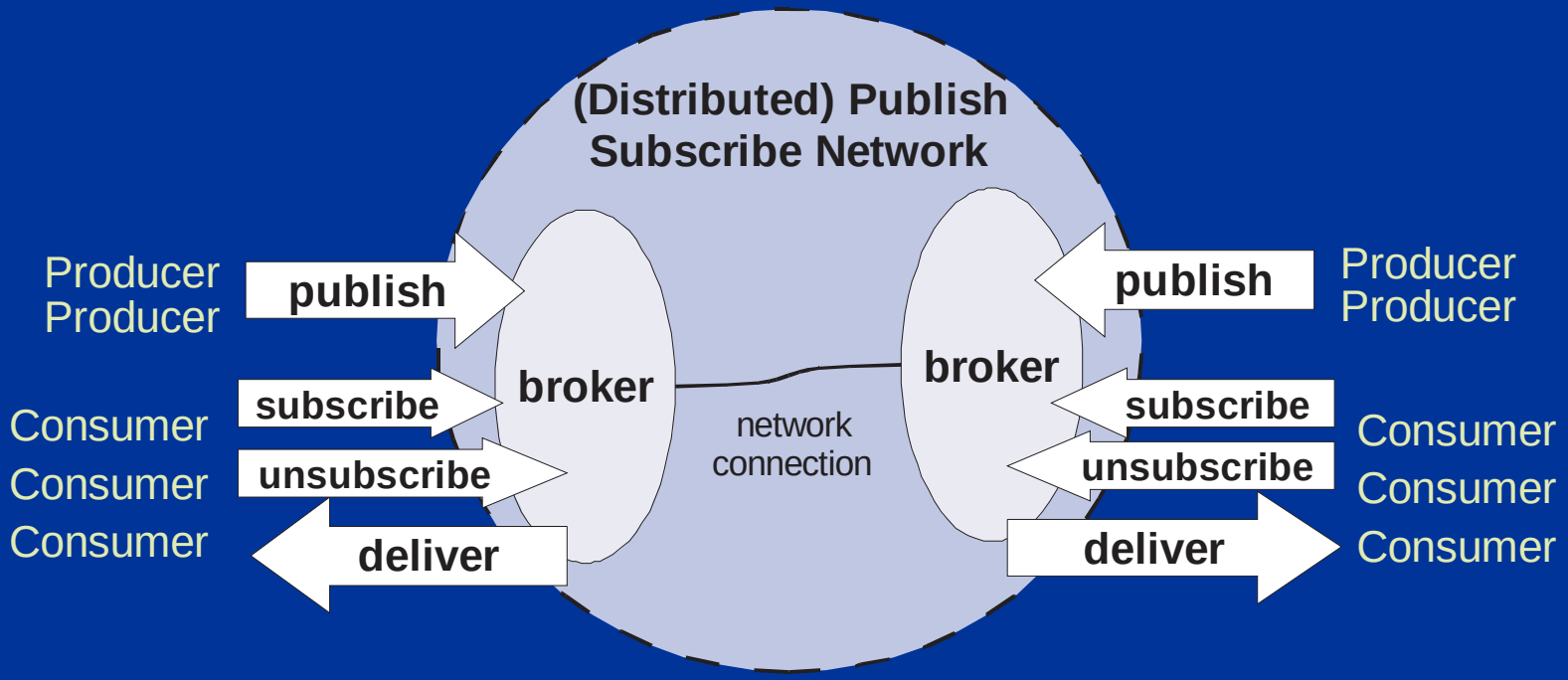
Zografoula Vagena - IBM

Vassilis J. Tsotras - UCR

INTRODUCTION

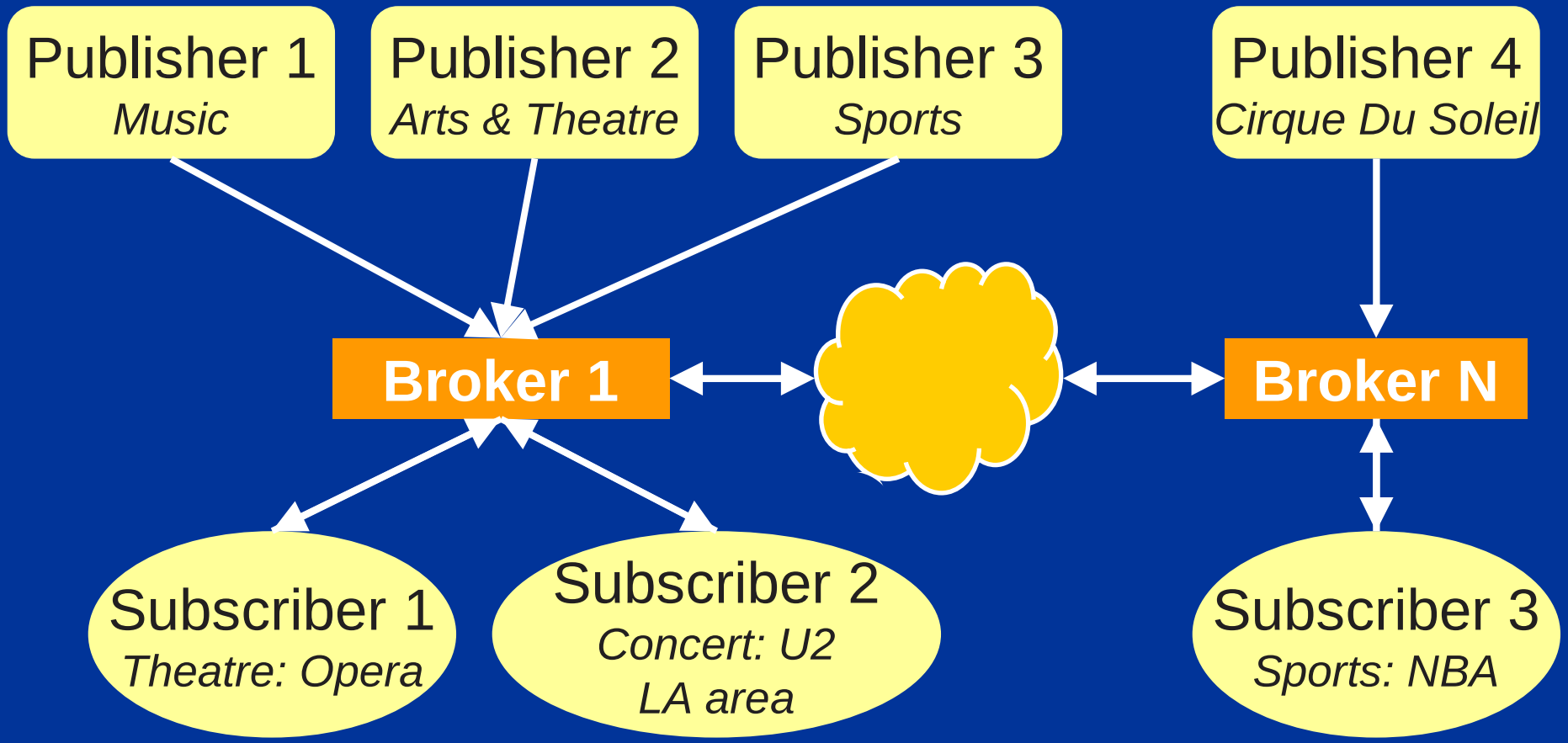
Content Based Routing

- Traditional delivery: destination IP address
- ➔ Data delivery defined by message content
- Publish/Subscribe applications (pub/sub)



