E-Fuice

Integration of E-Commerce Data
Agenda

- iFuice applications in the E-Commerce domain
- E-Fuice prototype
- Matching instances and ontologies using iFuice scripts
- First results of matching
- Outlook
E-Fuice: Product Search and Recommendations

- For end users:
  - Structured product search with query language:

    Find all products which conform to query conditions:

    (ObjectType="Product@Amazon" OR
     ObjectType="Product@Hugendubel")

    AND ProductType="Book" AND Category ="Fiction"

    AND (Cover="Hardcover" OR Audiobook=true)

    AND Price<80

  - Product recommendations based on different data sources:
    
    „Following products from our assortment received the best client reviews on Amazon.de“
E-Fuice: Product Comparison

For end users:

- Product comparison: compare characteristics of different products:

<table>
<thead>
<tr>
<th>Digital-Kamera DC 4400</th>
<th>Digital-Kamera DC 50slim</th>
</tr>
</thead>
<tbody>
<tr>
<td>4 Mio. Pixel</td>
<td>5 Mio. Pixel</td>
</tr>
<tr>
<td>3x optisches Zoom</td>
<td>3x optisches Zoom</td>
</tr>
<tr>
<td>4x digitales Zoom</td>
<td>4x digitales Zoom</td>
</tr>
<tr>
<td>€ 119.95</td>
<td>€ 179.95</td>
</tr>
</tbody>
</table>

- Price comparison: compare prices of identical/similar products in different shops:

<table>
<thead>
<tr>
<th>Quelle</th>
<th>Amazon</th>
<th>Ebay</th>
</tr>
</thead>
<tbody>
<tr>
<td>Canon IXUS 40</td>
<td>IXUS 40</td>
<td>CANON IXUS 40! NEUWARE! RECHNUNG! HÄNDLER! OVP</td>
</tr>
<tr>
<td>Preis: 289.00</td>
<td>Preis: 229.00</td>
<td>Preis: 249.00</td>
</tr>
</tbody>
</table>
E-Fuice: Analysis of the Market Competition

For business users:

- Analyze competitors‘ product assortment:
  - Product listings: which identical/similar products cost more/less at competitors‘ shops?
  - Price level: what is the average price of all products in the same category at competitors‘ shops?
  - Breadth of the price spectrum: how large is the price difference between the cheapest and the most expensive product in a product category? How large is the price variance?
E-Fuice Prototype

Integrating 4 Online-Shops:
www.amazon.com
www.ebay.de
www.softunity.com
www.bestpreis24.de

Product Assortment:
Software
Games, Game Console
Accessories
Movies (DVD, VHS)
E-Fuice Prototype

- For business users -- executing queries over integrated data from iFuice interface

- For end users -- Web portal featuring:
  - Navigation in the integrated data
  - Complex queries
  - Recommendations
E-Fuice: Source Mapping Model

|Ontology@Softunity|=466
|Ontology@Amazon|=1930, manually pruned
|Ontology@Ebay|=1141

|Product@Softunity|=2711
|Product@Amazon|=93897
|Product@Ebay|=? (many)

Product@Bestpreis24
Review@Amazon
#Loading ontologies
$suonto:=queryInstances(Ontology@Softunity,ALL)
$amonto:=queryInstances(Ontology@Amazon,ALL)

#Loading mapping from ontologies to products and product details
$amontoprod:=map($amonto,Amazon.OntoProd)
$suontoprod:=map($suonto,Softunity.OntoProd)
$amprod:=loadInstances(Product@Amazon)
$suprod:=loadInstances(Product@Softunity)

#Matching products
$su_am_instance_eqEAN:=joinMapResult($suprod,$amprod,[domain.ean]=[range.ean])
$su_am_instance_title_trigram09:=attrMatch($suprod,$amprod,0.9,[title],[title])
Matching Ontologies and Instances Using iFuice Scripts (continued)

#Loading manual ontology mapping
$am_su_onto_mm:=map($amonto,Softunity.ontoAmazon2SoftunityManualMerged)

#Executing COMA ontology mapping
$am_su_onto_COMA:=map($amonto,ontoAmazon2SoftunityCOMAContextsSyn)

#Creating instance-based match for ontologies
$su_am_onto_over_EAN:=composeMatchDice(compose($suontoprod,$su_am_instance_eqEAN), inverse($amontoprod),0.5)

#Creating combined mapping (COMA + Instances)
#Other combination algorithms -> further research
$su_am_combined:=union(inverse($am_su_onto_COMA), $su_am_onto_over_EAN)

#Checking correct correspondences
$correct_COMA:=intersect($su_am_onto_COMA, $am_su_onto_mm)
$correct_EAN:=intersect($su_am_onto_over_EAN, $am_su_onto_mm)
$correct_combined:=intersect($su_am_combined,$am_su_onto_mm)

