

## **Cross-lingual entity resolution and its applications**

Zinaida Apanovich<sup>1,2</sup> [0000-0002-5767-284X]

**Abstract.** Nowadays, entity resolution is being intensively investigated in the context of the integration of heterogeneous data sets. Collecting data from heterogeneous data sets and integrating them in a query able environment increases completeness and correctness as well as ensures a more effective analysis. Of special interest is the problem of cross-lingual entity resolution for multilingual data integration since local language data sources are often more complete and accurate than global data sources.

**Keywords:** multilingual knowledge bases, integration, cross-lingual entity resolution, clustering, interactive visualization, matrix ordering.

### **1. Introduction**

Nowadays, entity resolution is being intensively investigated in the context of the integration of heterogeneous data sets. Collecting data from heterogeneous data sets and integrating them in a query able environment increases completeness and correctness as well as ensures a more effective analysis. Of special interest is the problem of cross-lingual entity resolution for multilingual data integration since local language data sources are often more complete and accurate than global data sources.

Although English is the main language for research and the Internet, a great number of research publications belong to non-English authors and are translated from various foreign languages, which makes the task of integrating multiple data sources even more difficult. Naturally, this poses the problem of the cross-language disambiguation of named entities and, in particular, the cross-language matching of authors and publications [1, 2].

Our previous research has demonstrated that Russian names allowing several transliterations represent a challenge for international data bases and knowledge graphs. Experiments with several multilingual datasets have shown that Russian names admitting several transliterations are often treated as homonyms, and several different persons with identical name variations are treated as synonyms [3, 4]. This is especially annoying when errors occur in the resources calculating scientific ratings, such as Scopus and Web of Science.

For example, the papers by Виктор Карлович Сабельфельд (Viktor Karlovich Sabelfeld) are assigned in the Scopus data base [5] to three people with distinct Scopus identifiers and distinct lists of publications. Meanwhile, if we go to the

Russian site eLIBRARY.ru [6], we will realize that Victor K. Sabelfeld and V.K. Sabel'fel'd from the A.P. Ershov Institute of Informatics Systems, as well as Viktor Sabelfeld from the Karlsruhe Institute of Technology, are in fact the same person.

Thanks to the projects such as ORCID [7] and to the continual interaction of researchers with the of Web of Science and Scopus developers [8], the quality of data presented at these sites has improved. However, apart from them, there is a great number of other scholarly data sources, and it is impossible to verify all of them manually. For example, the Springer Nature SciGraph [9], a Linked Open Data platform, collecting information about conferences, publications, affiliations, and research projects, has the same problem with representing information on researchers having Russian names. There are two persons, named Victor Sabelfeld and VK Sabel'Fel'D, with distinct identifiers and distinct lists of papers. Besides, there are many papers authored by a real person, Viktor Karlovich Sabelfeld, which are not assigned to anybody in the SN SciGraph. Instead, the authors of such publications are represented by blank nodes and the properties of these blank nodes are described by means of several literal triples. For example, the property *shema:affiliation* of the blank node [https://scigraph.springernature.com/pub.10.1007/3-540-08065-1\\_5#N54dbe9edfb2343559c2d95777627d496](https://scigraph.springernature.com/pub.10.1007/3-540-08065-1_5#N54dbe9edfb2343559c2d95777627d496) is described by another blank node having the type *schema:Organization* and the name "Computing Center, Nowosibirsk, USSR," the organization that ceased to exist about thirty years ago. This example demonstrates another global problem of most international data bases: data incompleteness with regard to Russian organizations. The SN SciGraph uses the GRID data base [10] as the source of general information about research organizations. GRID contains information on 2,033 Russian research organizations while the Russian eLIBRARY.ru data set contains a list of more than 11,000 research organizations. For example, the GRID data base has no information about the A.P. Ershov Institute of Informatics Systems. Consequently, all the researchers of this institute are not affiliated in the SN SciGraph, which creates additional problems with the identification of the publications authors.

Since similar situations are quite common; our main objective is to investigate various methods to enrich and improve an English-language data source(s) by comparing its (their) content with national data sources in local languages, such as the Russian-language data sources.

An algorithm for the cross-language identity resolution using the SBRAS Open Archive is presented in [3]. The algorithm relied heavily on the information about Siberian researchers and their affiliations and for this reason had a very limited application. A possible solution to this problem would be using a larger data source such as eLIBRARY.ru.

