Matchmaking public procurement linked open data

Jindřich Mynarz[1], Vojtěch Svátek[1], and Tommaso Di Noia[2]

[1] Department of Information and Knowledge Engineering,
University of Economics, W. Churchill Sq. 4, 130 67 Prague 3, Czech Republic
jindrich.mynarz|svatek@vse.cz
[2] Polytechnic University of Bari – Bari, Italy
tommaso.dinoia@poliba.it

Abstract. An increasing amount of public procurement data is nowadays being ported to linked data format, in view of its exploitation by government, commercial as well as non-profit subjects. One of the crucial tasks in public procurement is matchmaking demand with supply. We conceived this task as that of finding a supplier with previous successful history of contracts similar to a current call for tenders. In this paper we show how to implement a portable matchmaking service that relies solely on the capability of SPARQL 1.1. In order to show its effectiveness, the proposed service has been tested and evaluated on the RDFized versions of 2 procurement databases: the European Union's Tenders Electronic Daily and the Czech public procurement register. We evaluate several factors influencing matchmaking accuracy, including score aggregation and weighting, query expansion, contribution of additional features obtained from linked data, data quality and volume.

Key words: public procurement, matchmaking, linked open data, SPARQL

1 Introduction

Public procurement constitutes a large share of countries' economy. For example, the financial volume of the public procurement market in the Czech Republic in 2013 accounted for 12.3 % of the country's gross domestic product.[3] The large volume of transactions in this domain gives rise to economies of scale, so that even minuscule improvements of public procurement processes can have a substantial impact. While releasing open data is frequently framed as a means to improve transparency of the public sector, it can also have a positive effect on its efficiency [8, p. 69], since the public sector itself is often the primary user of open data. Using open data can help streamline public sector processes [18, p. 90] and curb unnecessary expenditures [19, p. 4]. The publication of public procurement data is claimed to improve *"the quality of government investment decision-making"* [12, p. 2], as supervision enabled by access to data puts a

[3] http://www.portal-vz.cz/getmedia/8965ea38-8a96-490b-ad0f-ce4e1c0a32c9/Vyrocni-zprava-o-stavu-verejnych-zakazek-za-rok-2013.pdf

pressure on contracting authorities to follow fair and budget-wise contracting procedures. This affects not only the active waste with public resources that is often caused by corruption or clientelism. In a study of Italian public sector Bandiera et al. [3, p. 1282] observed that 83 % of inefficient spending in public procurement is due to passive waste that does not entail any benefit for the public decision-maker, and which is caused rather by a lack of skills or incentives. Releasing public procurement data also makes it possible to build applications on the data that assist contracting authorities to avoid passive waste and improve the quality of their decisions. Matchmaking public contracts to relevant suppliers is an example of such application that can contribute to better informed decisions that lead to more economically advantageous contracts.

In this paper we present an application of matchmaking in public procurement, in which calls for tenders (CFT) represent potential queries and potential suppliers are the resources to retrieve. The task is to support a contracting authority in preparing a CFT in terms of screening relevant suppliers for a future contract. Other matchmaking tasks are feasible as well, such as alerting businesses on relevant open CFTs, or helping contracting authorities fill in CFT's details based on past contracts, both being described in our earlier work in [16, p. 6]. In our case, matchmaking covers only the information phase of market transaction [21, p. 194] that corresponds to the preparation and tendering stages in public procurement life-cycle [17, p. 865], during which public bodies learn about relevant suppliers and companies learn about relevant open calls.

The presented research is a part of a larger effort originally carried out within the 'procurement linked data' use case of the EU LOD2 project.[4] The first phase of this effort involved extensive data extraction, transformation, and publication according to the *Public Contracts Ontology*.[5] The second phase exploited the linked procurement datasets for matchmaking and analytic tasks. An early version of procurement matchmaker had been embedded into a prototype tool, *Public Contract Filing Application* [24, p. 5–10], whose aim was to assist contracting authorities in preparation of new calls for tenders. Two contracting authorities helped to evaluate the tool including the matchmaking functionality [24, p. 33–36]. They found the functionality of recommending potential suppliers beneficial, in particular for contracts with the so-called restricted procurement method that allows suppliers to be directly invited by an authority. The research presented in this paper partly extends the work carried out in the LOD2 project and reflects the findings from the end-user evaluation, as regards the requirement of larger training data with better quality. We evaluate our SPARQL-based matchmaker for public procurement linked data using several factors influencing matchmaking accuracy, including score aggregation and weighing, query expansion, contribution of additional features obtained from linked data, data quality and volume. The evaluation has been performed on 2 public procurement datasets: the EU-wide register Tenders Electronic Daily and the Czech public procurement journal.

[4] `http://lod2.eu/WorkPackage/wp9a.html`

[5] `http://lov.okfn.org/dataset/lov/vocabs/pc`

2 Motivating Example

In order to illustrate the described matchmaking task we present a motivating example. The following examples describe 2 contracts using the Public Contracts Ontology. Their purpose is both to give an idea of a contract's representation in RDF as such and to demonstrate the added value of using identifiers and structured data rather than plain literals.