- Examples: hotwire, ticketmaster

Publish/Subscribe System



XML Aware Pub/Sub System

- Incoming Messages are encoded in XML
  - Many small documents
- User profiles are expressed using an (subset of) XML query language
  - Path expressions

/bibliography//book[./author="A_1"]

  - Two types of constraints
    - /bibliography//book and book/author
    - author = "A_1"

Related Work/Our Contribution

- Current work
  - Construction of overlay network
  - Dissemination/indexing of profiles (queries)
  - Processing of stream of messages
- Our proposal
  - RoXSum (Routing XML Summary) structure to aggregate messages [ICDE'07]
    - Batch processing of profiles over aggregated content
    - Structural constraints
  - Value-Aware RoXSum (VA-RoXSum)
    - Batch processing + Value aggregation

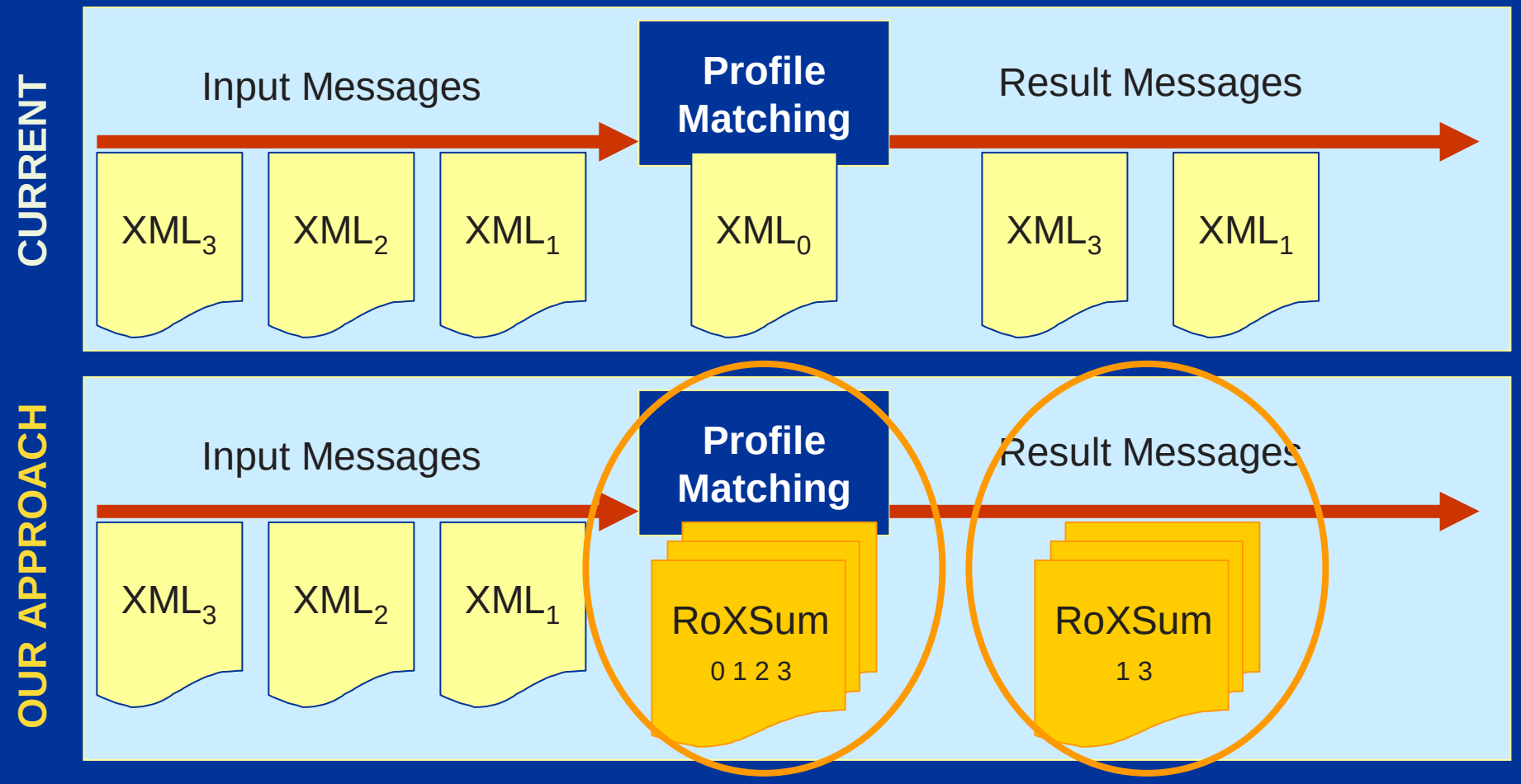
Outline

- RoXSum Data Structure
- VA-RoXSum
- Experimental Evaluation
- Final Remarks

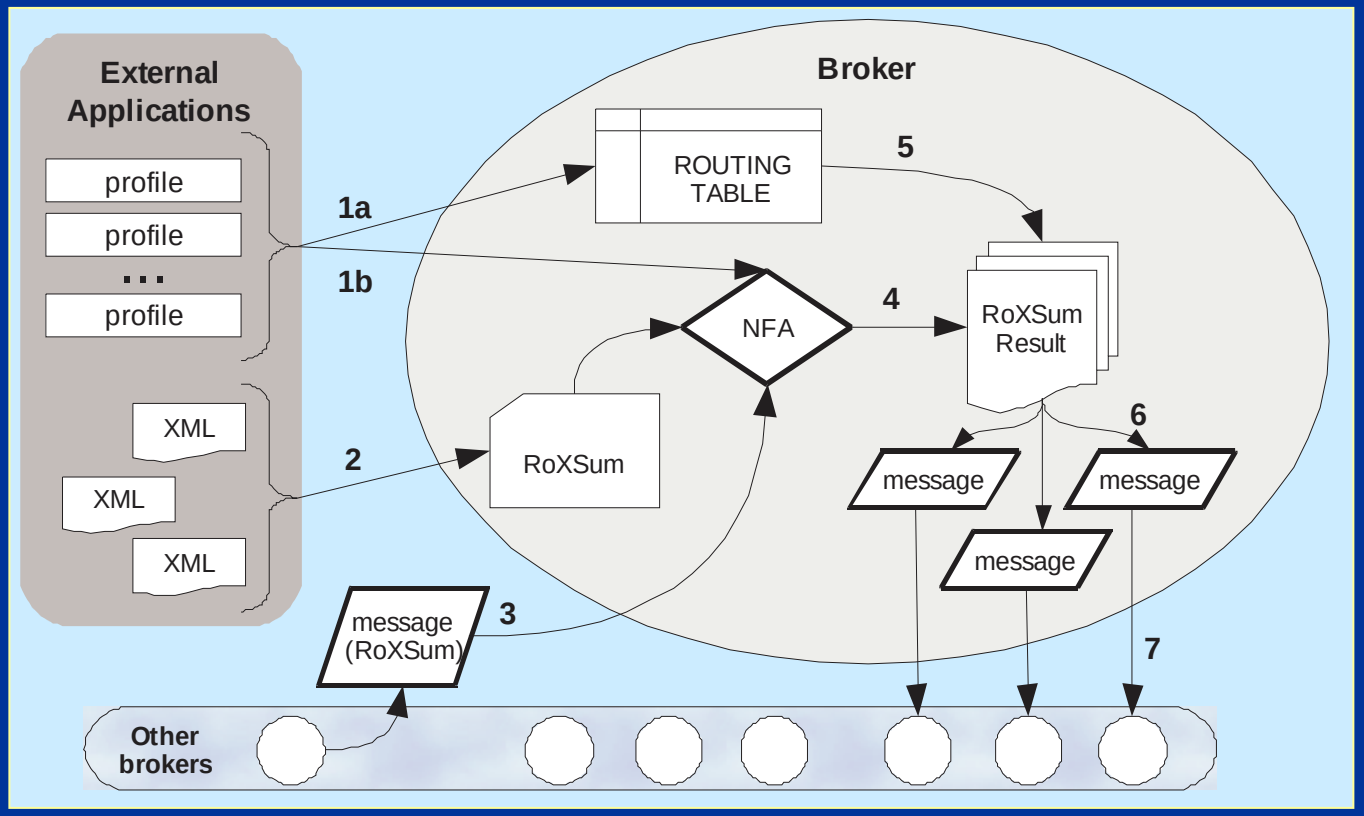
RoXSum Data Structure

RoXSum Processing

- XML content-based routing : profile matching on messages



Process Overview



Receiving Module

1. Brokers receive profiles:
   a. Create routing table
   b. Create an NFA
2. Create RoXSum
3. External RoXSum processing → NFA