To this end, a way of establishing correspondence between the Russian-named and English-named entities has to be developed. The transliteration-based matching of personal names was already described in [3]; our new algorithm, however,

has an additional matching step, enabling us to create groups of confirmed papers for an individual researcher. Another issue is establishing the correspondence between the titles of original Russian papers and their English translations as well as between journal titles in Russian language and their English equivalents. Due to this extended matching step, the new clustering algorithm for matching authors and publications has proven to be more efficient than the previous one. Finally, an interactive visualization algorithm provides comprehensible matching and clustering results and enables their analysis and modification. In particular, interactive visualization of similarity matrices based on different similarity measures has shown that the main source of improving the quality of the matching and clustering algorithm is extending the set of confirmed publications.

The paper is organized as follows: first, we outline the related works and then present the essential datasets and metadata. After that, the matching and clustering algorithm and implementation details are described. Finally, we demonstrate an interactive visualization, which facilitates the comprehension of the matching and clustering results and allows users to improve them.

## 2. Related works

Entity resolution – identity resolution, deduplication, record linkage, object matching –refers to the task of identifying different representations of the same real-world object. In the context of relational data bases, the term *entity reconciliation* has been used for a long time. Another related term is *link discovery*, which describes the process of automatic search for the overlapping parts of heterogeneous data sets and linking individual records of these data sets by exploiting their specific properties. When talking about entities from multiple sources, it is common to use the term *entity clustering*. There are several groups of works closely related to the topic of our paper.

Entity resolution is an important step in any data integration pipeline as well as in the area of information retrieval and question answering. Nowadays, data from any domain are available on the Web and a lot of research has been dedicated to this problem. A recent overview is presented in [11]. A semantic data integration technique exploiting the semantics encoded in the properties of the entities collected from the Web data sources is implemented in the FuhSeh query engine [12]. FuhSeh receives keyword based queries and produces knowledge graphs on demand at query time, making use of wrappers around the original data sources to generate RDF molecules and to merge several RDF molecules into a single one [13]. For more than two sources, a binary linking of entities is not sufficient: all matches of the same entity should be clustered together to derive a fused entity representation in the knowledge graph. Clustering is applied on a similarity graph, where entities are represented as vertices and edges link pairs of entities with a similarity above a predefined threshold. An approach to entity linking in multiple data sources is described in [14].

To avoid comparing new entities with all members of existing clusters, each cluster creates a cluster representative, which is fused from all the properties of the cluster members. The same idea is implemented as the DBpedia Global Id Management module [15], which assigns a global cluster ID to the entities linked by *owl:sameAs* links in different language editions of DBpedia. However, it is known that the number of the explicit inter-language links does not exceed 15 percent of the existing links.

Therefore, much more numerous are the studies of cross-lingual KG alignment aiming to match automatically entities in different languages from different data sources. Recently, several embedding-based approaches have been proposed for cross-lingual KG alignment, including MTransE [16] and JAPE [17].

More recent papers model the equivalent relations between entities by using the graph convolutional networks (GCN), since they are able to generate the neighborhood-aware embeddings of entities used to discover entity alignments [18].

Unlike the previous methods using entity embeddings to match entities, the paper [19] formulates the task of entity matching as a graph matching problem between the topic entity graphs. The latter approach is somewhat similar to the RDF-molecule based approach discussed in [13]. Finally, of special interest is the paper [20] describing an experience of interlinking the SN SciGraph with the English edition of DBpedia. However, none of these approaches takes into consideration different transliterations that may correspond to an entity having the same name.

### 3. Datasets and their metadata

The SpringerLink digital library [21] has been chosen as an English-language bibliographic data source mainly because of its continuously expanding set of metadata. SpringerLink is currently one of the largest digital libraries with over 13 million documents in various research fields including computer science, mathematics, life sciences, materials, philosophy, psychology, etc. It provides detailed meta-data about its publications, such as the paper title, list of authors, ISSN, authors' affiliations, publication date, venue (journal or conference title), key words, subject abstract, references, full texts in the pdf format, etc. One of the recent innovations is the "translated from" label for the papers written in foreign languages. These additional data make it possible to improve the disambiguation quality by matching the data of the original and translated paper versions.

Other important reasons for choosing this data source are open access to its data and the emergence of the related SN SciGraph Linked Open Data platform, which also demonstrates the problems mentioned above.

The eLIBRARY.ru data base is the largest bibliographic data set used for computing the scientific rating of Russian researchers. eLIBRARY.ru stores data in the fields of science, technology, medicine and education on more than thirty million publications, more than 900,000 researchers and over 13,000 organiza-

tions (including over 3,000 officially registered). The A.P. Ershov Institute of Informatics Systems of the Siberian Branch of the Russian Academy of Sciences is an organization registered at eLIBRARY.ru; it regularly inputs and updates information concerning its employees' publications. Moreover, you can find here a complete list of publications by Academician Andrei Petrovich Ershov created by Anna Andreyevna Bulyonkova.