The query contract (QC) is actually a CFT, for which its contracting authority seeks a supplier; the matched contract (MC) is a similar contract awarded in the past. The degree of similarity between them indicates whether the supplier of MC is suitable for QC as well. Listing 1 in the RDF Turtle syntax describes the contracts using mere keywords. We can see that the descriptions share the keyword "Onions", which constitutes an exact match:

```
@prefix : <http://purl.org/procurement/public-contracts#> .

<query-contract> a :Contract ;
  :mainObject "Carrots"@en ;
  :additionalObject "Onions"@en .

<matched-contract> a :Contract ;
  :mainObject "Vegetables"@en ;
  :additionalObject "Onions"@en,
    "Root and tuber vegetables"@en ;
  :awardedTender [ :bidder <matched-bidder> ] .
```

Listing 1: Keyword-based descriptions

Now, in Listing 2, we switch from keyword-based to concept-based descriptions of contracts, using the SKOS version of the Common Procurement Vocabulary (CPV, see Section 3). In the CPV taxonomy the concept of "Carrots" is narrower for a concept that is narrower for "Root and tuber vegetables", which is, in turn, narrower for "Vegetables" (see the bottom of the listing).

```
@prefix : <http://purl.org/procurement/public-contracts#> .
@prefix cpv: <http://linked.opendata.cz/resource/cpv-2008/
    concept/> .
@prefix skos: <http://www.w3.org/2004/02/skos/core#> .

<query-contract> a :Contract ;
  :mainObject cpv:03221112 ;
  :additionalObject cpv:03221113 .

<matched-contract> a :Contract ;
  :mainObject cpv:03221000 ;
  :additionalObject cpv:03221113, cpv:03221100 ;
  :awardedTender [ :bidder <matched-bidder> ] .

cpv:03221000 skos:prefLabel "Vegetables"@en .
cpv:03221100 skos:prefLabel "Root and tuber vegetables"@en .
cpv:03221112 skos:prefLabel "Carrots"@en .
cpv:03221113 skos:prefLabel "Onions"@en .

cpv:03221112 skos:broaderTransitive [
    skos:broaderTransitive cpv:03221100 ] .
```

Listing 2: Concept-based descriptions

QC and MC now become connected in 3 different ways:

- Additional object of QC ("Onions") is additional object of MC (as in the keyword approach)
- Main object of QC ("Carrots") is narrower (by 2 hops) of additional object of MC ("Root and tuber vegetables")
- Main object of QC ("Carrots") is narrower (by 3 hops) of main object of MC ("Vegetables")

Even in this simplified example, the similarity between the contracts, and, indirectly, the relevance of MC's supplier to QC, can take into account multiple inputs: taxonomic distances, relative importance of main vs. additional object, as well as the number of different connections. We will explain more details on the actual matchmaking method in Section 4.

3 Experimental Datasets

In this section we briefly describe 3 kinds of RDF resources we used, in turn. The first is the datasets describing public contracts proper. The second is the CPV dataset, which supplies the taxonomic structure for contract objects. The third, the zIndex dataset, contains a 'fairness score' of contracting authorities.

Public Procurement Linked Data. The presented matchmaker has been evaluated on 2 procurement datasets, including Czech (CZ) public procurement journal and the EU-wide register Tenders Electronic Daily (TED). These datasets expose public procurement notices informing about contracts above the financial thresholds for mandatory disclosure. The extraction and transformation of the selected datasets to linked data is described in [23, pp. 18–20] and [17, pp. 869–871]. Both are described using the Public Contracts Ontology, which contributes to the matchmaker's portability. We chose the dataset obtained from the Czech Public Procurement Journal as primary for the matchmaking experimentation because of its features (e.g., CPV codes or distinctions of contract lots) and an appropriate size. In Table 1 we provide a basic summary of these datasets.

Common Procurement Vocabulary. The most important linked dataset for the matchmaker is the Common Procurement Vocabulary (CPV):[6] a controlled vocabulary standardized by the EU for harmonizing the description of procured objects across the member states. The vocabulary has a mono-hierarchical structure in which the individual taxonomic links typically have the flavour of either subsumption[7] or part-whole[8] relations between CPV concepts. While the original CPV source expresses hierachical relations using the structure of numerical notations of the vocabulary's concepts, we transformed the CPV to RDF[9]

[6] `http://simap.europa.eu/codes-and-nomenclatures/codes-cpv/codes-cpv_en.htm`

[7] E.g., "Broccoli" has broader concept "Vegetables".

[8] E.g., "Vegetables" has broader concept "Vegetables, fruits and nuts".

[9] `https://github.com/opendatacz/cpv2rdf`

Table 1: Basic statistics of the datasets used in evaluation

	CZ	TED
Number of triples	11.7 M	46.6 M
Number of awarded contracts	73.9 k	173.3 k
Number of bidders	29 k	517 k
Temporal coverage	July 2006–August 2014	2012–June 2014
Average number of contracts won by bidder	3.82	1.27

that makes these relations explicit using the SKOS vocabulary. *Public Contracts Ontology* [14, p. 21] declares CPV concepts as range of the properties `pc:mainObject`[10] and `pc:additionalObject`, both of which are defined as subproperties of `dcterms:subject`. The property `pc:mainObject`, defined as functional, indicates the most important object procured in the contract, whereas `pc:additionalObject` describes supplementary objects related to the contract. Using the terminology of case-based reasoning, CPV provides a "bridge attribute" that allows to derive the similarity of contracts from the shared concepts in their descriptions. CPV is available in both of the used datasets.

Fairness Index. For the Czech dataset only we were able to use annotations with zIndex.[11] It is a rating of fairness of Czech contracting authorities computed from several indicators describing various issues in public procurement, such as the use of open tendering procedures. It is represented as a number z $(0 < z \leq 1)$, where 1 is assigned to contracting authorities that best adhere to the fair practices in public procurement, while the index of 0 is given to authorities that deviate from these practices the most. The zIndex is available for 63.1 % of contracting authorities present in the CZ dataset; for others we use 0.5 as a default value. There is no fairness index for the TED dataset.

4 Matchmaking Problem and Method

Matchmaking can be defined as an information retrieval task of searching the space of queries (demands) and resources (supplies), both of which are described using semi-structured data with comparable schemas, and the task results are ordered by the degree to which they fulfill the query [7, p. 278]. The presented matchmaker can be described as a case-based reasoning recommender system that provides a *"form of content-based recommendation that emphasizes the use*

[10] All vocabulary prefixes used in this paper can be resolved to their corresponding namespace URIs via `http://prefix.cc`.

