

Xcerpt and visXcerpt: **Integrating Web Querying**

Ludwig_____
Maximilians—
Universität_____
München_____

LMU

Sacha Berger
François Bry
Tim Furge
Benedikt Linse
Andreas Schroeder

Institute for Informatics
University of Munich

<http://www.pms.ifi.lmu.de/>

1

Data: Semi-structured Trees & Graphs

Graph data model for Xcerpt and visXcerpt

- as in RDF and semi-structured DBs like Lore
- great attention to **XML specificities** such as attributes and namespaces

Consistent Extension of XML

- children order may be irrelevant
- possible transparent resolution of **non-hierarchical relations**

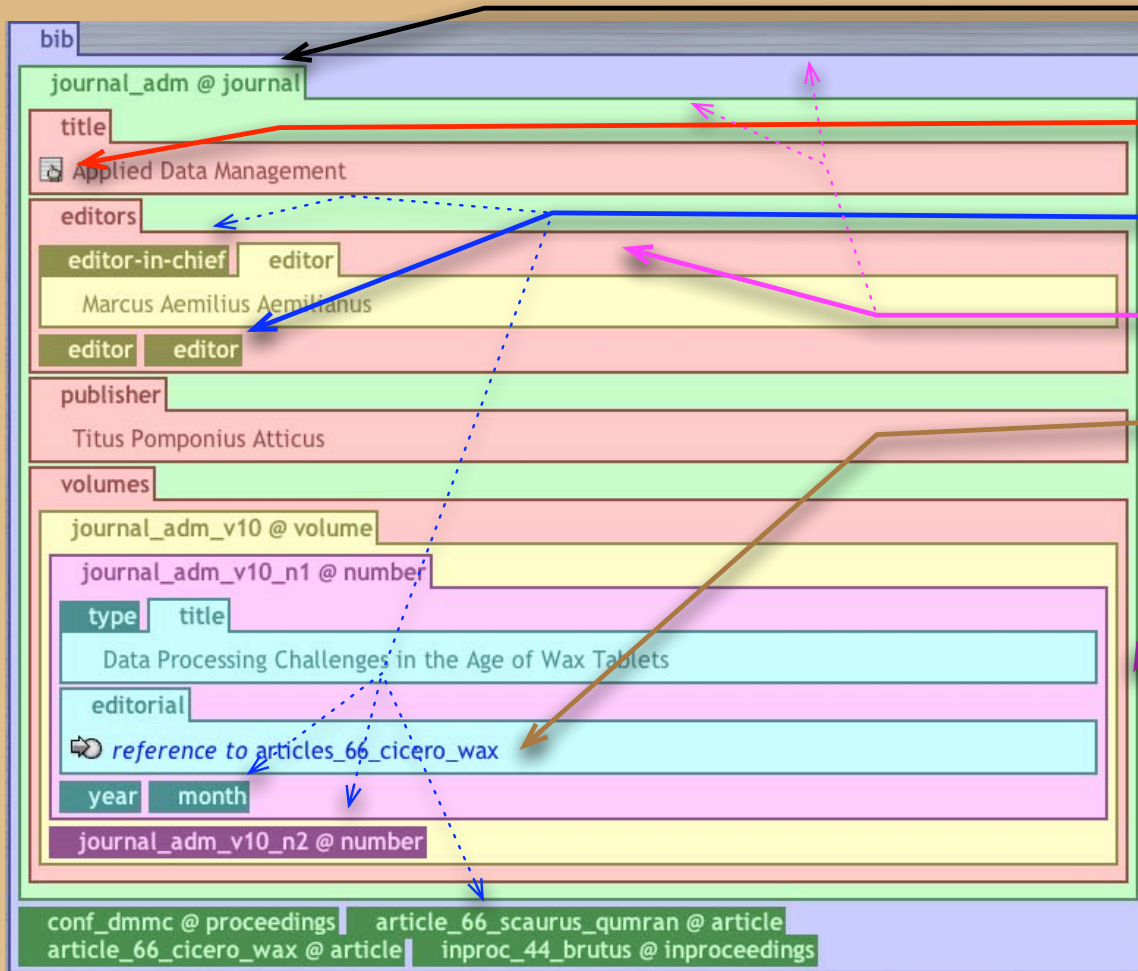
Overview

Data

Patterns

Rules

Bibliography Entries: DBLP-style



journal_adm @ journal

Identifier and label of elements



Context-Menu: Interactive Features

editor editor

Folding elements for information focus

Ordered vs. unordered children list

Non-hierarchical relations as hyperlinks



reference to articles_66_cicero_wax



Element nesting (child relation) becomes box nesting and colors

Bibliography Entries

- rather **regular** schema with **optionals**
- several **ordered** lists, otherwise keyed attributes