Filtering Module

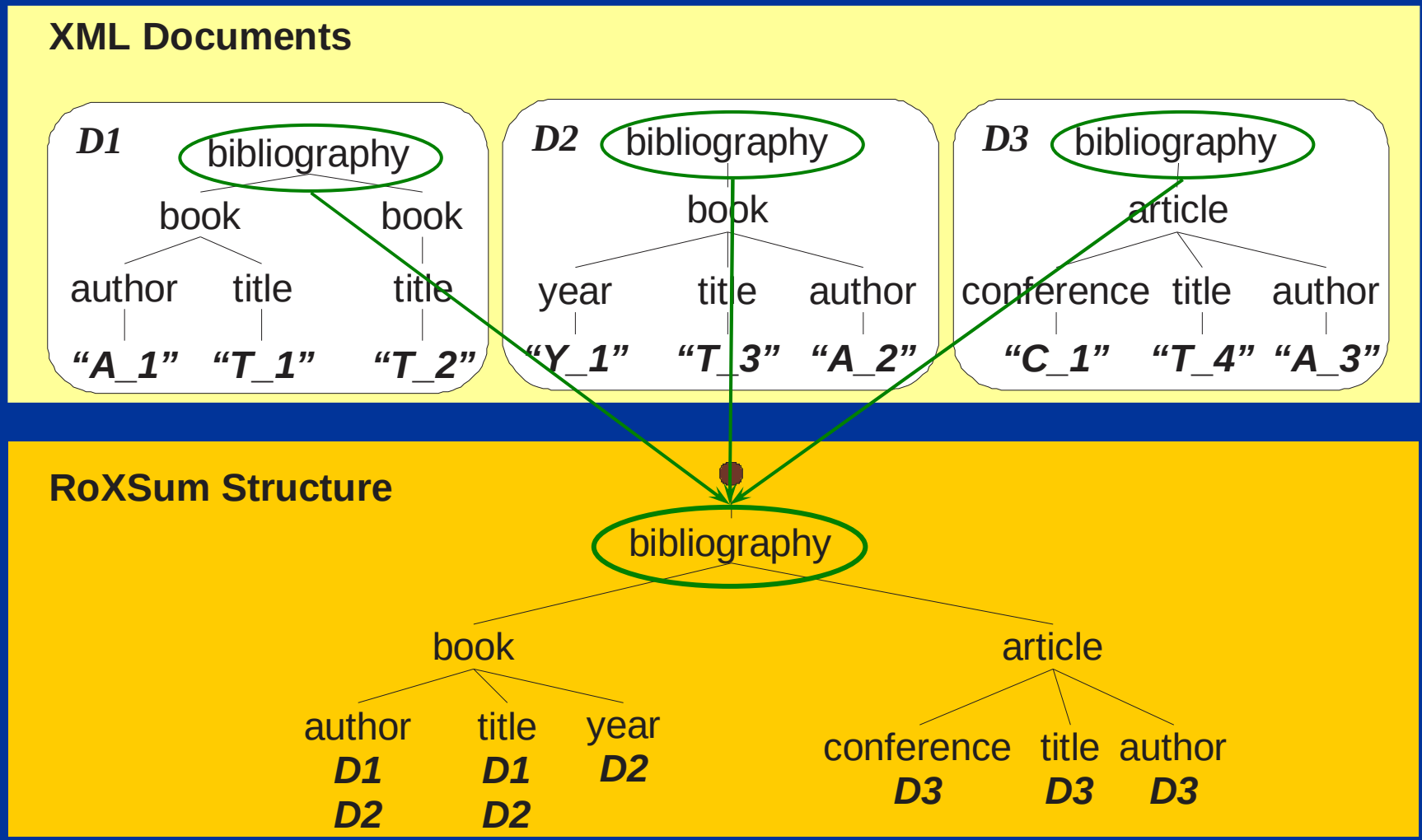
4. Match profiles:
   RoXSum → NFA

Forwarding Module

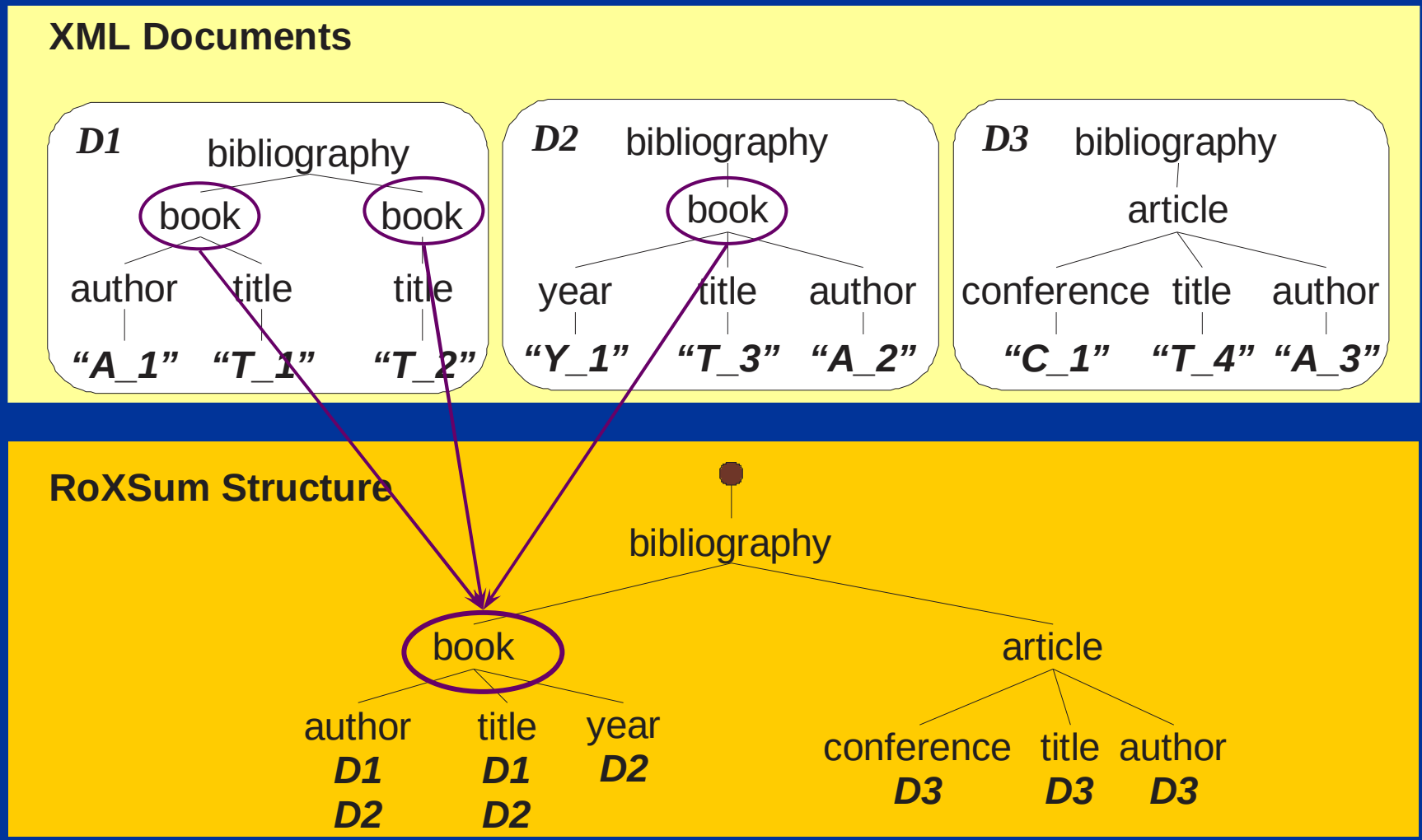
5. Decompose RoXSum
6. Create messages
7. Route

