Advanced Exploration of Public Procurement Data in Linked Data Paradigm

Krzysztof Wecel - I2G / Poznan University of Economics
Vojtech Svatek, Jindrich Mynarz – University of Economics, Prague
Motivation

Starting point:

- potential benefits for a wide range of players
- mandatory publication of public contract notices

Weak points:

- restrictions of search interfaces
- GUI in local languages
- no wider analyses available: aggregations, trends, patterns…
- lack of mechanical reasoning
- geographical context not leveraged
- no links to external information

Opportunity:

- representation in a form of linked data
Public Contracts Ontology
Public Contracts Ontology

pc:Notice
Public Contracts Ontology

pc:Notice

pc:Contract
Public Contracts Ontology

- pc:Contract
- pc:Notice
- vc:VCard
- gr:BusinessEntity
Public Contracts Ontology

- pc:Contract
- pc:Notice
- gr:BusinessEntity
- vc:VCard
- kind
- procedure
- CPV
Creating Knowledge out of Interlinked Data

Public Contracts Ontology

- pc:Contract
- pc:Notice
- vc:VCard
- gr:BusinessEntity
- gr:PriceSpecification
- pc:AwardCriteriaCombination
- kind
- procedure
- CPV
Creating Knowledge out of Interlinked Data

Public Contracts Ontology

- pc:Notice
- pc:Contract
- gr:Offering
- gr:PriceSpecification
- gr:BusinessEntity
- pc:AwardCriteriaCombination
- cp:VCard
- kind
- procedure
- CPV
Creating Knowledge out of Interlinked Data

Public Contracts Ontology

pc:Notice

vc:VCard

kind

procedure

CPV

pc:AwardCriteriaCombination

g:BusinessEntity

pc:Contract

gr:Offering

pc:Tender

gr:PriceSpecification
Public Contracts Ontology

Diagram:
- pc:Contract
  - pc:Notice
    - vc:VCard
    - kind
    - procedure
    - CPV
  - pc:AwardCriteriaCombination
  - pc:Tender
    - gr:Offering
    - gr:PriceSpecification
  - gr:BusinessEntity
Data sources and transformation

Public contract notices:
- HTML – navigation, scrapping
- XML – modeling approach
- mapping issues

Additional data – the real value
- business entities (ARES, CEIDG) and their codes (ICO, NIP, REGON)
- geographical codes (NUTS, TERYT)
- geographical coordinates (geocoding)
- CPV and other vocabularies
- optional external information
  - Czech Trade Inspection Authority
  - sentences of Polish National Board of Appeal
Polish dataset characteristics (2013)

number of triples: 28,8M
notices: 413,382
offerings: 922,038
contracting authorities: 17,648
contractors: 177,136
business entities: 194,784
unique CPV codes: 11,341
Polish dataset characteristics (2013)
Polish dataset characteristics (2013)
Polish dataset characteristics (2013)
Polish dataset characteristics (2013)
Data mining

leverage analytical and data mining techniques in order to find patterns, trends and anomalies in public contract data

Specific problems of graph data:
• multidimensionality
• big number of potential attributes
• overlap of the classes
• unbalanced learning data (different counts of classes)
• loss of information during transformation from graph to tabular data
CLUSTERING: looking for similar contracts

beneficiary: bidder

case:
• identify contracts from the past that would be most suitable
• monitor new notices similar the contracts they have already realised
• more expressive than typical search language

beneficiary: contracting authority

case:
• help in preparation of specific contract notice (not only CPV)
• aggregated demand opportunities
ASSOCIATIONS: ties between various market players

beneficiary: supervisory bodies

case:

• discover anomalies
• contractor-product association: stability of the offer; the tighter the relationship, the more reliable the contractor is
• contractor-authority association: signals the need to check for corruption
• analysis of the depth of the market
PREDICTIVE MODELS: number of bidders

leverages the link between notices concerning the same procurement process

beneficiary: contracting authority

case:
• the bigger number of bidders, the better
• one bidder can mean overspecified contract notice
Creating Knowledge out of Interlinked Data

Geography vs. value association

Case: Relations of number of contracts average estimated value by province
Number of tenders - long tail

- 38% of contracts had just one offer
- one bidder = overspecified contract notice?
- one contract had 610 offers
- in some cases large numbers of rejected offers: 298, 245, 111

<table>
<thead>
<tr>
<th>Number of Tenders</th>
<th>count</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>129910</td>
<td>38,2%</td>
</tr>
<tr>
<td>2</td>
<td>67929</td>
<td>20,0%</td>
</tr>
<tr>
<td>3</td>
<td>46412</td>
<td>13,6%</td>
</tr>
<tr>
<td>4</td>
<td>29850</td>
<td>8,8%</td>
</tr>
<tr>
<td>5</td>
<td>19367</td>
<td>5,7%</td>
</tr>
<tr>
<td>6</td>
<td>12679</td>
<td>3,7%</td>
</tr>
<tr>
<td>7</td>
<td>8277</td>
<td>2,4%</td>
</tr>
<tr>
<td>8</td>
<td>5639</td>
<td>1,7%</td>
</tr>
<tr>
<td>9</td>
<td>3920</td>
<td>1,2%</td>
</tr>
<tr>
<td>10</td>
<td>2799</td>
<td>0,8%</td>
</tr>
</tbody>
</table>
Some other discoveries

Type of contract vs. number of bidding entities
• supply contracts are the most popular
• construction work was the least popular
• areas of low competitiveness are more susceptible to abuse

Number of contractors to number of notices ratio
• should be compared to typical value in similar contracting authorities
Conclusions

Public procurement in Czech Republic and Poland
• ontology has been elaborated
• significant amount of data gathered
• we are looking for other interested parties

Data mining
specific issues of graph data have to be addressed
old and new tools applied
• similar contracts by clustering
• ties between various market players
• prediction of the number of bidders