[11] `http://wiki.zindex.cz/doku.php?id=en:start`

of structured representations and similarity-based retrieval during recommendation" [22, p. 369]. It recasts public contracts awarded in the past as cases to learn from. In this sense, the awarded contracts represent experiences of solved problems [20, p. 17]. The downside of this approach is that it favors larger and longer-established bidders that were awarded with more contracts than newcomers to the procurement market. The matchmaker produces a ranked list of top-k most relevant matched resources for a given query resource. Fig. 1 depicts a matchmaking scenario, with the resources to be matched in bold:

- The query resource is a *call for tenders* (CFT), marked as QC (for 'query CFT'), i.e. an incomplete contract object with no awarded tender.
- The matched resources are potential bidders, as business entities that may supply what the contract demands. The diagram represents them by MB_1 and MB_2 (for 'matched business entities').
- We only consider the business entities awarded with at least 1 contract. These contracts, through which the matchmaking is carried out ($MC_1 - MC_3$), are analyzed with respect to the similarity to the current CFT.
- The contract-CFT similarity calculation relies primarily on the similarity of their *objects*, in terms of CPV concepts linked by the `pc:mainObject` (MO) and `pc:additionalObject` (AO) property.
- The current contracting authority (CA) preparing the CFT (QCA, for 'query CA') is only implicit in the process. However, we consider the CAs responsible for the relevant past contracts (such as MCA_1, for 'matched CA', in the diagram), in particular in terms of their *fairness scores* (as FS_1). The better the fairness score of a CA, the higher the contribution of the contract awarded by it to the recommendation of the awarded supplier.

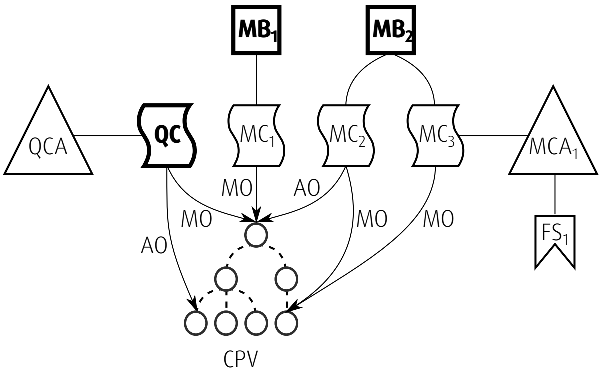

Fig. 1: Overall diagram of similarity-based contract matchmaking

4.1 Generic Method

We present a generic descriptive formalism of matching contracts (precisely, contract notices) to potential bidders, and gradually refine it to cover the notions specific to our approach. We start with the notions of contract and call for tenders. A public *contract* is a quadruple

$$c = (ca, su, obj, ctx)$$

where ca is a contracting authority (buyer), su is the supplier, obj is the object of the contract (what is provided by the supplier to the buyer) and ctx is the context of the contract, including, for simplicity, all of its aspects other than buyer, supplier and object: the conditions under which the contract is fulfilled (time, place, legal framework, etc.), but also the tendering procedure and the inherent characteristics of the contract, e.g., whether it is a standalone contract or just a part of contract, called 'lot'. Analogously, a *call for tenders* is a triple

$$cft = (ca, obj, ctx)$$

where ca is a contracting authority (potential buyer), obj is the demanded object of the anticipated contract (for which the tenders are to be submitted) and ctx is the context of the contract envisaged in the call (with similar but possibly reduced scope as for an awarded contract).

Let C be the *contract base*, i.e. the pool of all known contracts (with all possible contracting authorities) considered in the task, S is the (known) *supplier base*, i.e. the set of all past suppliers of all contracts from C, and CFT is the set of calls for tenders under consideration. From the practical point of view, it should consist of the calls of the given contracting authority for which the matchmaking task is to be carried out, namely, the *published calls* (not yet awarded to a bidder) and the *calls under preparation*. The task of matchmaking a call for tenders to potential suppliers is then cast as computation of match score from some ordinal range of values V, allowing subsequent ranking. It can be expressed by means of the following match-score abstract function:

$$mscore : CFT \times S \rightarrow V$$

Let $cft_q = (ca_q, obj_q)$ be the 'query' call for tenders for which we want to retrieve potential suppliers. In our particular model, formed by the actual data sources available, we primarily exploit the similarity at the level of contracts (call for tenders being just a reduced form of implemented contract). Therefore we introduce the notion of *contract-similarity match-score* function

$$mscore_{CS}(cft_q, su) = \underset{i}{\boldsymbol{agg}^c}(\boldsymbol{adj}(\boldsymbol{sim^c}(cft_q, c_i), \boldsymbol{imp}^c, \boldsymbol{qual}(ca_i)))$$

where

- $\boldsymbol{imp}^c$ is a function returning the importance of a contract, typically based on its type (e.g., a lot is less important than a complete contract)

- $\boldsymbol{qual}$ is a function returning the contracting authority quality, such as a 'fairness/transparency index' of the authority; it is uniformly computed for each contract of the same authority ca_i
- $\boldsymbol{sim}^c$ is a contract similarity metric, presumably a symmetric one
- $\boldsymbol{adj}$ is a contract-level score adjustment operator over the values of $\boldsymbol{sim}^c$, by $\boldsymbol{imp}^c$ and $\boldsymbol{qual}$
- $\underset{i}{\boldsymbol{agg}^c}$ is a score (adjusted by $\boldsymbol{adj}$) aggregation operator over all contracts $c_i = (ca_i, su, obj_i)$, i.e. those of supplier su.