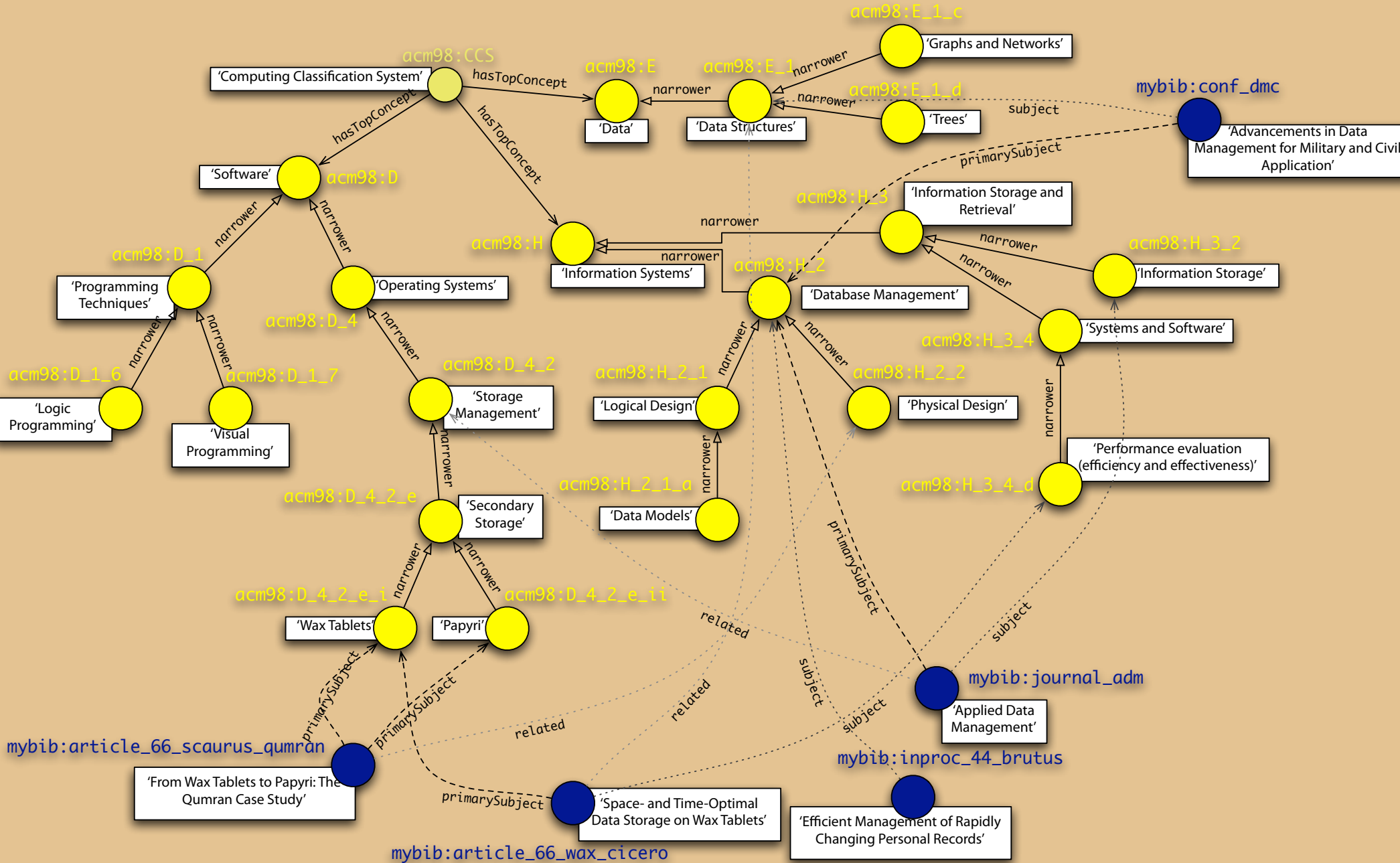
Overview

Data

Patterns

Rules

Topics and Themes: SKOS Ontology



2

Patterns: Examples for Selected Data

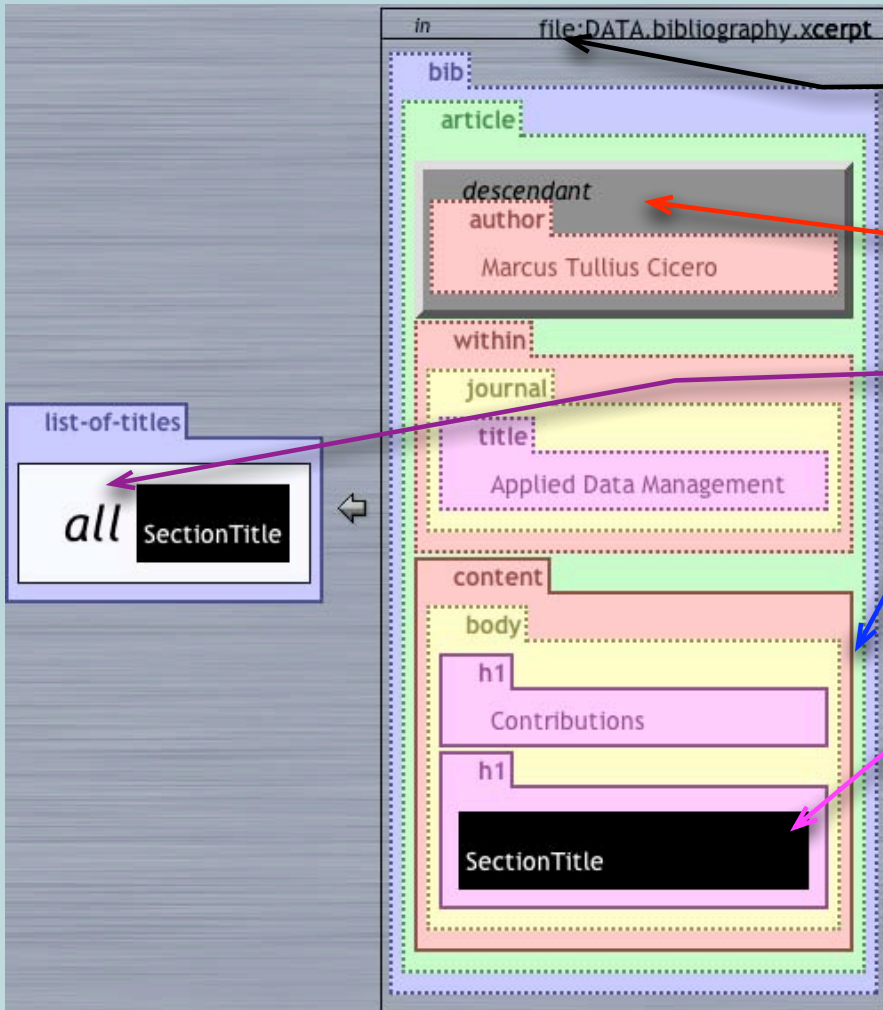
Query-by-Example paradigm

- queries just like data plus **variables**, **incompleteness**, **optionality**, **negation**
- patterns plus variables instead of navigation

