PatManQL: A language to manipulate patterns and data in hierarchical catalogs

Panagiotis Bouros, Theodore Dalamagas, Timos Sellis, Manolis Terrovitis

Knowledge and Database Systems Lab
School of Electrical and Computer Engineering
National Technical University of Athens
{pbour,dalamag,timos,mter}@dblab.ece.ntua.gr
Outline

- Introduction
- Contribution
- Structures
- Operators
- Prototype
- Related work
- Conclusion
Introduction

- Huge volumes of data on the Web
- Hierarchical structures and catalogs
- Paths $\rightarrow$ **knowledge artifacts**
  - Represent group of data
    - Conceptual clustering of raw data based on common properties
  - Semantic guides
- Example: **Portal catalogs**
Introduction

- Paths → **alternative pattern versions** for the same group of data
- Example: searching for lenses
  - `/cameras & lenses/lenses (adorama)`
  - `/photo/35mm systems/lenses (B&H)`

---

<table>
<thead>
<tr>
<th>brand</th>
<th>model</th>
<th>price</th>
</tr>
</thead>
<tbody>
<tr>
<td>Canon</td>
<td>EOS-3</td>
<td>990</td>
</tr>
<tr>
<td>Nikon</td>
<td>N65</td>
<td>300</td>
</tr>
<tr>
<td>Pentax</td>
<td>ZX-M</td>
<td>350</td>
</tr>
</tbody>
</table>

---

<table>
<thead>
<tr>
<th>brand</th>
<th>focald</th>
<th>cam_model</th>
<th>price</th>
</tr>
</thead>
<tbody>
<tr>
<td>Canon</td>
<td>50</td>
<td>EOS-3</td>
<td>400</td>
</tr>
<tr>
<td>Canon</td>
<td>80</td>
<td>EOS-3</td>
<td>450</td>
</tr>
<tr>
<td>Sigma</td>
<td>28</td>
<td>N65</td>
<td>150</td>
</tr>
</tbody>
</table>

---

PathMANQL
Introduction

- Paths → **complex pattern**
- Example: searching for integrated photo systems
  - /cameras & lenses/35mm SLR (adorama)
  - /photo/35mm systems/lenses (B&H)
Contribution

- A model to represent paths as knowledge artifacts
- The **PatManQL** language:
  - Operators to manipulate path-like patterns
  - Relational operators for data
- A prototype
Catalog Schema

- A tree with:
  - a root ($\boxtimes$)
  - a set of non-leaf nodes ($\mathcal{O}$)
  - a set of resource items as leaves ($\square$)

- Data: instances (records) of resource item

  - Resource item: Relation $R(a_1, a_2, \ldots, a_n)$, where $a_1, a_2, \ldots$ attributes
Catalog Schema

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Hierarchy

Resource items

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Data

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Catalog Schema

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Brand</th>
<th>Model</th>
<th>Price</th>
</tr>
</thead>
<tbody>
<tr>
<td>Canon</td>
<td>EOS-3</td>
<td>990</td>
</tr>
<tr>
<td>Nikon</td>
<td>N65</td>
<td>205</td>
</tr>
<tr>
<td>Pentax</td>
<td>ZX-M</td>
<td>148.50</td>
</tr>
</tbody>
</table>
Tree-Structure Relations (TSRs)

- Combining catalog schemas with common resource item
- Tree-Structure Relation (AND/OR-like graph):
  - **One** resource item
  - Paths organized in **OR components**
    - OR component: group of one or more paths (AND group)
    - OR components are alternative ways to access the common resource item
  - **Paths = patterns**
Tree-Structure Relations (TSRs)

(a) camera & lenses
   └─── cameras
      └─── 35mm SLR
          └─── SLR cameras

(b) photo
    └─── 35mm systems
        └─── SLR systems
            └─── lens_id
                └─── brand
                    └─── model
                        └─── price
                            └─── lens_id
Operators

Select ($\sigma$)

- $\sigma<$attribute condition$><$path condition$>$ (TSR)
  - attribute condition: {$=$, $\neq$, $<$}
  - path condition: {$=$, $\neq$, $\subset$, $\angle$}