Of the boldfaced template operators or metrics, $\boldsymbol{sim}^c$ is elaborated on below. For all, specific instantiations are provided in Section 4.2.

The contract similarity metric $\boldsymbol{sim}^c$ can be based on both components we distinguish for a contract: object and context. In our approach we currently only exploit the contract object, since the availability of machine-readable data is limited in the sources we address. We thus specialize the similarity metric to

$$\boldsymbol{sim}^c(cft_q, c_i) = \boldsymbol{sim}^{obj}(\boldsymbol{exp}(obj_q), obj_i)$$

where $\boldsymbol{sim}^{obj}$ is a contract *object* similarity metric, and $\boldsymbol{exp}$ is a query expansion operator for contract objects.

Let us now define more precisely the notion of *object* of a contract. Rather than as 'physical' products or services, the objects of public procurement are mostly observed at the level of concepts, declared in controlled vocabularies such as CPV, some of which may play dominant and some other only marginal role in the same contract. Therefore we decompose obj_i (as object in contract c_i), in our model, to a multiset of *concept associations*

$$obj_i = \{(con_{i,j}, str_{i,k}))\}$$

where all concepts $con_{i,j}$ belong to a common set of concepts Con and all $str_{i,k}$ are numerical values from a certain range, presumably $0 \leq str_{i,k} \leq 1$. Each $(con_{i,j}, str_{i,k})$ pair represents an association between contract c_i and a concept, equipped with specified numerical strength. Since c_i can be associated to the same concept in multiple ways with different or equal strengths, the same $con_{i,j}$ as well as $str_{i,k}$ may repeat in obj_i, giving the possibility of multiset rather than set.

We assume that the set Con is internally (especially, hierarchically) structured such that relevance of one concept may entail relevance of some other concepts in its neighborhood. The expansion operator $\boldsymbol{exp}$ takes as input a contract object obj and returns a contract object $\boldsymbol{exp}(obj)$. At the general level, we will only require $\boldsymbol{exp}$ to be monotonous with respect to the concept association multiset, i.e. for every concept $con \in Con$ holds

$$(con, str) \in obj \quad \Rightarrow \quad (con, str) \in \boldsymbol{exp}(obj)$$

As regards the similarity of two contract objects, we assume that it is to be computed at the level of concepts shared by their concept sets, taking into account the respective *strengths* of the concepts in the objects and also the prior *importance* of the concepts, obtained via the function $\boldsymbol{imp}^{con}$. The importance can be based on the concept's statistical discrimination power (which is what we exploit in our method described later), on its position in the graph structure of Con, on its assessment by a human oracle, etc. Formally,

$$\boldsymbol{sim}^{obj}(obj_q, obj_i) = \underset{j}{\boldsymbol{agg}^{con}}(\boldsymbol{comb}(str_{q,j}, str_{i,j}, \boldsymbol{imp}^{con}(con_j)))$$

for all j such that

$$(con_j, str_{q,j}) \in obj_q \ \wedge \ (con_j, str_{i,j}) \in obj_i$$

Here, $\boldsymbol{comb}$ is a function for combining the 3 weights into a partial similarity measure (in terms of one shared concept). Note that the 3 weights correspond to the 'strengths' of the 3 parts of the *path* connecting the query call for tenders cft_q and the matched contract c_i: edge from cft_q to con_j, edge from c_i to con_j, and concept con_j in the middle. Furthermore, $\boldsymbol{agg}^{con}$ is another aggregation operator, this time (in contrast to $\boldsymbol{agg}^{c}$) over the multiple shared concepts of contracts rather than over contracts themselves.

4.2 Function Instantiation

The boldfaced abstract functions from the previous subsection can be instantiated in many ways. The summary of our particular instantiation (aside $\boldsymbol{sim}^{c}$, already instantiated above) and of the choice/range of parameter values in the experiments, is in Table 2. Note that 'modifiers' (such as quality-based adjustment) have only been used in some of the experimental settings. The nature of these inferential procedures can be understood as reasoning under uncertainty [10]. Since $\boldsymbol{adj}$ and $\boldsymbol{comb}$ aggregate multiple 'sequential' weights over an inferential path, they are 'fuzzy-conjunctive' and can be modeled as *t-norms*. Analogously, $\boldsymbol{agg}^{c}$ and $\boldsymbol{agg}^{con}$ aggregate multiple 'parallel' weights affecting an inferential node, they are thus 'fuzzy-disjunctive' and can be viewed as *t-conorms*. Product ($a * b$) and probabilistic sum ($a + b - a * b$) are the most commonly used types of t-norm and t-conorm, respectively. However, probabilistic sum requires aggregation by multiplication, which cannot be implemented directly in SPARQL since it lacks an operator to multiply grouped bindings. Therefore, we implemented the aggregation via post-processing of SPARQL results. Eventually, since the difference on the evaluated metrics between the probabilistic sum and summation ($a+b$) turned out to be statistically insignificant, we opted for summation, which can be computed in SPARQL and is marginally faster. We also experimented with alternative t-norms and t-conorms: Gödel's and Łukasiewicz's methods [4, p. 27]. Their impact on the target metrics was however negative, as seen in the evaluation results in Section 7; we thus omit them from Table 2, for simplicity.

One more instantiation feature is the setting of association strength (str). While we always set it to 1 for the main object of the contract, it was lowered, in some configurations, for the additional object ('AO weight' in result tables).