Value-Aware RoXSum

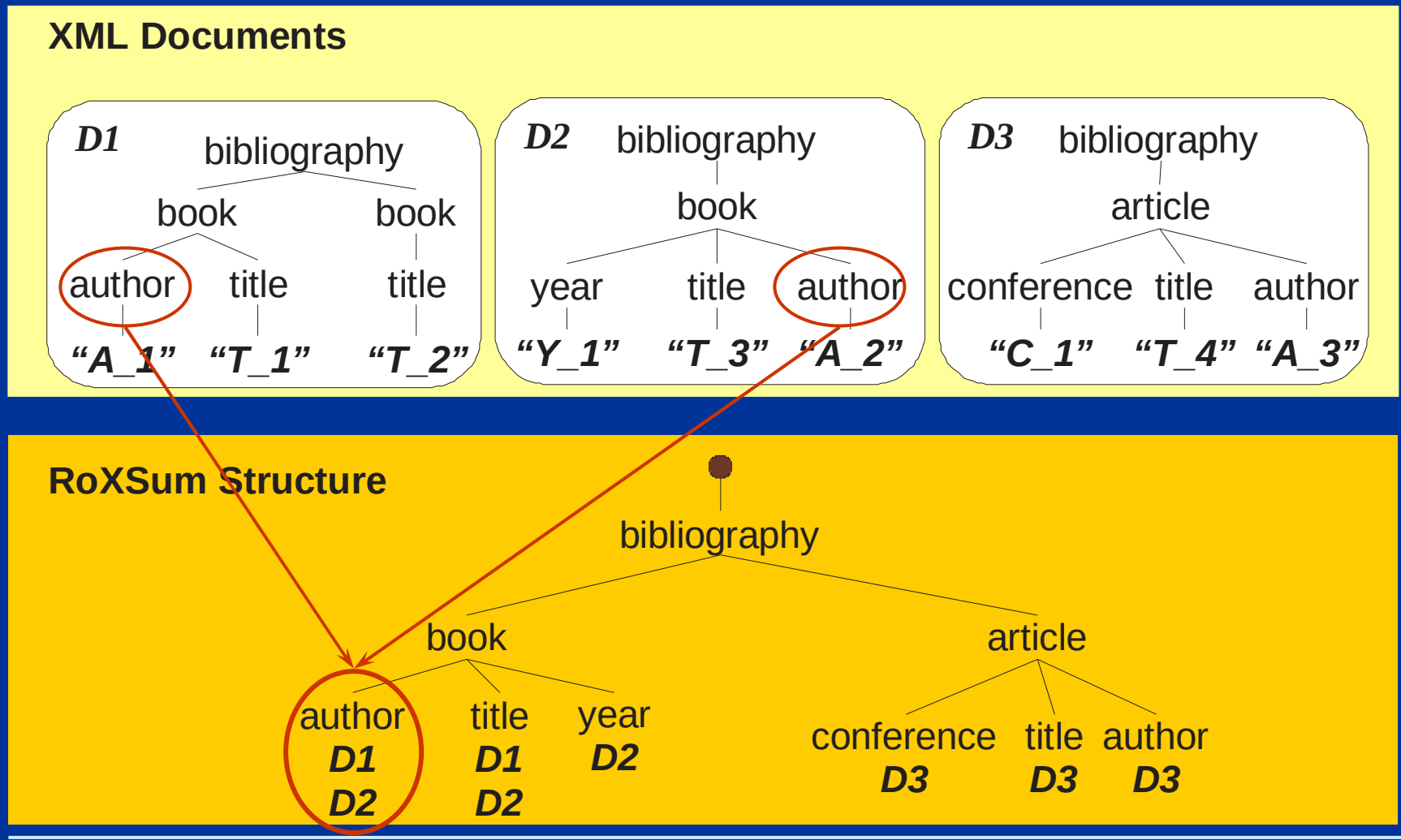
Documents and values



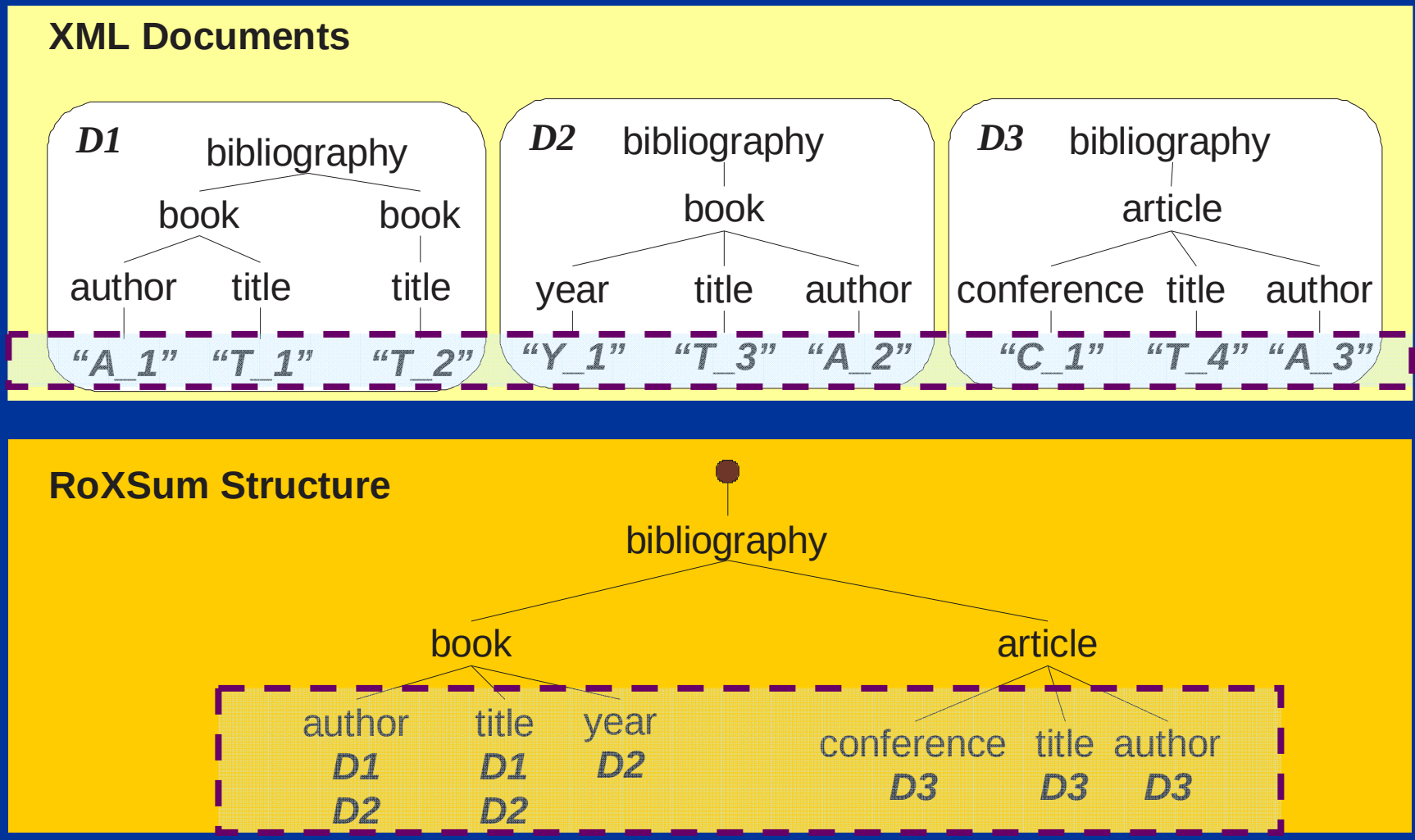
Documents and values



