FAKULTÄT
FÜR !NFORMATIK
Faculty of Informatics

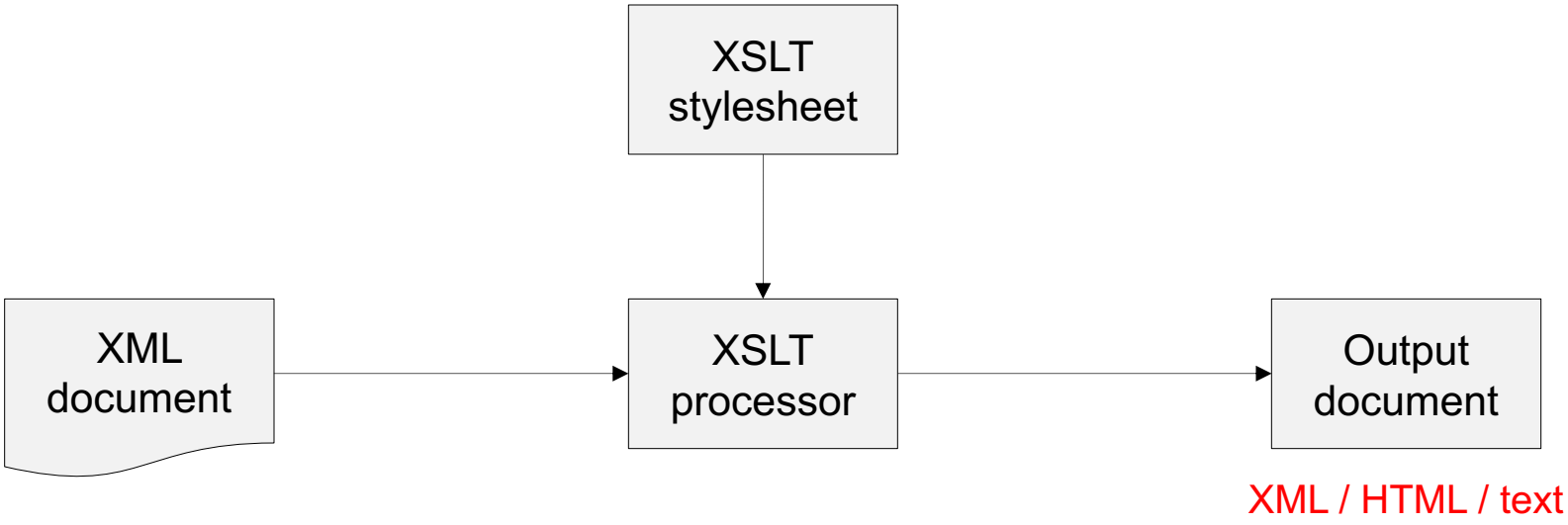
dbai

Semi-structured Data

7 – XSLT
Examples

Mantas Šimkus

How XSLT Works?



- Define a transformation with an XSLT document (which is an XML document)
- Apply this transformation on an input document using an XSLT processor

How to execute XSL Transformations

- Use the Saxon library
  - Download: http://www.dbai.tuwien.ac.at/education/ssd/SS14/saxon/saxon9he.jar
    http://saxon.sourceforge.net/
  - Add *saxon9he.jar* to the classpath
  - **java net.sf.saxon.Transform -s:example.xml -xsl:example.xsl -o:output.xml**
- Use an online tool
  - e.g. http://www.xpathtester.com/xslt
- Use a Webbrowser
  - Often only XSLT 1.0 / XPATH 1.0 support
- For the exercise use the provided Ant script

Linking an XML file to an XSL file

```
<?xml version="1.0"?>
<?xml-stylesheet type="text/xsl" href="simple.xsl"?>
<courses>
    <SSD  day="Thursday">
        Starts at
        09:15
    </SSD>
    <Databases day="Tuesday">
        Starts at
        09:45
    </Databases>
</courses>
```

A simple example

```
<courses>
    <SSD  day="Thursday">
        Starts at
        09:15
    </SSD>
    <Databases day="Tuesday">
        Starts at
        09:45
    </Databases>
</courses>
```

```
<?xml version="1.0" encoding="UTF-8"?>
```