- Filters instances of resource items and OR components
Select example

'Select all non Pentax cameras with price greater than 200 Euros, having "/photo/35mm systems" in their paths':

$\sigma<brand \neq \text{"Pentax"}, \text{price} > 200><\text{"/photo/35mm systems" \subset \_}> (\text{SLR systems})$
Operators

- **Project** ($\pi$)
  - $\pi<$attribute list>$<$variable list$>$ (TSR)
  - Attribute list: \{attribute\}
  - Variable list: \{$i (path variable),
  #i (OR variable)$\}

- Keeps attributes of resource item and paths of each OR component or OR components on the whole
Project example

'Cameras with only the model and lens_id attributes and the rightmost component':

\[ \Pi_{\langle\text{model}, \text{lens_id}\rangle\langle\#2\rangle}(\text{SLR systems}) \]

(a) SLR systems

<table>
<thead>
<tr>
<th>brand</th>
<th>model</th>
<th>price</th>
<th>lens_id</th>
</tr>
</thead>
<tbody>
<tr>
<td>Canon</td>
<td>EOS-3</td>
<td>990</td>
<td>1</td>
</tr>
<tr>
<td>Nikon</td>
<td>N65</td>
<td>205</td>
<td>2</td>
</tr>
<tr>
<td>Pentax</td>
<td>ZX-M</td>
<td>148.5</td>
<td>2</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>

(b) SLR systems

<table>
<thead>
<tr>
<th>model</th>
<th>lens_id</th>
</tr>
</thead>
<tbody>
<tr>
<td>EOS-3</td>
<td>1</td>
</tr>
<tr>
<td>N65</td>
<td>2</td>
</tr>
<tr>
<td>ZX-M</td>
<td>2</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>
Operators

- Cartesian product \((X)\)
  - \((\text{TSR1}) \times (\text{TSR2})\)
  - Combine instances of resources and OR components
Cartesian product example

(SLR systems) \times (Lenses)
Operators

- **Union** (U)
  - (TSR) U (TSR)
  - Union of instances and all OR components

- **Intersection** (∩)
  - (TSR) ∩ (TSR)
  - Intersection of instances and all OR components

- **Difference** (–)
  - (TSR) – (TSR)
  - Instances of the first TSR not present in the second one and all OR components of the first TSR
Union example

\[(\text{SLR systems}) \cup (\text{SLR systems})\]

(a) SLR systems

<table>
<thead>
<tr>
<th>cbrand</th>
<th>cmodel</th>
<th>cprice</th>
<th>clenid</th>
</tr>
</thead>
<tbody>
<tr>
<td>Canon</td>
<td>EOS-3</td>
<td>990</td>
<td>1</td>
</tr>
<tr>
<td>Nikon</td>
<td>N65</td>
<td>205</td>
<td>1</td>
</tr>
<tr>
<td>Pentax</td>
<td>ZX-M</td>
<td>148.5</td>
<td>2</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>

(b) SLR systems

<table>
<thead>
<tr>
<th>cbrand</th>
<th>cmodel</th>
<th>cprice</th>
<th>clenid</th>
</tr>
</thead>
<tbody>
<tr>
<td>Canon</td>
<td>EOS-3</td>
<td>990</td>
<td>1</td>
</tr>
<tr>
<td>Nikon</td>
<td>FM2</td>
<td>800</td>
<td>1</td>
</tr>
<tr>
<td>Pentax</td>
<td>ZX-M</td>
<td>148.5</td>
<td>2</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>

(c) SLR systems

<table>
<thead>
<tr>
<th>cbrand</th>
<th>cmodel</th>
<th>cprice</th>
<th>clenid</th>
</tr>
</thead>
<tbody>
<tr>
<td>Canon</td>
<td>EOS-3</td>
<td>990</td>
<td>1</td>
</tr>
<tr>
<td>Nikon</td>
<td>N65</td>
<td>205</td>
<td>1</td>
</tr>
<tr>
<td>Pentax</td>
<td>ZX-M</td>
<td>148.5</td>
<td>2</td>
</tr>
<tr>
<td>Nikon</td>
<td>FM2</td>
<td>800</td>
<td>1</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>
Prototype