#Calculate recall/precision
$recall_COMA:=count($correct_COMA)/count($am_su_onto_mm)
$precision_COMA:=count($correct_COMA)/count($am_su_onto_COMA)

...
Matching Amazon and Ebay Ontologies with COMA++

Both ontologies loaded as OWL files into COMA++:

Ebay Ontology:  1137 nodes, Amazon Ontology: 1930 nodes (manually pruned)

<table>
<thead>
<tr>
<th>Match Strategy</th>
<th>Nr of corr.</th>
<th>Nr of corr. correct</th>
<th>Recall</th>
<th>Precision</th>
</tr>
</thead>
<tbody>
<tr>
<td>MANUAL</td>
<td>374</td>
<td>374</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>FILTEREDCONTEXT: NODES/COMA</td>
<td>179</td>
<td>51</td>
<td>0.1363</td>
<td>0.2849</td>
</tr>
<tr>
<td>FILTEREDCONTEXT: NODES/CONTEXT</td>
<td>123</td>
<td>51</td>
<td>0.1363</td>
<td>0.4146</td>
</tr>
<tr>
<td>ALLCONTEXT: TAXONOMY (softunity as global taxonomy)</td>
<td>2846</td>
<td>189</td>
<td>0.5121</td>
<td>0.0664</td>
</tr>
<tr>
<td>ALLCONTEXT: TAXONOMY_AVERAGE (softunity as global taxonomy)</td>
<td>2264</td>
<td>171</td>
<td>0.4634</td>
<td>0.0755</td>
</tr>
<tr>
<td>ALLCONTEXT: COMA</td>
<td>244</td>
<td>101</td>
<td>0.27</td>
<td>0.4139</td>
</tr>
<tr>
<td>ALLCONTEXT: CONTEXTS</td>
<td>336</td>
<td>185</td>
<td>0.4946</td>
<td>0.5505</td>
</tr>
<tr>
<td>ALLCONTEXT: CONTEXTS with SYNONYMS, SELECTION 0, 0.01, 0.5 (DEFAULT)</td>
<td>341</td>
<td>202</td>
<td>0.5401</td>
<td>0.5923</td>
</tr>
<tr>
<td>ALLCONTEXT: CONTEXTS with SYNONYMS, SELECTION 2, 0.03, 0.7</td>
<td>347</td>
<td>218</td>
<td>0.5828</td>
<td>0.6282</td>
</tr>
</tbody>
</table>
Matching Softunity and Amazon Ontologies with COMA++

Both ontologies loaded as OWL files into COMA:

Softunity Ontology: 466 nodes, Amazon Ontology: 1930 nodes (manually pruned)

<table>
<thead>
<tr>
<th>Match Strategy</th>
<th>Nr of corr.</th>
<th>Nr of corr. correct</th>
<th>Recall</th>
<th>Precision</th>
</tr>
</thead>
<tbody>
<tr>
<td>MANUAL</td>
<td>212</td>
<td>212</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>ALLCONTEXT: COMA</td>
<td>64</td>
<td>25</td>
<td>0.11792453</td>
<td>0.390625</td>
</tr>
<tr>
<td>ALLCONTEXT: CONTEXTS_SYN SEL 0.0,0.01,0.5 (default)</td>
<td>156</td>
<td>112</td>
<td>0.5283019</td>
<td>0.71794873</td>
</tr>
<tr>
<td>ALLCONTEXT: CONTEXTS_SYN SEL 0.0,0.02,0.5</td>
<td>194</td>
<td>117</td>
<td>0.5518868</td>
<td>0.6030928</td>
</tr>
<tr>
<td>ALLCONTEXT: CONTEXTS_SYN SEL 0.0,0.008,0.5</td>
<td>174</td>
<td>114</td>
<td>0.5377358</td>
<td>0.6551724</td>
</tr>
<tr>
<td>ALLCONTEXT: CONTEXTS_SYN SEL 2,0.03,0.7</td>
<td>45</td>
<td>32</td>
<td>0.1509434</td>
<td>0.7111111</td>
</tr>
</tbody>
</table>
Matching Products: Softunity vs. Amazon

Products matched using IFuice.

Softunity: 2711 products, Amazon: 42942 (Software and Games)