Logical Variables in Patterns

- select relevant data (***n*-ary queries**)
- **group** and **aggregate** data
- **join** different data items

Basic Patterns: Variables and Incompleteness



in file:DATA.bibliography.xcerpt

Accessing Web resources: arbitrary XML documents can be accessed using their URL

descendant author

Incomplete patterns in depth:
descendant allows additional intermediary elements

all SectionTitle

Grouping collects alternative bindings for variables: essential for structural assembly

SectionTitle

Incomplete patterns in breadth:
partial patterns allow additional child elements

SectionTitle

Variables are used in lieu of data: express selection, joins, or arithmetic conditions

Basic Pattern

"return the titles of all top-level sections in articles by Marcus Tullius Cicero and published in 'Applied Data Management'."

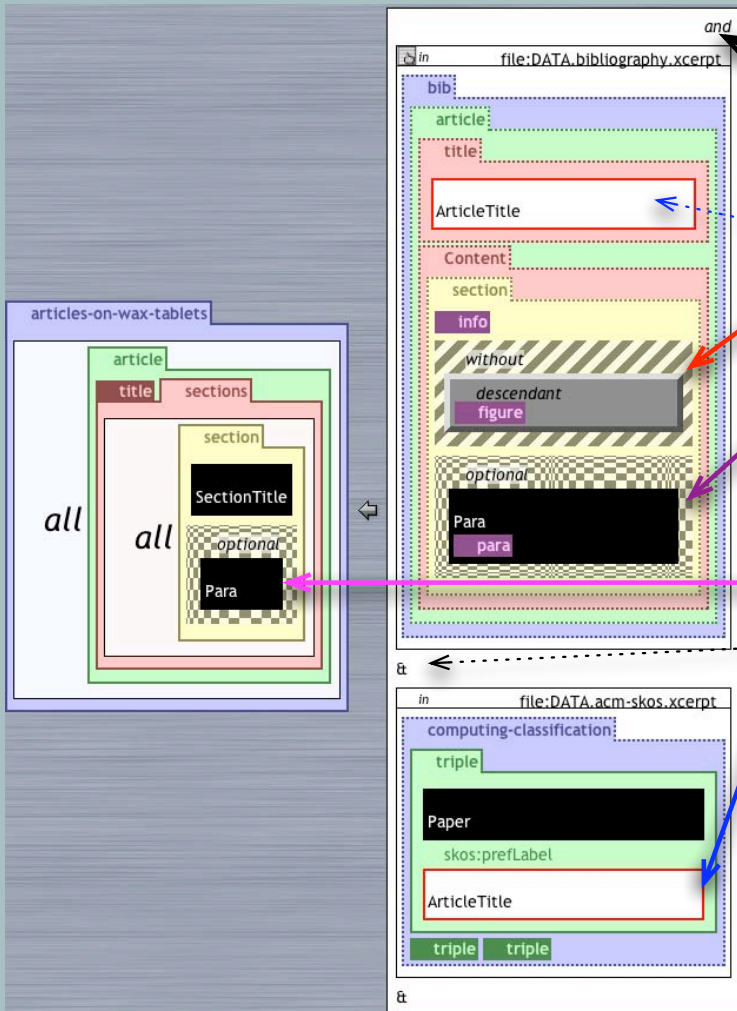
Overview

Data

Patterns

Rules

Complex Patterns: Formulas, Join, Optionality



and &

Terms as **formulas**:
Terms may contain **boolean** connectives, variables, **negation**, etc.