The sets of metadata provided by eLIBRARY.ru are similar to those of SpringerLink, though access to these metadata is restricted. To be more specific, the list of the authors publications are freely available, but detailed metadata on their papers are not free. Therefore, our disambiguation algorithm is based on the data freely available at eLIBRARY.ru. Another essential difference between these two data sources is that the language of SpringerLink is English, and that of eLIBRARY.ru is Russian, even when it stores data on the English publications of Russian researchers. The main problem, hence, is how to match entities described in different languages.

#### 4. Problem formulation and algorithm to solve it

The problem to solve is formulated as follows. Given an English language bibliographic data source(s), extract all publications potentially belonging to the person(s) specified by a Russian-language keyword and divide the extracted articles into subsets  $S_1, S_2, \dots, S_n$  so that each subset of articles belongs to one real person. The steps of the matching and clustering algorithm are:

1. Given a full Russian name, a set of *extended transliterations* is generated.
2. Elements of this set are used for the keyword search of publications in the English-language data source (for example, SpringerLink digital library).
3. An extended set of potential homonyms of the person, specified by the full Russian name, is used to extract groups of publications from a Russia-language data source (for example, eLIBRARY.ru).
4. All the publications extracted from SpringerLink are matched against the eLIBRARY.ru groups of publications.
5. The papers unmatched at the previous step are further analyzed and clustered.
6. Interactive visualization of the clustering results makes it possible to analyze and to further refine them.

The extended transliteration and data extraction from the English-language source SpringerLink are described in detail in [3]. Therefore, we will describe in more detail creating the groups of confirmed publications, a clustering algorithm, and interactive visualization.

The authors of the publications extracted from SpringerLink can be either homonyms or synonyms. The entity identification algorithm should process the list of publications and determine which of their authors are synonyms and which

are homonyms. In other words, the list of publications should be clustered into the subsets  $S_1, S_2, \dots, S_n$  such that each subset of articles is authored by a single person and all his or her name variations are synonyms. The subset  $S_1$  should contain the articles authored by the person under consideration.

To this end, the list of publications  $S$  extracted from SpringerLink is matched against the lists of publications  $E$  extracted from eLIBRARY.ru. Note that the papers of eLIBRARY.ru are already clustered into the groups  $E_1, E_2, \dots, E_m$  corresponding to individual authors. Therefore, if a paper  $s_i \in S$  is recognized as identical to a paper  $e_j$  belonging to a group  $E_m$  from eLIBRARY.ru, it is assigned to a group  $S_m$ .

eLIBRARY.ru specifies persons by their full Russian name in <LastName First Name Middle Name> format, affiliation and location of the employing organization.

Note that each person can have several homonyms and "partial" homonyms, when a short form of his/her name coincides with the short form of another person's name. For example, five full homonyms, having the same name Петров Евгений Сергеевич, are described in eLIBRARY.ru along with two partial homonyms having two distinct middle names. The persons having identical full or short forms of their names can be erroneously identified as synonyms.

To prevent this kind of errors, our algorithm creates groups of confirmed eLIBRARY.ru papers for each potential homonym of a given author. Since we consider eLIBRARY.ru to be a reliable source of information about publications written by Russian authors, we use it to create confirmed groups of publications for an English data source. That is why the confirmed groups of papers are created by comparing the papers from SpringerLink and eLIBRARY.ru.

When comparing publications from the two data sources, two main possibilities are considered.

1. Publications in both data sources are described in English. The situation when a publication in eLIBRARY.ru contains a description in Russian and an English-language version of the publication is described in the "Versions" field of the publication also falls into this category.
2. The English data source contains an English description of the publication, and the Russian data source only contains a description of the Russian version.

In the first case, when the descriptions of both articles are given in English, the names of the publications and the lists of authors are compared. A paper  $s_i \in S$  is considered to be identical to a paper  $e_j \in E$  if  $Title(s_i) = Title(e_j)$  AND  $Authors(s_i) = Authors(e_j)$ . The title cannot identify a paper uniquely as some authors can have several publications with the same title. Nevertheless, the exact match of titles and author names can be considered as evidence that the papers were authored by the same person. However, some paper titles differ in SpringerLink and eLIBRARY.ru due to scanning errors. For example, the paper titled as *SCHEMATOLOGY IN A MULTI-LANGUAGE OPTIMIZER* in eLIBRARY.ru

appears as *Schematology in a MJ I/T I-language OPT imizer* in SpringerLink. In the absence of an exact match of the paper titles, both titles are stemmed by the Porter stemmer and their overlap score is calculated. If this score exceeds a threshold value, the titles are considered coinciding. The discovered matching is written in a special file for further user control.