Starts at

09:15

Starts at

09:45

```
<?xml version="1.0"?>
<xsl:stylesheet  version="2.0"
                xmlns:xsl="http://www.w3.org/1999/XSL/Transform">
</xsl:stylesheet>
```

xsl:output

```
<courses>
    <SSD  day="Thursday">
        Starts at
        09:15
    </SSD>
    <Databases day="Tuesday">
        Starts at
        09:45
    </Databases>
</courses>
```

Starts at

09:15

Starts at

09:45

```
<?xml version="1.0"?>
<xsl:stylesheet  version="2.0"
                xmlns:xsl="http://www.w3.org/1999/XSL/Transform">
    <xsl:output  method="html"/>
</xsl:stylesheet>
```

Priorities of Template rules

```
<?xml version="1.0"?>
<xsl:stylesheet  version="2.0"   xmlns:xsl="http://www.w3.org/1999/XSL/Transform">
  <xsl:output method="text"/>

  <xsl:template match="name">
    <xsl:value-of select="."/>
  </xsl:template>

  <xsl:template match="author/name">
    AUTHOR: <xsl:value-of select="."/>
  </xsl:template>
</xsl:stylesheet>
```

```
<books>
 <author>
   <name>Douglas Adams</name>
 </author>
 <book>
  <name>The Hitchhiker's
   Guide to the Galaxy</name>
 </book>
</books>
```

Priorities of Template rules

```
<?xml version="1.0"?>
<xsl:stylesheet  version="2.0"   xmlns:xsl="http://www.w3.org/1999/XSL/Transform">
  <xsl:output method="text"/>

  <xsl:template match="name">
    <xsl:value-of select="."/>
  </xsl:template>

  <xsl:template match="author/name">
    AUTHOR: <xsl:value-of select="."/>
  </xsl:template>
</xsl:stylesheet>
```

Output:

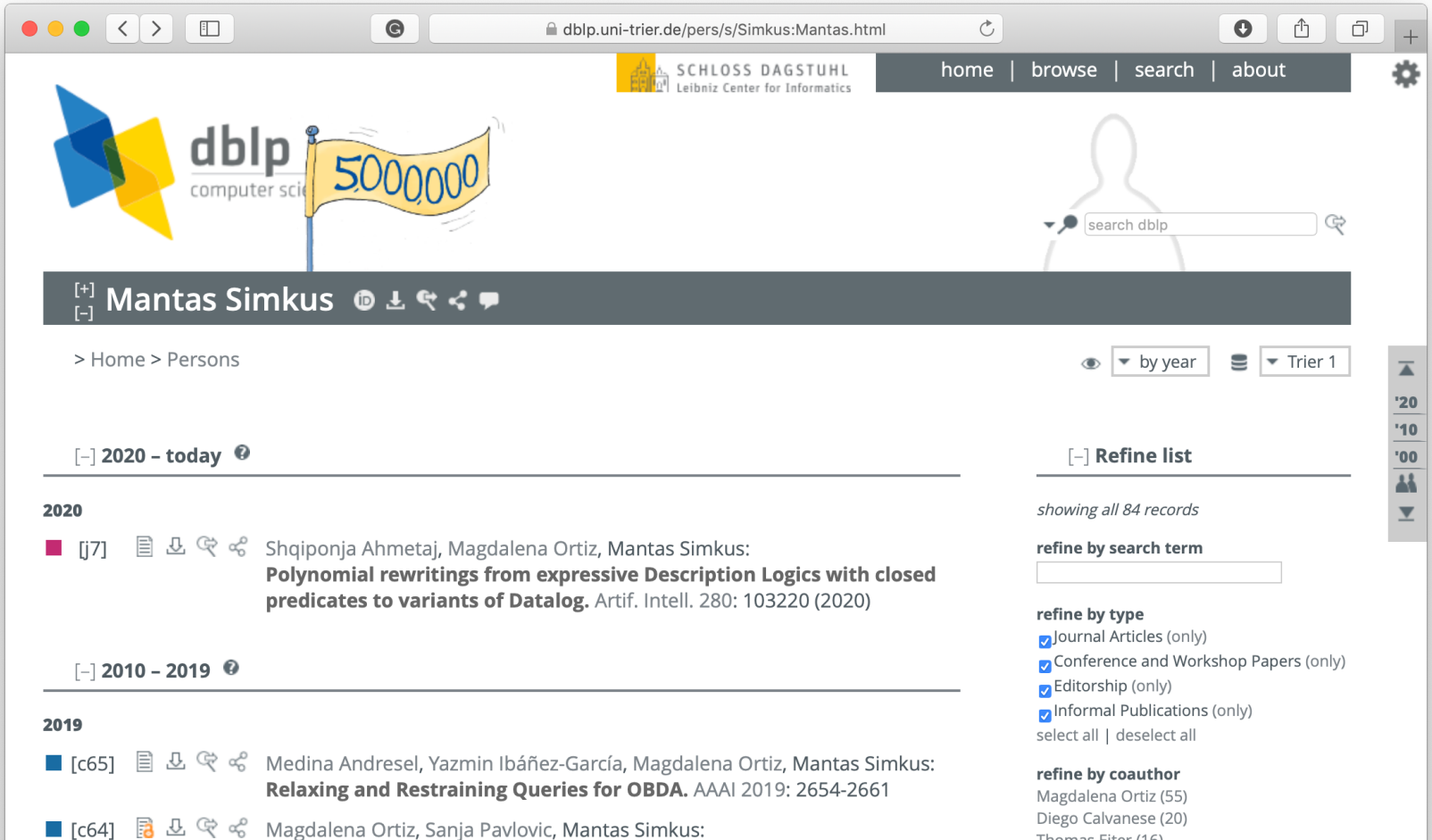
AUTHOR: Douglas Adams

The Hitchhiker's Guide to the Galaxy

A more complex example

DBLP Computer science bibliography

- Is a an online computer science bibliography listing all major journals and conferences in computer science
  - http://dblp.org/



Our DBLP scenario

- DBLP provides its data (e.g. publication lists of authors) as xml data
- We aim to use these files to create publication lists for the websites of several authors.
- As first step we get a dataset from the dblp website:
  - https://dblp.uni-trier.de/pers/xx/s/Simkus:Mantas.xml
- The dataset contains
  - Meta-information about the author
  - A list of publications
  - Some information about the co-authors (which we ignore)

Our DBLP scenario - Meta-information for the author

```
<dblpperson name="Mantas Simkus" pid="01/5959" n="84">
        <person key="homepages/01/5959" mdate="2009-06-10">
                <author pid="01/5959">Mantas Simkus</author>
                <note type="affiliation">TU Wien, Austria</note>
        </person>
[…]
</dblpperson>
```

Sometimes inaccurate and/or outdated

Our DBLP scenario - A list of publications

```
<dblpperson name="Mantas Simkus" pid="01/5959" n="84">
[…]
    <r>
        <article key="journals/ai/AhmetajOS20" mdate="2020-02-11">
        <author pid="144/7445">Shqiponja Ahmetaj</author>
        <author pid="71/3644">Magdalena Ortiz</author>
        <author pid="01/5959">Mantas Simkus</author>
        <title>Polynomial rewritings from expressive Description Logics with closed predicates to
            variants of Datalog.</title>
        <pages>103-220</pages><year>2020</year><volume>280</volume><journal>Artif.
            Intell.</journal>
        <ee>https://doi.org/10.1016/j.artint.2019.103220</ee>
        <url>db/journals/ai/ai280.html#AhmetajOS20</url>
        </article>
    </r>
[…]
</dblpperson>
```

A journal article

list of authors

Our DBLP scenario - A list of publications

```
<dblpperson name="Mantas Simkus" pid="01/5959" n="84">
[…]
        <r>
            <inproceedings key="conf/www/AhmetajFPSS15" mdate="2019-10-19">
                <author pid="144/7445">Shqiponja Ahmetaj</author>
                <author pid="148/7353">Wolfgang Fischl</author>
                <author pid="26/5252">Reinhard Pichler</author>
                <author orcid="0000-0003-0632-0294" pid="01/5959">Mantas Simkus</author>
                <author orcid="0000-0003-3054-7683" pid="22/8384">Sebastian Skritek</author>
                <title>Towards Reconciling SPARQL and Certain Answers.</title>
                <pages>23-33</pages><year>2015</year><booktitle>WWW</booktitle>
                <ee>https://doi.org/10.1145/2736277.2741636</ee>
                <crossref>conf/www/2015</crossref>
                <url>db/conf/www/www2015.html#AhmetajFPSS15</url>
            </inproceedings>
        </r>
[…]
</dblpperson>
```

A paper in conference proceddings

list of authors

Our DBLP scenario – first XSLT transformation

- We first give a simple XSLT transformation that creates a webpage with
  - A list of journal publications, and
  - A list of conference papers
- and later improve it in several directions