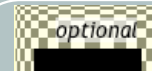
Subterm negation:
Some subterms may be required not to occur in matching data



Optional subterms:
Local form of disjunction essential for variable schema data



Value Joins:
Expressed through multiple variable occurrences



Optional construction:
Limited form of conditional construction based on variable bindings

Complex Pattern

"return titles and optionally paragraphs of all top-level sections without figures in articles on the topic 'Wax Tablets.'"

Overview

Data

Patterns

Rules

3

Rules: Separation of Concern by Views

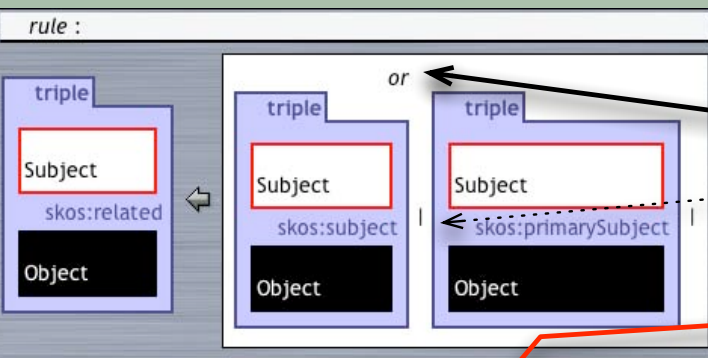
Separation of Query and Construction

- two separate parts in rules
- no mixing of construction and querying
 - instead chaining where necessary

Separation of Concern by Views

- separate **tasks** of a query in rules
- efficient evaluation of chained queries
 - memoization and unfolding

Rules: Inference, Views, and Chaining

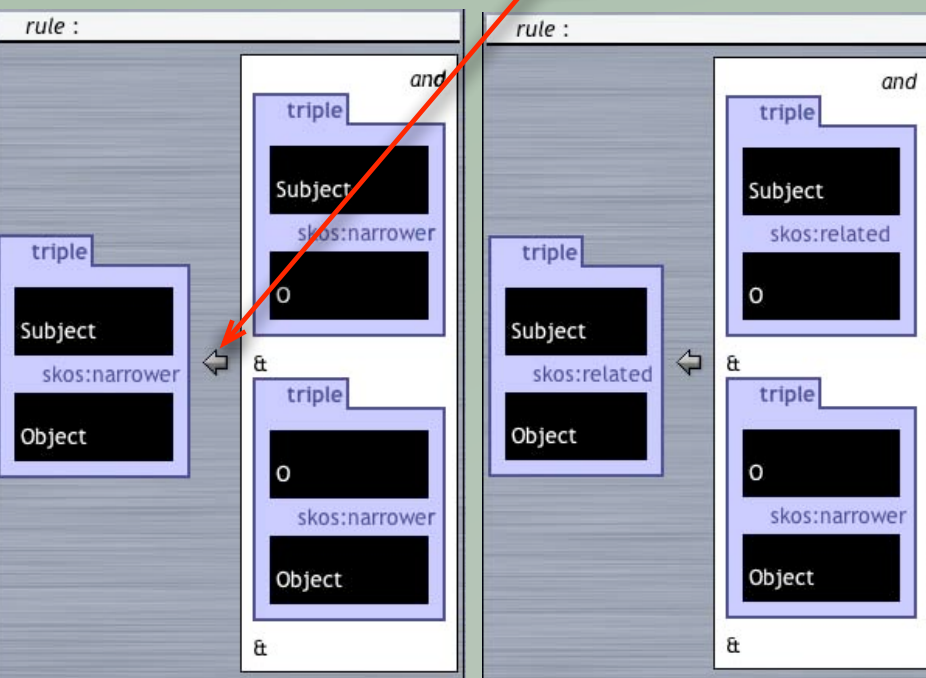


or |

Terms as **formulas**:
Terms may contain boolean connectives, including disjunctions



Rules separate construction from querying
and allow for procedural abstraction in query programs



Rules and Chaining

“close the skos:related relation on the provided data by adding skos:subject and traversing the closure of skos:narrower”

Overview

Data

Patterns

Rules