Function	Instantiation	Value/range
$\boldsymbol{imp}^c$	Two-valued function	1 for complete contract, 0.5 for lot
$\boldsymbol{qual}$	For Czech contracts only: *Z-index* of the authority	$(0, 1]$
$\boldsymbol{adj}$	Multiplication	
$\boldsymbol{agg}^c$	Summation	
$\boldsymbol{exp}$	Addition of broader/narrower concepts n hops away from the current concept	n varying from 1 to 3; strength same as for current
$\boldsymbol{imp}^{con}$	IDF measure wrt. use in contracts	$(0, 1]$
$\boldsymbol{comb}$	Multiplication	
$\boldsymbol{agg}^{con}$	Summation	

Table 2: Instantiations of abstract functions used in experiments

5 Implementation

Matchmaking public contract to suitable bidders starts with retrieving similar contracts awarded in the past. For each awarded contract a similarity score is computed and the contracts are grouped by bidders that won them. Scores of each group are aggregated and sorted in descending order. In this way, matchmaking uses both semantic and statistical properties of data on which it operates. While the semantics of contracts' descriptions is employed in similarity measurement, the aggregation of scores reflects the statistics about past participation of bidders in public procurement [2, p. 122].

The matchmaker is implemented using SPARQL [1] as a native way of RDF data processing. Indeed, as it operates directly on the RDF data model, there is no need for data re-formatting. In this way, the matchmaker avoids the initialization time needed for pre-processing data or training a model. Given that it needs only RDF store's indices for operation, it is well suited for streaming data that requires real-time processing. Such requirement is to a certain degree also present in the public procurement domain because its data becomes quickly obsolete due to its currency bound on fixed deadlines for tender submission. Moreover, since the matchmaker is limited to the standard SPARQL without proprietary addons or extension functions, it is portable across RDF stores compliant with the SPARQL specification. The implementation of our matchmaker bases exclusively on querying a SPARQL endpoint without any previous data preprocessing. Our tool can thus be easily deployed by any public administration exposing its data via a SPARQL endpoint with no further tool or service needed.

While RDF stores in general suffer from performance penalty compared to relational databases, recent advancements in the application of column store

```
?queryCFT (pc:mainObject|pc:additionalObject)/
  (skos:broaderTransitive|skos:narrowerTransitive)*/
  ^(pc:mainObject|pc:additionalObject)/
  pc:awardedTender/pc:bidder ?matchedBidder .
```

Listing 3: Matchmaker's SPARQL property path

technology for RDF data improved things a lot [5, p. 23]. Yet, in order to get the best performance of SPARQL, the matchmaker is limited to exact joins. Fuzzy joins over literal ranges or overlapping substrings significantly decrease the matchmaker's performance and are therefore avoided.

The basic graph pattern considered in most configurations of the matchmaker is illustrated in Listing 3 using the SPARQL 1.1 Property Path syntax.

The actual implementation of the matchmaker in SPARQL is based on nested sub-queries and `VALUES` clauses used to associate the considered properties with weights. Score aggregation is done using SPARQL 1.1 aggregates. More detailed description of the matchmaker's implementation and the API it exposes is available in [16, p. 5]. Matchmaker source code is available in a public repository[12] licensed as open source under the terms of Eclipse Public License. Example SPARQL queries used by the matchmaker can be found at `https://github.com/opendatacz/matchmaker/wiki/SPARQL-query-examples`. A demo instance of the matchmaker's JSON-LD API configured for the data from TED is available at `http://lod2-dev.vse.cz:8080/matchmaker/`.

6 Evaluation Protocol

We evaluated the matchmaker's configurations on the task of predicting the contracts' winning bidders. In our setting the matchmaker predicts from the space of all bidders and not only from those that bid for a given contract, since only the identity of the awarder bidder is available in the datasets used for evaluation. In this context, winning bidders are used as ground truth and the matchmaker attempts to mimick the selection of contracting authorities. However, it is clear that the awarded bidder may not be the best match in all cases. The choice of the winner may not only reflect its suitability for a particular contract, but it can also be influenced by adverse factors including corruption, lack of information or cartel agreements. Attempting to address these factors we included zIndex as a weight in one of the matchmaker's configurations in order to take into account how well contracting authorities adhere to fair contract award procedures. The evaluation was done using offline experimental setup.

The datasets on which we evaluated the matchmaker are described in Section 3. Our evaluation setup was restricted to public contracts that announced 1 winner. However, some contracts in the considered datasets reported more that

[12] `https://github.com/opendatacz/matchmaker`

a single winner, typically because their descriptions did not distinguish between lots the contracts were composed of, so it was not possible to assign winners to the lots they were awarded. While in the Czech dataset this was the case for only a handful of contracts, in TED it accounted for 9.8 % of awarded contracts. Due to our evaluation setup we removed these contracts from the datasets used for experimentation. The datasets' statistics in Table 1 reflect this change.

We used 5-fold cross validation on the complete datasets without respecting the temporal order of contracts (i.e.we also use future contracts for matching the query contract). The folds were not overlapping, so each contract was evaluated exactly once. The results obtained for the adopted metrics were averaged over individual folds. Evaluation of statistical significance was done using the Student's t-test.

6.1 Metrics

The chosen evaluation metrics reflect the matchmaker's accuracy and variance in results. We adopted Hit-Rate at 10 (HR@10) [6, p. 159] as our principal metric. This metric is defined as $\frac{number\ of\ hits}{n}$, where hits account for contracts for which the awarded bidder is found in the first 10 results of the matchmaker and n is the total number of the evaluated contracts. We prioritized this metric because the first 10 results are typically the only ones users consider.[13]

A complementary metric employed in the evaluation was mean Average Rank at 100 (AR@100). This metric is defined as $\frac{1}{n}\sum_{i=1}^{n} r_i$, where r_i $(1 \leq r_i \leq 100)$ is the rank of the awarded bidder in the first 100 results of the matchmaker and n is the number of cases when the awarded contract was found in the first 100 results of the matchmaker. Note that AR cannot be used alone, since it does not penalize complete non-matches and matches below the threshold rank. AR is thus meant as a fine-grained adjustment for distinguishing between cases that all have rather high HR, which, in turn, does not account for the actual rank at all. The threshold for AR@100 was set as more relaxed as that for HR@10 such that 'less visible but not yet hopeless' matches (considering a more patient user) would only lose their weight gradually.