Our DBLP scenario – first XSLT transformation

```
<?xml version="1.0"?>
<xsl:stylesheet  version="2.0"   xmlns:xsl="http://www.w3.org/1999/XSL/Transform">
   <xsl:output method="html"/>

    […]

</xsl:stylesheet>
```

Our DBLP scenario – first XSLT transformation

A template for the root element:

```
<xsl:template match="dblpperson">
  <html><head><title>Publications</title></head>
    <body>
      <h1>Publications</h1>
        <h2>Journal Publications</h2>
          <ol>
            <xsl:apply-templates select=".//article"/>
          </ol>
        <h2>Conference Papers</h2>
          <ol>
            <xsl:apply-templates select=".//inproceedings"/>
          </ol>
        </body>
  </html>
 </xsl:template>
```

Our DBLP scenario – first XSLT transformation

A template for formatting article entries:

```
<xsl:template match="article">
    <li>
        <xsl:apply-templates select="author"/><br/>
        <b><xsl:value-of select="title"/></b><br/>
        <xsl:value-of select="journal"/>, <xsl:value-of select="volume"/>,
        <xsl:value-of select="pages"/>, <xsl:value-of select="year"/>
    </li>
</xsl:template>
```

Our DBLP scenario – first XSLT transformation

A template for formatting conference paper entries:

```
<xsl:template match="inproceedings">
    <li>
        <xsl:apply-templates select="author"/><br/>
        <b><xsl:value-of select="title"/></b><br/>
        In Proceedings of <xsl:value-of select="booktitle"/>,
        <xsl:value-of select="pages"/> (<xsl:value-of select="year"/>)
    </li>
</xsl:template>
```

```
<xsl:template match="author">
    <xsl:value-of select="."/>,
</xsl:template>
```

Our DBLP scenario – first XSLT transformation

```
<r>
    <article key="journals/ai/AhmetajOS20" mdate="2020-02-11">
        <author pid="144/7445">Shqiponja Ahmetaj</author>
        <author pid="71/3644">Magdalena Ortiz</author>
        <author pid="01/5959">Mantas Simkus</author>
        <title>Polynomial rewritings from expressive Description Logics with closed
        predicates to variants of Datalog.</title>
        <pages>103220</pages>
        <year>2020</year>
        <volume>280</volume>
        <journal>Artif. Intell.</journal>
        <ee>https://doi.org/10.1016/j.artint.2019.103220</ee>
        <url>db/journals/ai/ai280.html#AhmetajOS20</url>
    </article>
</r>
```

```
<li>Shqiponja Ahmetaj,
        Magdalena Ortiz,
        Mantas Simkus,
        <br><b>Polynomial rewritings from expressive Description Logics with closed
predicates to variants of Datalog.</b><br>Artif. Intell., 280, 103220, 2020
 </li>
```

Our DBLP scenario – first XSLT transformation

```
<r>
    <inproceedings key="conf/www/AhmetajFPSS15" mdate="2019-10-19">
        <author pid="144/7445">Shqiponja Ahmetaj</author>
        <author pid="148/7353">Wolfgang Fischl</author>
        <author pid="26/5252">Reinhard Pichler</author>
        <author orcid="0000-0003-0632-0294" pid="01/5959">Mantas Simkus</author>
        <author orcid="0000-0003-3054-7683" pid="22/8384">Sebastian Skritek</author>
        <title>Towards Reconciling SPARQL and Certain Answers.</title>
        <pages>23-33</pages>
        <year>2015</year>
        <booktitle>WWW</booktitle>
        <ee>https://doi.org/10.1145/2736277.2741636</ee>
        <crossref>conf/www/2015</crossref>
        <url>db/conf/www/www2015.html#AhmetajFPSS15</url>
    </inproceedings>
</r>
```

```
<li>Shqiponja Ahmetaj,
    Wolfgang Fischl,
    Reinhard Pichler,
    Mantas Simkus,
    Sebastian Skritek,
    <br><b>Towards Reconciling SPARQL and Certain Answers.</b><br>
        In Proceedings of WWW, 23-33 (2015)
    </li>
```

Our DBLP scenario – first XSLT transformation

- This XSLT transformation yields `dblp-simple.html` (see TUWEL)

- Issues with that webpage:
  - In does not distingush between articles in peer reviewed journals and informal publications
  - It does not deal with missing values (we get several commas in a row)
  - The webpage does not mention the author for whom we created the publication list
  - No css-stylesheets for the webpage, no Ids for list entries, etc.
  - ...