In the second case, cross-language identification of paper and journal titles should be applied. Many Russian journals are first published in Russian and then translated in English. A typical example is the *Программирование (Programmirovanie)* journal, which is published in English as *Programming and Computer Software*. About 40% of eLIBRARY.ru older entries have only Russian description and do not have an English counterpart. These publications, however, are very important for making confirmed paper groups as large as possible. This situation raises several problems. First, it is impossible to compare papers by the title when the title of an original paper is in Russian and the title of a translated paper is in English. Besides, the original and translated papers have disjoint sets of attributes, such as venue, ISSN, publication data, page numbers, etc.

Although SpringerLink provides information about the journal titles in the Latin alphabet only, every translated paper in the database mentions its Russian original. For example, the paper by V.E. Kotov *Parallel programming with types of control* has the label “Translated from *Kibernetika*, No. 3, pp. 1–13, May-June, 1979” in SpringerLink. Moreover, the SpringerLink database provides the ISSN of the translated version. This information suffices to find the Russian version of the paper if it is available in eLIBRARY.ru. The corresponding English-language article is marked as matching and the pair of papers is saved for further processing.

The average number of papers assigned to the confirmed groups during the matching step was about 69%, while the number of erroneously attributed publications was close to zero. The main reason why the system cannot assign some papers to their authors is data sparsity. To extend the set of the identified authors of papers, a clustering algorithm was applied to the unmatched papers.

## 5. Unmatched papers clustering

The publications unmatched at the previous step are considered to be unconfirmed and should be further analyzed and clustered. The clustering algorithm for unconfirmed publications is based on comparing each unconfirmed publication with groups of confirmed publications based on various similarity metrics.

The unconfirmed publications are compared with groups of confirmed publications using the following attributes: titles of publications, lists of authors of publications, topics and keywords, dates of publication, venue of publication (journal title or conference title), and similarity of texts of publications (TF-IDF).

All the attributes are compared pairwise, which results in a number of scores, that are summarized in the final step. When calculating the similarity scores the program adheres to the following rules.

**Titles of papers similarity.** If an exact match of the paper titles A and B is found, the *title\_similarity\_score* is set to 1.0. Otherwise, the titles of the papers A and B are stemmed, and the *title\_similarity\_score* is set to the overlap ratio of their word lists.

**Co-authors similarity.** For each confirmed group of publications, a list of all its co-authors is created, and the more often an author appears in the list of co-authors for a certain group of publications, the greater is his weight. Also, for each confirmed group of publications, a list of the author's affiliations is created, which characterizes each group, and when comparing the authors of publications, not only the names of the authors, but also their affiliations are compared.

**Subjects and keywords similarity.** The *subject\_similarity\_score* and *keyword\_similarity\_score* use Jaccard Index to evaluate the overlap ratio of the respective lists.

**Date similarity.** The *date\_similarity\_score* is set to 0.1 if the timestamp difference of the papers A and B is less than five years. If the timestamps difference of the papers A and B is more than twenty five years, it is set to - 0.1.

**Venue similarity.** The *publication\_venue\_score* (i.e., conference/journal title) is set to 0.1 if there is an exact match between their titles.

**Text similarity.** *Text\_similarity\_score* is evaluated by TF\_IDF and cosin similarity measure.

The final assignment likelihood is calculated as the sum of all the above scores.

All unconfirmed publications initially obtain a group number of -1. An unconfirmed publication joins a group of confirmed publications if its similarity with a given group exceeds a threshold value. Two groups of unconfirmed publications can be merged if the similarity value of publications inside two groups exceeds the threshold value. When merging two groups, the algorithm monitors that both groups do not belong to the set of the confirmed groups. If this happens, the merging does not occur, since the confirmed groups correspond to the publications by distinct authors. The weights of all the attributes involved in the comparison, as well as the threshold value, can be adjusted at the stage of interactive visualization.

## 6. Interactive visualization for analyzing and modifying the matching and clustering results

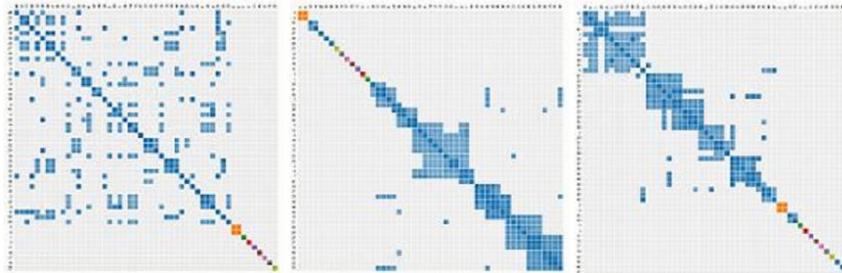
To simplify the understanding and modification of the matching and clustering results, several interlinked visualizations have been developed; they are described in detail in [4]. A global view of the obtained groups of publications is represent-

ed as a pie chart. An example of the pie chart produced by using the full Russian name *Сабельфельд Виктор Карлович* (*Sabelfeld Viktor Karlovich*) is shown in Figure 1.

