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 :

Digital-Kamera DC 4400	Digital-Kamera DC 50slim
4 Mio. Pixel	5 Mio. Pixel
3x optisches Zoom	3x optisches Zoom
4x digitales Zoom	4x digitales Zoom
€ 119.95	€ 179.95

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

Quelle	Amazon	Ebay
Canon IXUS 40	IXUS 40	CANON IXUS 40! NEUWARE! RECHNUNG! HÄNDLER! OVP
Preis: 289.00	Preis: 229.00	Preis: 249.00

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

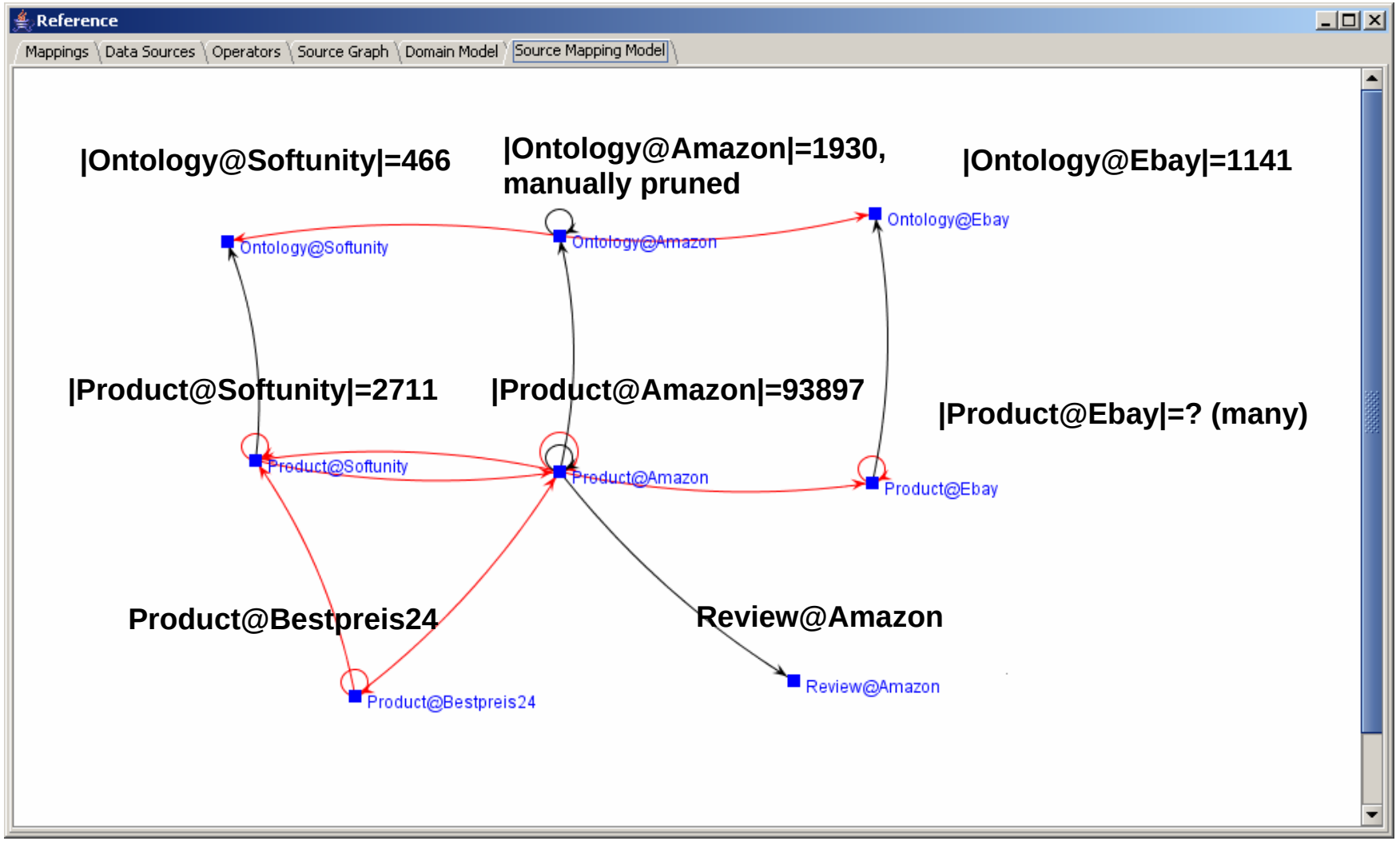
Games, Game Console Accessories

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



Matching Ontologies and Instances Using iFuice Scripts

#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)

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

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)

Match Strategy	Nr of corr.	Nr of corr. correct	Recall	Precision
MANUAL	212	212	1	1
ALLCONTEXT: COMA	64	25	0.11792453	0.390625
ALLCONTEXT: CONTEXTS_SYN SEL 0,0.01,0.5 (default)	156	112	0.5283019	0.71794873
ALLCONTEXT: CONTEXTS_SYN SEL 0,0.02,0.5	194	117	0.5518868	0.6030928
ALLCONTEXT: CONTEXTS_SYN SEL 0,0.008,0.5	174	114	0.5377358	0.6551724
ALLCONTEXT: CONTEXTS_SYN SEL 2,0.03,0.7	45	32	0.1509434	0.7111111

Matching Products: Softunity vs. Amazon

Products matched using IFuice.

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

Match Strategy	Nr of corr.	Nr of corr. correct	Recall	Precision
UNAMBIGUOUS MATCHING (With EAN)	2133	2133	1	1
Exact matching based on product title	925	779	0,3652	0,8421
Trigram matching based on title, threshold 0.99	926	780	0,3656	0,8423
Trigram matching based on title, threshold 0.9	1296	801	0,3755	0,6180
Trigram matching based on title, threshold 0.7	3915	1500	0,7032	0,3831
Trigram matching based on title, threshold 0.5	32647	1947	0,9127	0,0596
MSSQL Fuzzy Matching, threshold 0.9	1684	932	0,4369	0,5534
MSSQL Fuzzy Matching, threshold 0.7	3942	1435	0,6727	0,3655

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

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

• Dice coefficient: D= 2Nc/(Na+Nb), where Nc – 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: *Unterhaltung->Fun,* etc.)

- Orthogonal subontologies: often correspond to attribute names and values in objects.
  (example: *„Nach Produktionsland“* -> attribute *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.