Apart from metrics of accuracy we consider catalog coverage [9, p. 258], which measures variance of the matchmaker's results. Catalog coverage is ratio of distinct items that are effectively presented in matchmaking results I_r over all items I. If we denote the number of queries as N and the items presented for a query j as I_r^j, then catalog coverage can be computed as:

$$\frac{\left|\bigcup_{j=1\ldots N} I_r^j\right|}{|I|}$$

Using the described evaluation protocol we conducted several experiments with different configurations of the matchmaker. We discuss the results of these experiments in the following section.

[13] A study claims that this is the only part considered by 91 % of search engine users (`http://www.seo-takeover.com/case-study-click-through-rate-google/`).

7 Comparison of the Evaluated Approaches

From the large pool of possible configurations of the matchmaker we only selected a few for evaluation of the metrics. These configurations either differ in the matchmaking method or produce the best result improvement. The following tables summarize the results obtained.

7.1 Evaluation Results

The results are presented for different matchmaker settings and varying additional object weight (AO weight). Tables 3 and 4 refer to the CZ and TED dataset, respectively. The number of configurations is higher for CZ due to its smaller size, allowing to effectively explore the space of configurations, and also due to availability of fairness index (zIndex) annotations for it; for TED we also did not perform additional deduplication. The best results for each metric are in bold font. All differences in evaluation results that we report further on as statistically significant were tested for p-values < 0.01. Large size of the evaluation datasets allows us to recognize even minor but statistically significant differences.

Table 3: Comparison of selected matchmaker's configurations on the CZ dataset

Matchmaker	AO weight	HR@10	AR@100	CC@10
Exact CPV	1	0.234	19.046	0.304
Exact CPV	0.1	0.237	18.564	0.333
Exact, Gödel comb.	0.5	0.084	32.15	0.324
Exact, Łukasiewicz comb.	0.5	0.076	32.805	0.326
Exact, product comb.	0.5	0.235	19.104	0.305
Exact, distinguishing lots	1	0.23	19.399	0.3
Exact, with zIndex	1	0.233	18.867	0.307
Exact, better deduplication	1	0.27	18.354	0.309
Exact, better deduplication	0.1	**0.273**	**18.052**	**0.337**
Expand 1 hop to broader	1	0.227	20.32	0.268
Expand 1 hop to broader, with IDF	1	0.235	19.877	0.286
Expand 1 hop to narrower	1	0.234	19.044	0.304
Expand 1 hop to narrower, with IDF	1	0.236	19.278	0.298

Table 4: Comparison of selected matchmaker's configurations on TED

Matchmaker	AO weight	HR@10	AR@100	CC@10
Exact CPV	1	0.06	**19.159**	0.065
Exact CPV	0.1	**0.06**	19.423	**0.081**
Exact, Gödel comb.	0.5	0.014	36.347	0.06
Exact, Łukasiewicz comb.	0.5	0.014	36.516	0.06
Exact, product comb.	0.5	0.059	19.744	0.069
Expand 1 hop to broader	1	0.053	21.833	0.049
Expand 1 hop to broader, with IDF	1	0.058	20.193	0.057
Expand 1 hop to narrower	1	**0.06**	19.16	0.065

7.2 Discussion

We applied the matchmaker using exact matches of CPV concepts with *association strenghts* set to 1 as our baseline. When varying the additional object inhibition (AO weight), we found that the best results for the CZ dataset can be obtained by adjusting it to 0.1. With this setting the HR@10 is increased by 0.27 %, while AR@100 improves by 2.53 %. This configuration also has a positive impact on catalog coverage, which for CC@10 increases by 2.89 %. Using this setting also improved the evaluation results for the TED dataset, as can be seen in Table 4. Having a small influence of additional objects on the similarity scores surpasses ignoring the additional objects and is also better than assigning higher weights to additional objects. We experimented with using combinations of different AO weight for query CFTs and matched contracts, but the results were worse than when the weight was set to 0.1 in both cases.

When testing different *t-norms* and *t-conorms* we set the weights to 0.5 instead of 1 because so as to allow the differences in some combination methods to manifest. The product combination clearly outperformed the others, because it reflects statistical properties of data better. If we distinguish lots from complete contracts by a decreased weight, the evaluation results get worse. When we applied zIndex fairness score as a weight for contracting authorities, we did not observe any statistically significant difference in the evaluation results.

The largest improvement of the matchmaker's evaluation results was achieved by better *deduplication* of bidders in the CZ dataset. Deduplication and fusion of bidders reduces the search space of possible matches and thereby increases the probability of finding the correct match. Deduplication led to 14.97 % decrease in the number of bidders, which in turn accounted for 3.56 % improvement in HR@10, 3.63 % improvement in AR@10, and 0.46 % improvement in CC@10. These results indicate that influence of data quality is greater than the impact of the parameters varied in the evaluation.

Query expansion did not improve the matchmaker's performance. We experimented with 1-3 hops for expansion to broader and narrower concepts. Whether we applied IDF as a weight for the expanded concepts or not, the results did not improve significantly and often even got worse. Since we did not obtain any gain by query expansion, its computational overhead clearly does not pay off.