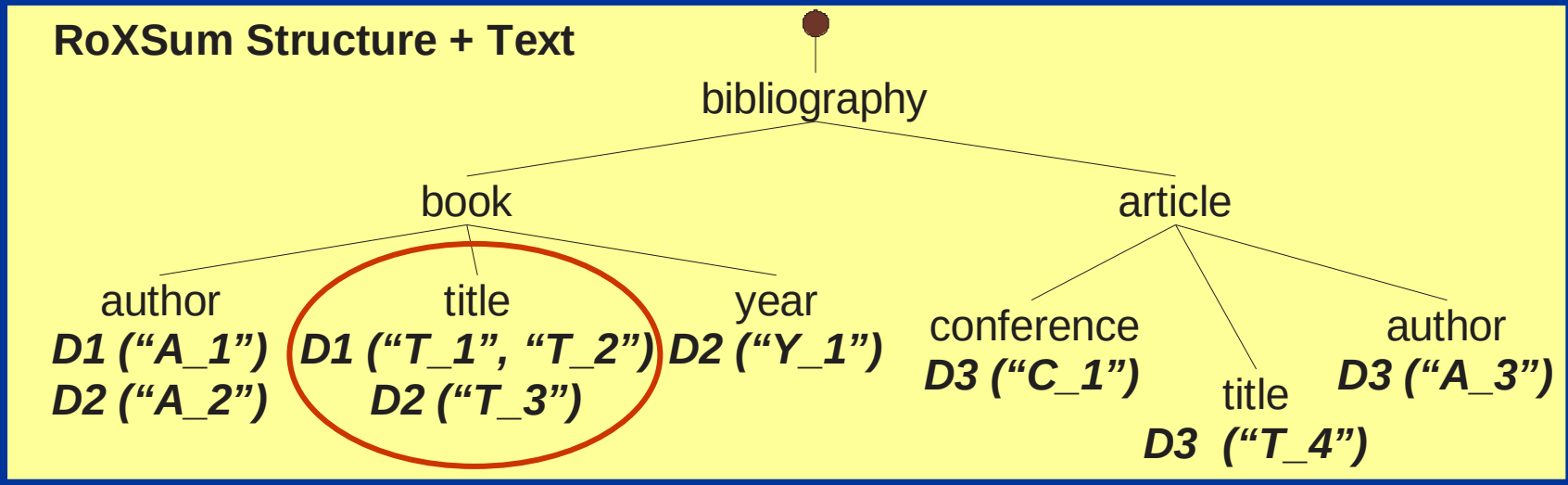
Documents and values



Documents and values

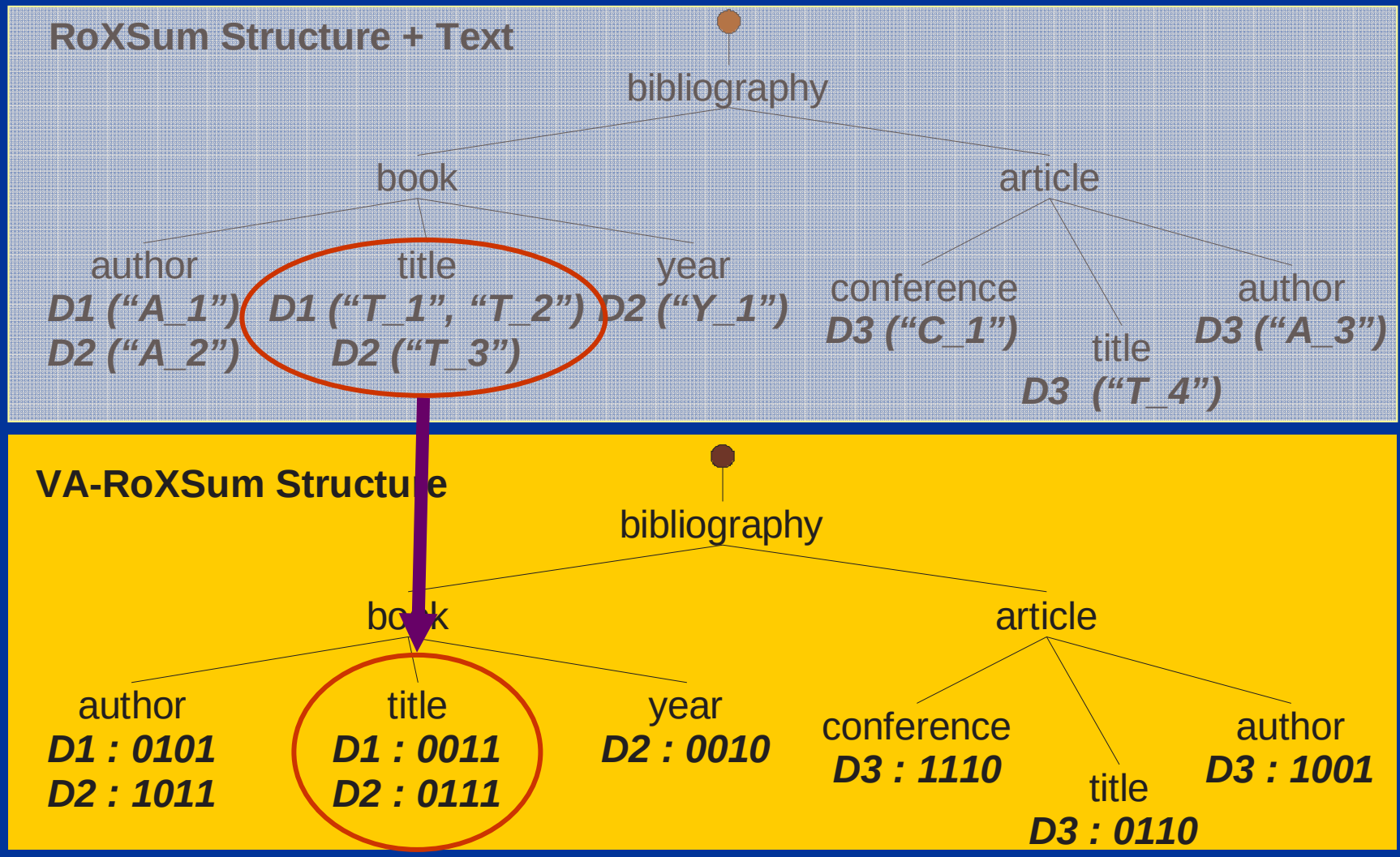