Our DBLP scenario – informal publications

```
<dblpperson name="Mantas Simkus" pid="01/5959" n="84">
[…]
   <r>
      <article publtype="informal" key="journals/corr/BienvenuCOS14" mdate="2018-08-13">
            <author pid="80/28">Meghyn Bienvenu</author>
            <author pid="c/DiegoCalvanese">Diego Calvanese</author>
            <author pid="71/3644">Magdalena Ortiz</author>
            <author pid="01/5959">Mantas Simkus</author>
            <title>Nested Regular Path Queries in Description Logics.</title>
            <year>2014</year>
            <volume>abs/1402.7122</volume>
            <journal>CoRR</journal>
            <ee type="oa">http://arxiv.org/abs/1402.7122</ee>
            <url>db/journals/corr/corr1402.html#BienvenuCOS14</url>
      </article>
   </r>
[…]
</dblpperson>
```

Informal publications have an attribute publtype

Our XSLT– Template for the root element

```
<xsl:template match="dblpperson">
  <html>
    <head>
      <title>Publications - <xsl:value-of select="@name"/></title>
      <link rel="stylesheet" href="all.css" media="all" type="text/css"/>
      <link rel="stylesheet" href="print.css" media="print" type="text/css"/>
    </head>
    <body>
      <h1>Publications of <xsl:value-of select="@name"/></h1>
          <h2  id="journal">Journal Publications</h2>
              <ol><xsl:apply-templates select=".//article[not(@publtype='informal publication')]"/></ol>
          <h2  id="conference">Conference Papers</h2>
              <ol><xsl:apply-templates select=".//inproceedings"/></ol>
          <h2 id="informal">Informal Publications</h2>
              <ol><xsl:apply-templates select=".//article[@publtype='informal publication']"
                    mode="informal"/></ol>
    </body>
  </html>
</xsl:template>
```

Our XSLT– Templates for article elements

```
<xsl:template match="article">
    <li id="{@key}">
        <xsl:call-template name="author_and_title"/>
        <xsl:value-of select="journal"/>, <xsl:value-of select="volume"/>,
        <xsl:value-of select="pages"/>, <xsl:value-of select="year"/>.
    </li>
</xsl:template>

<xsl:template match="article“ mode=“informal”>
    <li id="{@key}">
        <xsl:call-template name="author_and_title"/>
        <xsl:value-of select="journal"/>, <xsl:value-of select="volume"/>,
        <xsl:value-of select="year"/>.
    </li>
</xsl:template>
```

Our XSLT– Template for Conference papers

```
<xsl:template match="inproceedings">
 <li id="{@key}">
              <xsl:call-template name="author_and_title"/>
               In Proceedings of <xsl:value-of select="booktitle"/>,
               <xsl:value-of select="pages"/> (<xsl:value-of select="year"/>)
    </li>
</xsl:template>
```

```
 <xsl:template name="author_and_title">
               <xsl:apply-templates select="author"/><br/>
               <b><xsl:value-of select="title"/></b><br/>
  </xsl:template>
```

Our XSLT– Template for Authors

```
<xsl:template match="author">
  <xsl:choose>
      <xsl:when test="./text()='Mantas Simkus' ">
                Mantas Šimkus,
      </xsl:when>
      <xsl:otherwise>
                <xsl:value-of select="."/>,
      </xsl:otherwise>
  </xsl:choose>
</xsl:template>
```

Our DBLP scenario

```
<r>
        <article key="journals/ai/AhmetajOS20" mdate="2020-02-11">
                <author pid="144/7445">Shqiponja Ahmetaj</author>
                <author pid="71/3644">Magdalena Ortiz</author>
                <author pid="01/5959">Mantas Simkus</author>
                <title>Polynomial rewritings from expressive Description Logics with closed
                predicates to variants of Datalog.</title>
                <pages>103220</pages>
                <year>2020</year>
                <volume>280</volume>
                <journal>Artif. Intell.</journal>
                <ee>https://doi.org/10.1016/j.artint.2019.103220</ee>
                <url>db/journals/ai/ai280.html#AhmetajOS20</url>
        </article>
</r>
```

```
<li id="journals/ai/AhmetajOS20">Shqiponja Ahmetaj,
                Magdalena Ortiz,
                Mantas Šimkus,
                <br><b>Polynomial rewritings from expressive Description Logics with closed
predicates to variants of Datalog.</b><br>Artif. Intell., 280, 103220, 2020.

</li>
```