Apart from varying single parameters, we evaluated a couple of *combinations* of the best-performing parameter settings. When we combined better deduplicated dataset with using 0.1 as AO weight, we got the best HR@10, which surpassed the baseline by 3.93 %, AR@10 improved by 5.22 %, and CC@10 outperformed the baseline by 3.3 %. In case this configuration was combined with 1-hop expansion to broader concepts the evaluation results got worse.

Overall, the results for the TED dataset were worse than for the CZ dataset, which is likely due to greater heterogeneity and duplicity. In effect, these results are barely usable in practice.

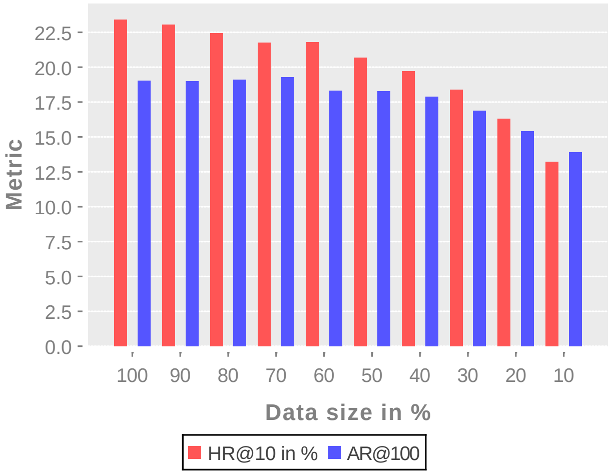

Fig. 2: Impact of data reduction

In order to assess the impact of data volume on the matchmaker's results we performed evaluation with reduced sample dataset. Before each experiment we temporarily removed a part of the dataset containing the links between public contracts and awarded bidders and run the 5-fold cross validation on the reduced dataset. The results show that if we reduce the Czech dataset by half, HR@10 drops by 2.73 % for the baseline configuration. If we reduce the dataset to 10 % of the original size, HR@10 decreases by 10.16 % compared to the baseline. It is worth noticing that sometimes while HR@10 decreases, AR@100 improves, highlighting that these metrics must be evaluated together. The impact of data reduction is shown in Figure 2. This demonstrates how the matchmaker leverages the volume of data using aggregations along with data semantics.

8 Related Work

In order to illustrate the progress beyond the state of the art made by the presented matchmaker, we compare it to the features provided by the publishers of public procurement data, reusers of this data, and the related research in matchmaking in general.

The search interfaces provided by the publishers of datasets employed in this paper can be used to approximate the matchmaker's functionality, although doing so requires additional manual work for the tasks that cannot be automated via these interfaces. Besides simple search, TED allows to search its archives using the Common Command Language [11]. For instance, following the example matched contract from Section 2, we can express a query for awarded contracts described by the CPV concept "Root and tuber vegetables", its narrower concepts obtained by query expansing using a wildcard, or the concept "Onions" as `TD=["Contract award"] AND PC=[032211* OR 03221113]`. This query retrieves a list of contract award notices, which a user needs to go through individually to find the awarded bidders. Aggregation of search results is possible via a statistic mode, however, awarded bidder is not among the fields users are allowed to group the results by. Approximating the matchmaker's results with this interface thus requires additional manual work. Similar matchmaking functionality may be achieved by searching XML dumps of TED data,[14] but combining it with additional data and writing expressive queries can be difficult.

The Czech public procurement journal provides a search interface that allows to query multiple fields and provides full-text search features including Boolean operators and wildcards. The expressivity of this interface allows to perform queries analogous to those for TED. The journal also exposes an XML API,[15] providing machine-readable data. However, as is the case for TED, data aggregation and combination with other sources may require a lot of effort.

Given the availability of these datasets as open data, their commercial reusers can build services similar to our matchmaker. An example reuser of TED data is Euroalert.net [15], which provides alert services matching search profiles of companies to relevant contracts. However, as the description of this service[16] suggests, the matchmaking is keyword-based without using the semantics of contracts' descriptions. For the Czech public contracts, the maintainer of the official procurement journal provides the portal Zakázky+,[17] which offers an analogous alert service using cleaner and better deduplicated data.

If we survey the related research in matchmaking, the closest work to ours is likely from Alvarez-Rodríguez et al. [2], who used SPARQL for matchmaking organizations and public procurement notices along with several methods of query expansion [2, p. 118]. Unfortunately, it is difficult to determine the differences between their and our approach, because neither implementation details

[14] http://ted.europa.eu/TED/misc/xmlPackagesDownload.do

[15] http://vestnikverejnychzakazek.cz/en/PublishAForm/XMLInterfaceForISVZUS

[16] http://euroalert.net/en/10ders_alerts_government_contracts.aspx

[17] http://www.zakazky-plus.cz/

nor evaluation is revealed in the paper describing this work. SPARQL also served as a basis for extensions focused on similarity retrieval in iSPARQL by Kiefer el al. [13]. Unlike iSPARQL, our approach works without extending SPARQL and therefore maintains better compatibility.

9 Conclusions

Public procurement is an area where the benefits of linked data technology are potentially manifold and affect numerous parties: contracting authorities, suppliers, citizens as well as supervisory bodies. In the paper we focused on the contracting authority side, namely, on the exploitation of past contracts data when creating new CFTs and assessing which business entities could potentially become suppliers; the whole task is recast as CFT-to-supplier matchmaking mainly leveraging CFT-to-contract similarity. Based on prior feedback from contracting authorities, we carried out a number of experiments on 2 procurement datasets. In summary, the influence of data volume and quality in terms of better deduplication appears greater than the impact of the parameters varied in the evaluation. Future work will also address systematic exploitation of textual description associated with the contracts, as well as geo-spatial information, such as location of the suppliers, and recency of the past contracts.