<table>
<thead>
<tr>
<th>Match Strategy</th>
<th>Nr of corr.</th>
<th>Nr of corr. correct</th>
<th>Recall</th>
<th>Precision</th>
</tr>
</thead>
<tbody>
<tr>
<td>UNAMBIGUOUS MATCHING (With EAN)</td>
<td>2133</td>
<td>2133</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Exact matching based on product title</td>
<td>925</td>
<td>779</td>
<td>0.3652</td>
<td>0.8421</td>
</tr>
<tr>
<td>Trigram matching based on title, threshold 0.99</td>
<td>926</td>
<td>780</td>
<td>0.3656</td>
<td>0.8423</td>
</tr>
<tr>
<td>Trigram matching based on title, threshold 0.9</td>
<td>1296</td>
<td>801</td>
<td>0.3755</td>
<td>0.6180</td>
</tr>
<tr>
<td>Trigram matching based on title, threshold 0.7</td>
<td>3915</td>
<td>1500</td>
<td>0.7032</td>
<td>0.3831</td>
</tr>
<tr>
<td>Trigram matching based on title, threshold 0.5</td>
<td>32647</td>
<td>1947</td>
<td>0.9127</td>
<td>0.0596</td>
</tr>
<tr>
<td>MSSQL Fuzzy Matching, threshold 0.9</td>
<td>1684</td>
<td>932</td>
<td>0.4369</td>
<td>0.5534</td>
</tr>
<tr>
<td>MSSQL Fuzzy Matching, threshold 0.7</td>
<td>3942</td>
<td>1435</td>
<td>0.6727</td>
<td>0.3655</td>
</tr>
</tbody>
</table>
Combined Mapping: Softunity Ontology vs. Amazon Ontology Using Instance(Product) Mappings

Softunity Ontology: 466 nodes, Amazon Ontology: 1930 nodes (manually pruned)
Softunity: 2711 products, Amazon: 93897 products

<table>
<thead>
<tr>
<th>Match Strategy</th>
<th>Nr of corr.</th>
<th>Nr of corr. correct</th>
<th>Recall</th>
<th>Precision</th>
</tr>
</thead>
<tbody>
<tr>
<td>MANUAL COMBINED MAPPING</td>
<td>351</td>
<td>351</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>INSTANCE EAN Dice Sim. &gt;0.5</td>
<td>286</td>
<td>209</td>
<td>0.5954</td>
<td>0.7307</td>
</tr>
<tr>
<td>INSTANCE EAN Dice Sim. &gt;0.7</td>
<td>135</td>
<td>109</td>
<td>0.3105</td>
<td>0.8074</td>
</tr>
<tr>
<td>INSTANCE EAN Dice Sim. &gt;0.9</td>
<td>68</td>
<td>53</td>
<td>0.1509</td>
<td>0.7794</td>
</tr>
<tr>
<td>COMA CONTEXTS_SYN</td>
<td>156</td>
<td>112</td>
<td>0.3190</td>
<td>0.7179</td>
</tr>
<tr>
<td>COMBINED INSTANCE EAN Dice Sim. &gt;0.5 AND CONTEXTS_SYN</td>
<td>410</td>
<td>287</td>
<td>0.8176</td>
<td>0.7</td>
</tr>
<tr>
<td>COMBINED INSTANCE EAN Dice Sim. &gt;0.7 AND CONTEXTS_SYN</td>
<td>265</td>
<td>193</td>
<td>0.5498</td>
<td>0.7283</td>
</tr>
<tr>
<td>COMBINED INSTANCE EAN Dice Sim. &gt;0.9 AND CONTEXTS_SYN</td>
<td>210</td>
<td>148</td>
<td>0.4216</td>
<td>0.7047</td>
</tr>
</tbody>
</table>

- Dice coefficient: $D = \frac{2N_c}{Na + Nb}$, where $N_c$ – number of instance correspondences for a concept, $Na$, $Nb$ – number of instances in the concept in the ontologies being matched
Combined Mapping (continued)

- Overlap between COMA mapping and instance Mapping (Dice>0.5): ~7%.
- The correspondences which are found only in COMA mapping can be attributed to the fact that only 39% of the concepts in softunity ontology have instances.
- The correspondences which are found only in the Instance mapping are based on background knowledge, which is not explicitly found in the names of ontology concepts. Examples:

  *Unterhaltung- und Gesellschaftsspiele* -> *Fun-Spiele*
  *Spiele/Konsolen/Zubehör/Nach Hersteller/Microsoft* -> *Spiele/X-Box/Originalzubehör*

- Problem: orthogonal subontologies, for example:

  *Filme:*
  *Nach Genre*
  *Nach Produktionsland*

  Because of this, „Zeichentrick“ is mapped to „Japan“, „Classic Western Collection“ to „Italien“. Special heuristics are needed to resolve this.
Further Research Questions:

- E-Commerce data are highly dynamic (ontologies may change once in several month, instances change up to several times a week, ebay -- constantly):
  
  *Automatic ontology evolution support using new ontology and instance data, could be implemented as iFuice script*

- *(Reuse for negative results)*
Further Research Questions:

- Instance-based correspondences can be used not only directly for matching ontologies, but also to amend „synonym“ tables for reuse by name matchers. (example: \textit{Unterhaltung} -> \textit{Fun}, etc.)

- Orthogonal subontologies: often correspond to attribute names and values in objects. (example: „\textit{Nach Produktionsland}“ -> attribute \textit{Produktionsland})
Further Research Questions:

- Ontology-Correspondences in our implementation are very simple: no distinction between different types of relations is made (subclassing, equality). This is sufficient for web recommendations, but may be insufficient for structured queries.
Thank you for your attention.