Our DBLP scenario

```
<r>
        <inproceedings key="conf/www/AhmetajFPSS15" mdate="2019-10-19">
                <author pid="144/7445">Shqiponja Ahmetaj</author>
                <author pid="148/7353">Wolfgang Fischl</author>
                <author pid="26/5252">Reinhard Pichler</author>
                <author orcid="0000-0003-0632-0294" pid="01/5959">Mantas Simkus</author>
                <author orcid="0000-0003-3054-7683" pid="22/8384">Sebastian Skritek</author>
                <title>Towards Reconciling SPARQL and Certain Answers.</title>
                <pages>23-33</pages>
                <year>2015</year>
                <booktitle>WWW</booktitle>
                <ee>https://doi.org/10.1145/2736277.2741636</ee>
                <crossref>conf/www/2015</crossref>
                <url>db/conf/www/www2015.html#AhmetajFPSS15</url>
        </inproceedings>
</r>
```

```
<li id="conf/www/AhmetajFPSS15">Shqiponja Ahmetaj,
                  Wolfgang Fischl,
                  Reinhard Pichler,
                  Mantas Šimkus,
                  Sebastian Skritek,
                  <br><b>Towards Reconciling SPARQL and Certain Answers.</b><br>
                 In Proceedings of WWW, 23-33 (2015)

</li>
```

Our DBLP scenario

```
<r>
        <article publtype="informal" key="journals/corr/BienvenuCOS14" mdate="2018-08-13">
                <author pid="80/28">Meghyn Bienvenu</author>
                <author pid="c/DiegoCalvanese">Diego Calvanese</author>
                <author pid="71/3644">Magdalena Ortiz</author>
                <author pid="01/5959">Mantas Simkus</author>
                <title>Nested Regular Path Queries in Description Logics.</title>
                <year>2014</year>
                <volume>abs/1402.7122</volume>
                <journal>CoRR</journal>
                <ee type="oa">http://arxiv.org/abs/1402.7122</ee>
                <url>db/journals/corr/corr1402.html#BienvenuCOS14</url>
        </article>
</r>
```

```
<li id="journals/corr/BienvenuCOS14">Meghyn Bienvenu,
                Diego Calvanese,
                Magdalena Ortiz,
                Mantas Šimkus,
                <br><b>Nested Regular Path Queries in Description Logics.</b><br>CoRR,
abs/1402.7122, 2014.

</li>
```

Our DBLP scenario

- This XSLT results dblp-publications.html (see TUWEL)

Our DBLP scenario – co-authors

- Now as we have a publication list we are interested in a list of co-authors.

Two scientist are considered to be co-authors if they have written at least one joint publication.

Our XSLT– Template for the root element

```
<xsl:template match="dblpperson">
  <html>
    <head>
       <title>Publications - <xsl:apply-templates select="@name"/></title>
      <link rel="stylesheet" href="all.css" media="all" type="text/css"/>
      <link rel="stylesheet" href="print.css" media="print" type="text/css"/>
    </head>
    <body>
      <h1>Publications</h1>
          <h2 id="coAuthors">Co-authors of <xsl:apply-templates select="@name"/></h2>
           <ul>
             <xsl:variable name="author" select="@name"/>
             <xsl:apply-templates
                  select="//author[not(./text() = $author)][not(./text() = preceding::author/text())]"/>
          </ul>
    </body>
  </html>
</xsl:template>
```

List all co-authors (but not the author)

List all co-authors (but avoid duplicates)

Our XSLT– Template for author elements

```
<xsl:template match="author">
    <li>
            <b>
              <xsl:value-of select="."/>
            </b>
            <br/>
            […NEXT SLIDE…]
     </li>
</xsl:template>
```

Present the name of the author and start new line.