RoXSum + values

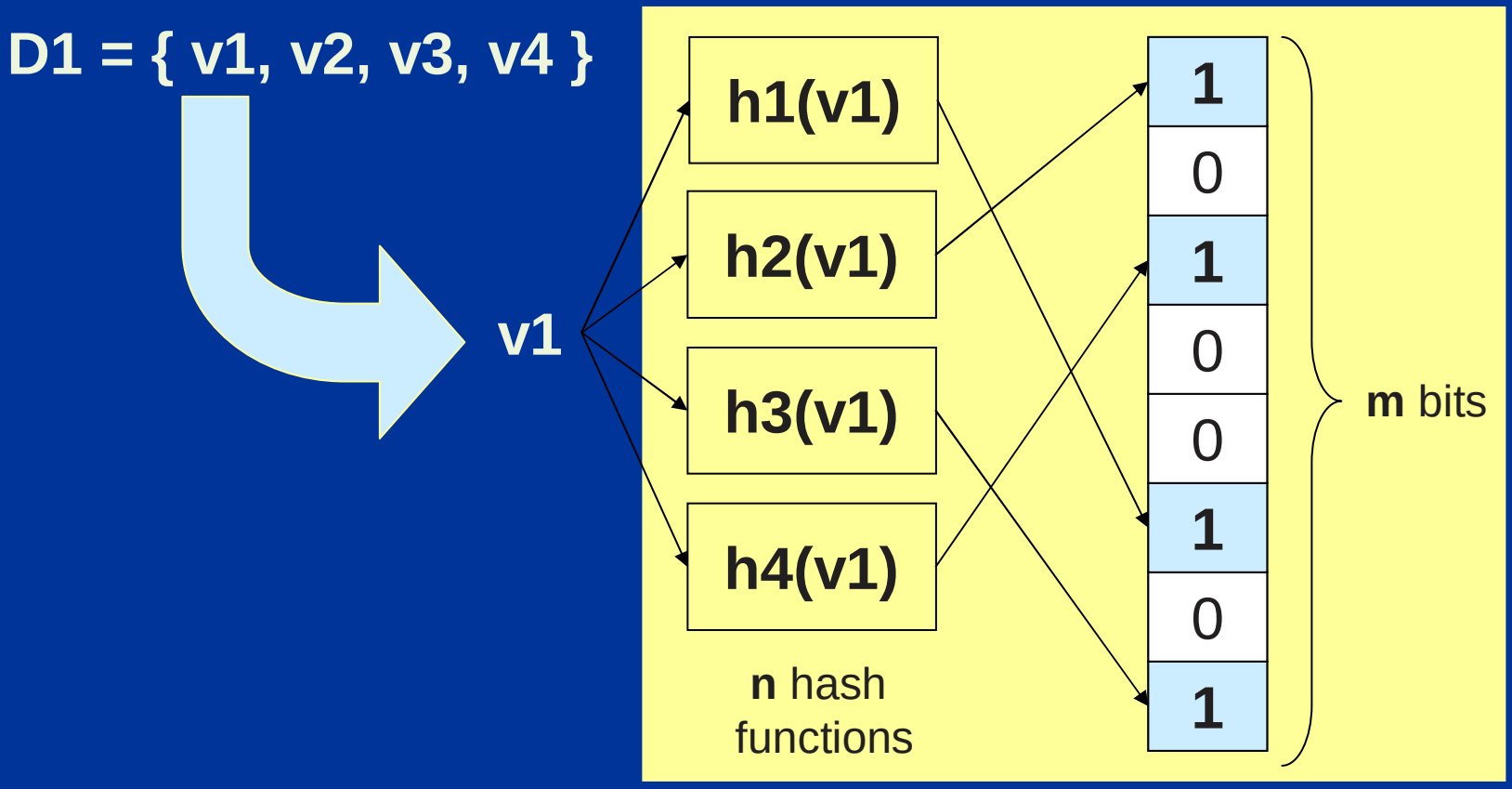


- All necessary information
- Space requirement
- No auxiliary structure → sequential scan