- Interpreter
- Query Execution Engine
- Storage mechanism
  - XML files
  - MySQL RDBMS
    - All-edges-in-one-table storage approach
- Graphical Interface
Related work

- Pattern management (PANDA project) (S. Rizzi et al.)
- Inductive databases framework (Tomasz Imielinski et al.)
  - DMQL (Jiawei Han et al.), MINE RULE (R. Meo et al.)
    - Descriptive rules
- Tree algebras
  - TAX (H. V. Jagadish et al.)
    - Selecting – reconstructing bulk XML data
  - YAT (V. Christophides et al.)
    - Tuple-based, not tree-based
Conclusion

- A model to represent paths as knowledge artifacts (patterns)
  - Catalog schema
  - Tree-Structure Relations (TSRs)
- The PatManQL language:
  - Operators to manipulate paths as patterns and data
- A prototype system
Future Work

- Properties of the Operators
- Restructure operators
- Join operator
Questions (?)
Tree-Structure Relations (TSRs)

(a) SLR cameras
- brand
- model
- price

(b) SLR systems
- brand
- model
- price
- lens_id

X $1

$1

35mm SLR

35mm systems

photo

cameras

camera & lenses

$1

$2

35mm systems

bodies

lenses

35mm SLR

photo

$1

$1

$1
Storage mechanism

**XML file**

```xml
<tsr name="SLR systems">
  <or>
    <and>/photo/35mm SLR/bodies</and>
    <and>/photo/lenses</and>
  </or>
  <or>
    <and>/photo/35mm systems</and>
  </or>
  <item>
    <attribute name="brand" type="...">
    <attribute name="model" type="...">
      ...
      <tuple>...</tuple>
      ...
    </item>
  </tsr>
```

---

PatManQL 25
Storage mechanism

- Database

<table>
<thead>
<tr>
<th>tid</th>
<th>name</th>
<th>file</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>SLR systems</td>
<td>portal.xml</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>brand</th>
<th>model</th>
<th>price</th>
<th>lens_id</th>
</tr>
</thead>
<tbody>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>tid</th>
<th>orid</th>
<th>andid</th>
<th>path</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>/photo/35mm SLR/bodies</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>2</td>
<td>/photo/lenses</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>1</td>
<td>/photo/35mm systems</td>
</tr>
</tbody>
</table>
Catalog Schemas examples

(a)

<table>
<thead>
<tr>
<th>brand</th>
<th>model</th>
<th>price</th>
</tr>
</thead>
<tbody>
<tr>
<td>Canon</td>
<td>EOS-3</td>
<td>990</td>
</tr>
<tr>
<td>Nikon</td>
<td>N65</td>
<td>300</td>
</tr>
<tr>
<td>Pentax</td>
<td>ZX-M</td>
<td>350</td>
</tr>
</tbody>
</table>

(b)

<table>
<thead>
<tr>
<th>brand</th>
<th>focald</th>
<th>cam_model</th>
<th>price</th>
</tr>
</thead>
<tbody>
<tr>
<td>Canon</td>
<td>50</td>
<td>EOS-3</td>
<td>400</td>
</tr>
<tr>
<td>Canon</td>
<td>80</td>
<td>EOS-3</td>
<td>450</td>
</tr>
<tr>
<td>Sigma</td>
<td>28</td>
<td>N65</td>
<td>150</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>
Catalog Schema Manipulation

- SLR integrated systems from X – fig. (a)
- SLR cameras from Adorama – fig. (b)
- Lenses from B&H – fig. (c)

Scenario for X:
- New lenses out in the market
- Lenses provided by B&H, that fit in Canon bodies provided by Adorama
- Above SLR systems not present in her stock
Catalog Schema Manipulation