Our XSLT– Template for author elements

```
<xsl:template match="author">
  <li>
      [… PREVIOUS SLIDE… .]
      <xsl:variable name="coauthor" select="."/>
      Number of joint papers:
      <xsl:value-of select="count(//r//author[./text()=$coauthor/text()])"/> <br/>
      Joint papers:
      <xsl:for-each select="//r[.//author/text()=$coauthor/text()]">
         <xsl:sort select="element()/year" order="ascending"/>
          <a href="dblp-publications.html#{element()/@key}">
              [<xsl:value-of select="position()"/>]
          </a>
          <xsl:if test="position()!=last()">, </xsl:if>
       </xsl:for-each><br/>
       […NEXT SLIDE…]
  </li>
</xsl:template>
```

Count joint publications

List joint publications and link to the entry in publications.html

Our XSLT– Template for author elements

```
<xsl:template match="author">
  <li>
      […PREVIOUS SLIDE…]
     Joint co-authors:
     <xsl:variable name="coauthors"
select="//author[not(.=//dblpperson/@name)][not(.=$coauthor)][../author=$coauthor]"/>
     <xsl:apply-templates select="$coauthors[index-of($coauthors/text(),./text())[1]]"
       mode="simple"/>
  </li>
</xsl:template>
```

List joint co-authors
(but avoid duplicates)

Our XSLT– Remaining templates

```
<xsl:template match="author" mode="simple">
        <xsl:value-of select="."/>,
</xsl:template>

<xsl:template match="@name">
    *** <xsl:value-of select="."/> ***
</xsl:template>
```

Our XSLT– Example output (see dblp-authors.html)

```
[…]
<h2 id="coAuthors">Co-authors of Mantas Simkus</h2>
    <ul>
      <li><b>Shqiponja Ahmetaj</b><br>
            Number of joint papers: 15<br>
            Joint papers:
            <a href="publications.html#conf/aaai/AhmetajCOS14">[1]</a>, <a
href="publications.html#conf/amw/AhmetajCOS14">[2]</a>, […] <a
href="publications.html#journals/ai/AhmetajOS20">[15]</a><br>
            Joint co-authors: Magdalena Ortiz, [….] Sebastian Skritek,
      </li>
[….]
```

Our DBLP scenario – co-authors

- Finally, we want to order the co-authors by the number of joint publications

- We only have to change the template for the root element

Our XSLT– Template for the root element

```
<xsl:template match="dblpperson">
  <html>
    <head>
      <title>Publications - <xsl:apply-templates select="@name"/></title>
      <link rel="stylesheet" href="all.css" media="all" type="text/css"/>
      <link rel="stylesheet" href="print.css" media="print" type="text/css"/>
    </head>
    <body>
      <h1>Publications</h1>
        <h2 id="coAuthors">Co-authors of <xsl:apply-templates select="@name"/></h2>
        <ul>
            [next slide]
        </ul>
    </body>
  </html>
</xsl:template>
```

Sofar unchanged

Our XSLT– Template for the root element

```
<xsl:template match="dblpperson">
[…]
<ul>
   <xsl:variable name="author" select="@name"/>

   <xsl:variable name="coauthors">
        <xsl:for-each-group select="//r//author" group-by="text()">
          <xsl:variable name="coauthor" select="."/>
           <author count="{count(//r//author[./text()=$coauthor/text()])}">
           <xsl:value-of select="./text()"/></author>
        </xsl:for-each-group>
   </xsl:variable>
   […]
</ul>
[…]
</xsl:template>
```

In the variable *coauthors* we store new author elements that already have the number of joint publications as attribute

Our XSLT– Template for the root element

```
<xsl:template match="dblpperson">
[…]
<ul>
   <xsl:variable name="author" select="@name"/>
   <xsl:variable name="coauthors">
      […]
   </xsl:variable>
   <xsl:variable name="dblpperson" select="."/>
   <xsl:for-each select="$coauthors/author">
      <xsl:sort select="@count" data-type="number" order="descending"/>
       <xsl:variable name="coauthor" select="."/>
       <xsl:apply-templates
select="$dblpperson//author[not(./text() = $author)][./text()=$coauthor/text()][not(./text() =
preceding::author/text())]"/>
   </xsl:for-each>
</ul>
[…]
</xsl:template>
```

In *coauthors* we store author elements that have the number of joint papers as attribute

Reference to the root element

One additional predicate that selects only elements that correspond to the current co-author

Our DBLP scenario – co-authors

- Finally, we ordered the co-authors by the number of joint publications
- The other templates are unchanged
- An example output is given in `dblp-authors2.html` (TUWEL)