VA-RoXSum Structure



Bloom-Filters



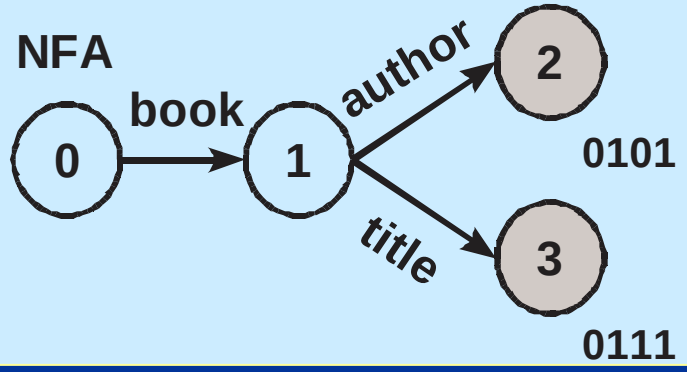
- Compact → Encode a set of values
- Efficient → Check whether a value belongs to set

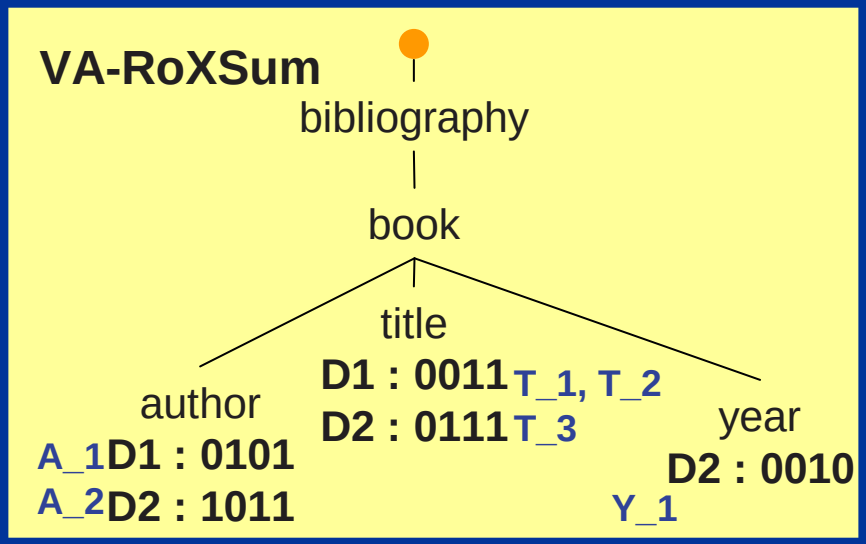
Profile Matching

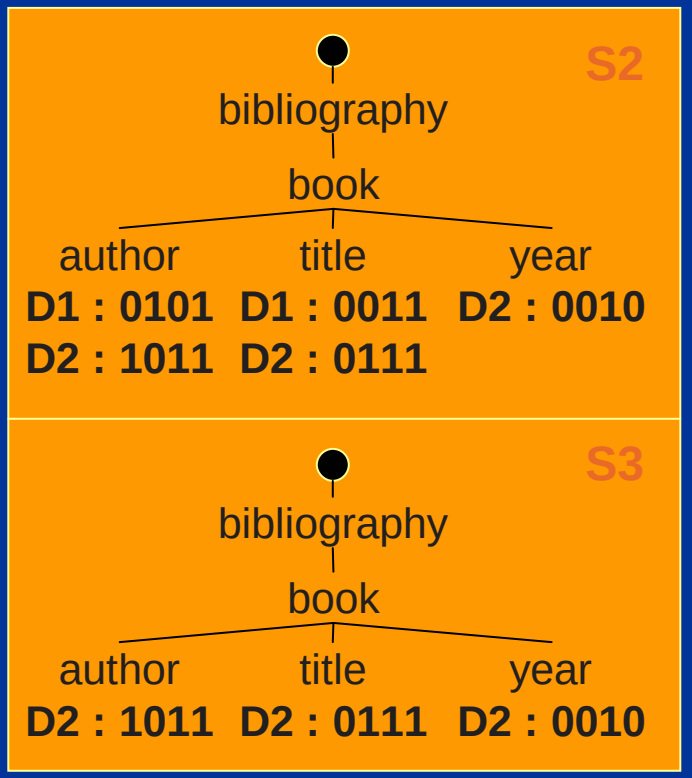
- VA-RoXSum construction and partition
- Profile matching
  - NFA states + value constraints
    (bloom filter signatures)
  - Traverse VA-RoXSum structure
  - Value constraint → signature containment