(a) SLR systems

- photo
- 35mm systems

(b) SLR cameras

- camera & lenses
- 35mm SLR

(c) Lenses

- photo
- 35mm systems
- lenses

### SLR systems

<table>
<thead>
<tr>
<th>cbrand</th>
<th>cmodel</th>
<th>cprice</th>
<th>lmodel</th>
<th>lprice</th>
</tr>
</thead>
<tbody>
<tr>
<td>Canon</td>
<td>EOS-3</td>
<td>990</td>
<td>100</td>
<td>390</td>
</tr>
<tr>
<td>Nikon</td>
<td>N65</td>
<td>205</td>
<td>340</td>
<td>160</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>

### SLR cameras

<table>
<thead>
<tr>
<th>cbrand</th>
<th>cmodel</th>
<th>cprice</th>
</tr>
</thead>
<tbody>
<tr>
<td>Canon</td>
<td>EOS-3</td>
<td>990</td>
</tr>
<tr>
<td>Nikon</td>
<td>N65</td>
<td>300</td>
</tr>
<tr>
<td>Pentax</td>
<td>ZX-M</td>
<td>350</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>

### Lenses

<table>
<thead>
<tr>
<th>lbrand</th>
<th>lmodel</th>
<th>focald</th>
<th>cam_model</th>
<th>lprice</th>
</tr>
</thead>
<tbody>
<tr>
<td>Canon</td>
<td>100</td>
<td>50</td>
<td>EOS-3</td>
<td>400</td>
</tr>
<tr>
<td>Canon</td>
<td>110</td>
<td>80</td>
<td>EOS-3</td>
<td>450</td>
</tr>
<tr>
<td>Sigma</td>
<td>340</td>
<td>28</td>
<td>N65</td>
<td>150</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>
Catalog Schema Manipulation

- Systems with Canon bodies from Adorama and lenses from B&H – fig. (d):
  - \[ q_1 = \pi_{<\text{cbrand},\text{cmodel},\text{lmodel}> < >} (\sigma_{<\text{cmodel}=\text{cam}_\text{model}, \text{cbrand}="\text{Canon}" < >} ((\text{SLR cameras}) \times (\text{lenses}))) \]

- Systems with Canon bodies from Adorama and lenses from B&H which are not in X's catalog – fig. (e):
  - \[ q_2 = (q_1) - \pi_{<\text{cbrand},\text{cmodel},\text{lmodel}> < >} (\text{SLR cameras}) \]

- Lenses only without the appropriate camera bodies – fig. (f):
  - \[ \pi_{<\text{lmodel}> < $2 >}<q_2> \]
Catalog Schema Manipulation

(d) camera & lenses
   / 
  /   \  
35mm systems / 
   /     \  
35mm SLR lenses

<table>
<thead>
<tr>
<th>cbrand</th>
<th>cmodel</th>
<th>lmodel</th>
</tr>
</thead>
<tbody>
<tr>
<td>Canon</td>
<td>EOS-3</td>
<td>100</td>
</tr>
<tr>
<td>Canon</td>
<td>EOS-3</td>
<td>110</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>

(e) camera & lenses
   / 
  /   \  
35mm systems / 
   /     \  
35mm SLR lenses

<table>
<thead>
<tr>
<th>cbrand</th>
<th>cmodel</th>
<th>lmodel</th>
</tr>
</thead>
<tbody>
<tr>
<td>Canon</td>
<td>EOS-3</td>
<td>110</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>

(f) 35mm systems
   / 
  /   \  
lenses

Lenses

<table>
<thead>
<tr>
<th>lmodel</th>
</tr>
</thead>
<tbody>
<tr>
<td>110</td>
</tr>
<tr>
<td>...</td>
</tr>
</tbody>
</table>
Prototype Architecture
XML File Manager (XFM)
Database Manager (DM)

- Database
  - OR components, paths and resource items retrieval
  - OR components, paths and resource items storage

- TSR
  - Interpreter
  - Graphic Result Interface
Query Execution Engine (QE)

Interpreter

TSR
Parameters list

Paths construction and OR components creation

TSR

Resource item construction Attributes and resords

Interpreter
Interpreter

Parser

Parameter collection

Query execution engine

Print error message

Error message

Query

Query parameters

Database manager

Graphic result interface

XML file manager

Database manager

TSR

PatManQL
Graphic Result Interface (GRI)