Acknowledgements: The presented research has been partially supported by the EU ICT FP7 project no. 257943, LOD2 project, by the H2020 project no. 645833 (OpenBudgets.eu), by the Italian projects PON01_03113 ERMES, PON02_00563_3470993 Vincente and PON02_00563_ 3446857 KHIRA, and by the long-term institutional support of research activities by Faculty of Informatics and Statistics, University of Economics, Prague.

References

1. SPARQL 1.1 Query Language. W3C recommendation, W3C (March 2013), http://www.w3.org/TR/sparql11-query/
2. Alvarez-Rodríguez, J.M., Labra Gayo, J.E., Ordoñez de Pablos, P.: Enabling the matchmaking of organizations and public procurement notices by means of linked open data. In: Cases on Open-Linked Data and Semantic Web Applications, pp. 105–131. IGI Global (2013)
3. Bandiera, O., Prat, A., Valletti, T.: Active and passive waste in government spending: evidence from a policy experiment. American Economic Review 99(4), 1278–1308 (2009), http://www.aeaweb.org/articles.php?doi=10.1257/aer.99.4.1278
4. Beliakov, G., Pradera, A., Calvo, T.: Aggregation functions: a guide for practitioners. Springer (2007)
5. Boncz, P., Erling, O., Pham, M.D.: Advances in large-scale RDF data management. In: Auer, S., Bryl, V., Tramp, S. (eds.) Linked open data: creating knowledge out of interlinked data, pp. 21–44. Lecture Notes in Computer Science, Springer (2014)
6. Deshpande, M., Karypis, G.: Item-based top-n recommendation algorithms. ACM Transactions on Information Systems 22(1), 143–177 (January 2004)

7. Di Noia, T., Di Sciascio, E., Donini, F.M.: Semantic matchmaking as non-monotonic reasoning: a description logic approach. Journal of Artificial Intelligence Research 29(1), 269–307 (May 2007)
8. Europe, A.I., Foundation, O.K.: Beyond access: open government data & the right to (re)use public information. Tech. rep. (2011), http://www.access-info.org/documents/Access_Docs/Advancing/Beyond_Access_7_January_2011_web.pdf
9. Ge, M., Delgado-Battenfeld, C., Jannach, D.: Beyond accuracy: evaluating recommender systems by coverage and serendipity. In: Proceedings of the Fourth ACM Conference on Recommender Systems. pp. 257–260. ACM, New York (NY) (2010)
10. Hajek, P.: The Metamathematics of Fuzzy Logic. Kluwer (1998)
11. ISO: Information and documentation: commands for interactive text searching. ISO 8777:1993, International Organization for Standardization (1993)
12. Kenny, C.: Publish what you buy: the case for routine publication of government contracts. Tech. Rep. 011, Center for Global Development, Washington DC (August 2012), http://www.cgdev.org/content/publications/detail/1426431
13. Kiefer, C., Bernstein, A., Stocker, M.: The fundamentals of iSPARQL: a virtual triple approach for similarity-based semantic web tasks. In: The semantic web, Lecture notes in computer science, vol. 4825, pp. 295–309. Springer (2007)
14. Klímek, J., Knap, T., Mynarz, J., Nečaský, M., Svátek, V.: LOD2 deliverable 9a.1.1: Framework for creating linked data in the domain of public sector contracts. Tech. rep., LOD2 EU Project, Prague (2012), http://static.lod2.eu/Deliverables/deliverable-9a.1.1.pdf
15. Marín, J.L., et al.: Euroalert.net: aggregating public procurement data to deliver commercial services to SMEs, pp. 114–130. IGI (2013), http://www.igi-global.com/chapter/euroalert-net-aggregating-public-procurement/69590
16. Mynarz, J., Zeman, V., Dudáš, M.: LOD2 deliverable 9a.2.2: Stable implementation of matching functionality into web application for filing public contracts. Deliverable D9a.2.2, LOD2 EU Project (2014), http://svn.aksw.org/lod2/D9a.2.2/public.pdf
17. Nečaský, M., Klímek, J., Mynarz, J., Knap, T., Svátek, V., Stárka, J.: Linked data support for filing public contracts. Computers in Industry 65(5), 862–877 (2014), special Issue: New trends on E-Procurement applying Semantic Technologies
18. Parycek, P., Höchtl, J., Ginner, M.: Open government data implementation evaluation. Journal of Theoretical and Applied Electronic Commerce Research 9(2), 80–99 (May 2014), http://www.scielo.cl/pdf/jtaer/v9n2/art07.pdf
19. Prešern, M., Žejn, G.: Supervizor: an indispensable open government application. In: Share-PSI 2.0 workshop on uses of open data within government for innovation and efficiency (2014), https://www.w3.org/2013/share-psi/wiki/images/6/6b/Supervizor_Slovenia_description_pdf.pdf
20. Richter, M.M., Weber, R.O.: Case-based reasoning: a textbook. Springer (2013)
21. Schmid, B.F., Lindemann, M.A.: Elements of a reference model for electronic markets. In: Proceedings of the Thirty-First Hawaii International Conference on System Sciences. vol. 4, pp. 193–201 (January 1998)
22. Smyth, B.: Case-based recommendation. In: The Adaptive Web, Lecture Notes in Computer Science, vol. 4321, pp. 342–376. Springer (2007)
23. Svátek, V., et al.: LOD2 deliverable 9a.3.1: application of data analytics methods of linked data in the domain of PSC. Deliverable D9a.3.1, LOD2 EU Project (2014), http://svn.aksw.org/lod2/D9a.3.1/public.pdf
24. Svátek, V., et al.: LOD2 deliverable 9a.3.2: Implementation of data analytics in the public contract filing application. Deliverable D9a.3.2, LOD2 EU Project (2014), http://svn.aksw.org/lod2/D9a.3.2/public.pdf