Routing Table

Profile	Target
//book[//author="A_1"]	B2
//book[//title="T_3"]	B2,B3







Results

Broker	Document
B2	D1, D2
B3	D2

Experimental Evaluation

Experimental Evaluation

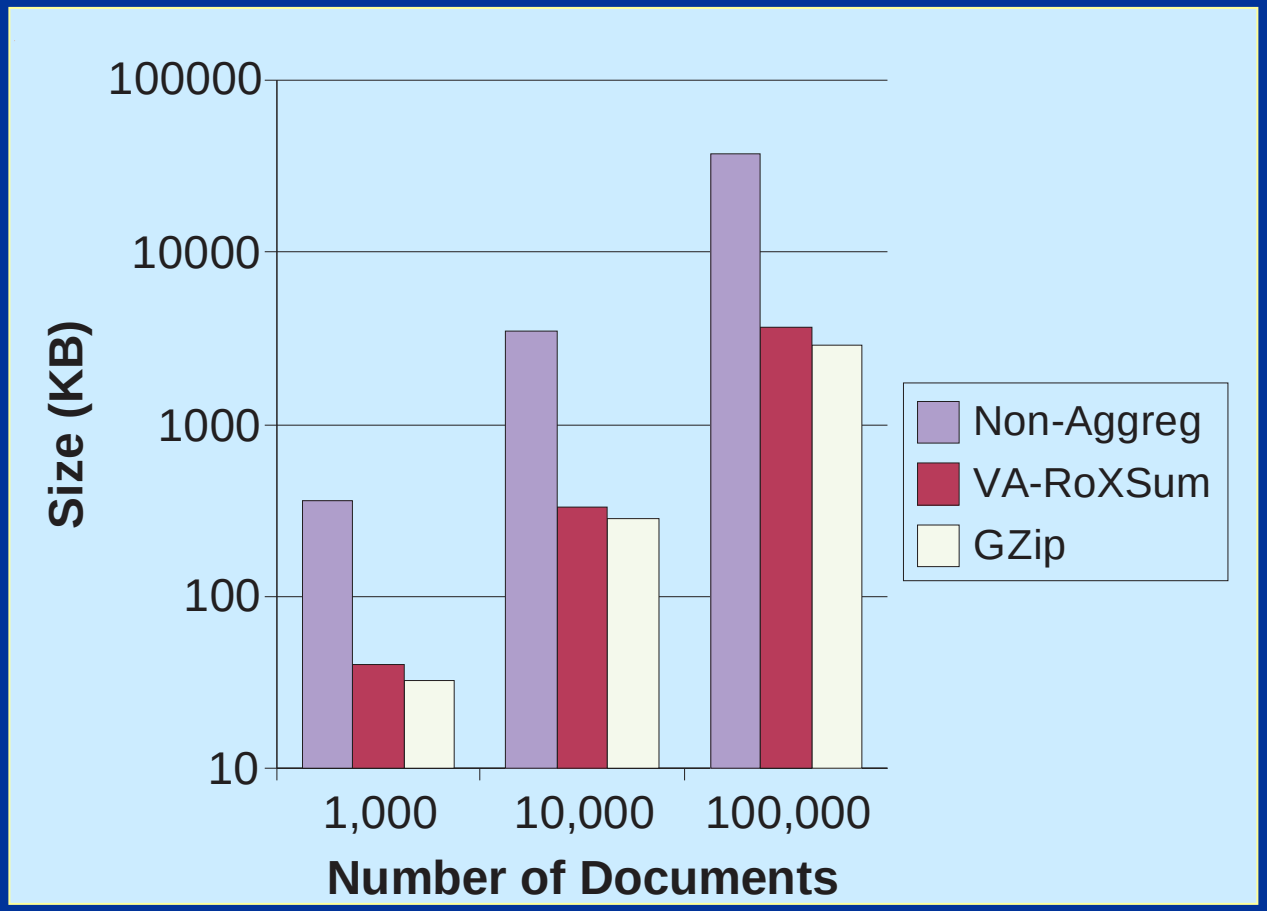
- Pub/Sub system + VA-RoXSum
- 100,000 path queries
- 100,000 small documents (25 DTDs)
- Evaluate
  - Number of false positives (bloom filter parameters)
  - Message size
  - Query evaluation processing time
- Against state-of-the-art stream processor: non-aggregated solution

Number of False Positives

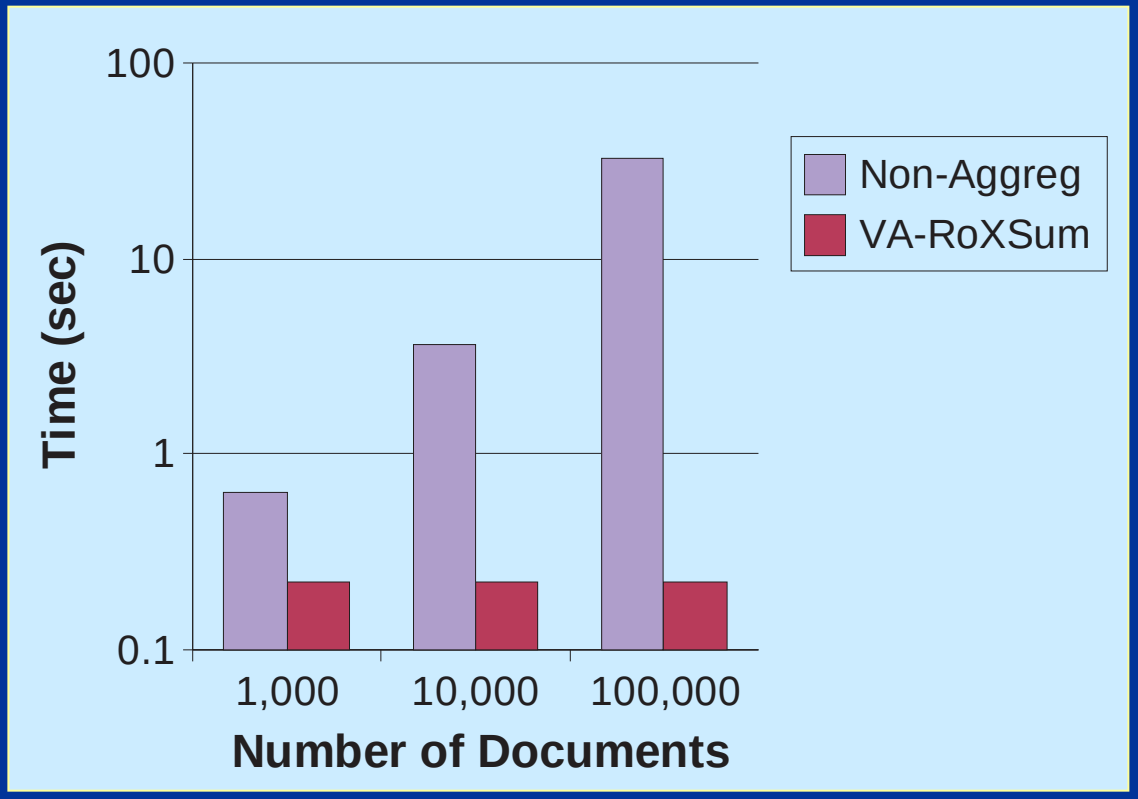
Hash Functions \ Bytes	2	3	4	5	6
2	8	6	4	3	0
3	4	1	0	0	0
4	15	1	6	0	0
5	7	0	0	0	0
6	0	0	0	0	0

100,000 messages (XML documents) + 100,000 profiles (XML path queries)
3 bytes + 3 hash functions → insignificant number of false positives

Message Size



Processing Time



100,000 queries
VA-RoXSum: much smaller size
use of bitwise operators to match value predicates

FINAL REMARKS

Summary

- Matching of messages and profiles within XML-aware pub/sub systems
- VA-RoXSum
  - Aggregate structure + values
  - Profile matching on aggregated content
  - NFA + bloom filters
- Performance
  - Improvement: two orders of magnitude

Value-Aware RoXSum:
Effective Message Aggregation for
XML-Aware Information Dissemination

?? Questions ??

Contact: mirella@inf.ufrgs.br