Experiments have shown that to verify the matching and clustering results users need to compare papers clustered in distinct groups. To this end, they need to see all the attributes of the entire set of publications in a single view. Therefore, a set of visualizations was extended by a similarity matrix reordering module based on the `reorder.js` library [22]. The `reorder.js` library comprises several matrix re-ordering algorithms



**Figure 1.** The clustering result produced when using the full Russian name *Сабельфельд Виктор Карлович* (*Sabelfeld Viktor Karlovich*) as a keyword



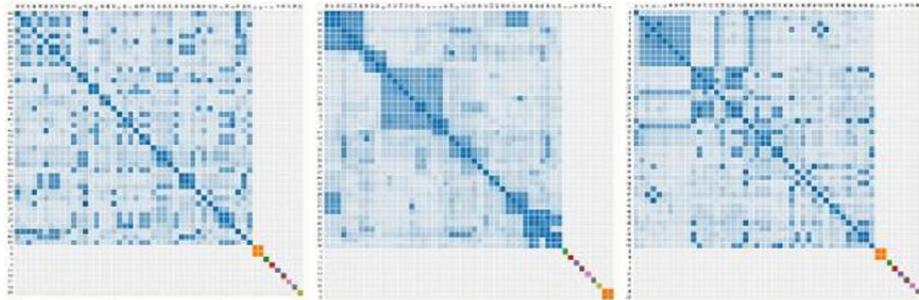
**Figure 2.** Groups of publications created by the NLP Clustering, Leaf Order and Barycenter algorithms for a similarity matrix produced by our matching and clustering algorithm

such as Barycenter heuristic, Optimal Leaf Ordering, Principal Component Analysis, Reverse Cuthill-McKee, and Spectral Ordering. An example of visualization created by the similarity matrix ordering module is shown in Figure 2. All

publications assigned to the same group by our matching and clustering algorithm have the same color. When selecting an entry of the similarity matrix with the mouse, the user can get complete information about all the attributes of the two compared publications and all the terms of their similarity score. The diagonal blocks of similarity matrices correspond to the groups of publications most similar to each other. The matrix ordering module uses the same similarity values as our clustering algorithm implemented in the matching and clustering program.

Two groups of experiments have been carried out. First, the available matrix ordering algorithms were applied to the similarity matrices which contained information about the similarity of various publications in the English source, but did not contain information about the results of the comparison of the English-language and Russian-language data sources. In this case, all ordering algorithms showed different results ( Figure 2).

Then the similarity matrices were modified by adding a large similarity score to the publications that were classified as identical at the stage of the comparison of the English-language and Russian-language data sources. After that, all the matrix ordering algorithms started to output the equivalent results ( Figure 3)



**Figure 3.** Groups of publications created by the NLP Clustering, Leaf Order and Barycenter algorithms for a similarity matrix produced by our matching and clustering algorithm

## 7. Conclusion

The newly developed matching procedure provides the algorithm presented in this paper with the ability not only to cluster the papers correctly, but also to determine the exact identity of authors, including the name and location of the affiliating organization.

The program implementing the algorithm has been tested on a dataset of 100 people employed by the IIS SB RAS at various time periods. Also, this dataset contains Academician A.P. Ershov, whose papers have been input into eLIBRARY.ru by the IIS SB RAS. The total number of papers found in SpringerLink for all Russian names in this dataset was 3,175. All the results ob-

tained by the program were verified manually. For each person listed in the test dataset the following values were calculated:

- total number of papers found in SpringerLink for each Russian full name listed in the test dataset;
- number of articles actually authored by a researcher specified in the test dataset;
- number of papers that correctly recognized by the matching algorithm;
- number of papers that correctly recognized by the matching + clustering algorithm;

Experiments have shown that the main source of improving the quality of the cross-lingual entity resolution algorithm is extending the set of confirmed publications by matching publications and authors in Russian and English data sources. The combination of the matching algorithm and the clustering algorithm allows us to correctly recognize from 92 to 93 percent of publications.

This algorithm can be used to match any multilingual knowledge bases. The further development of these studies is supposed to focus on implementing new algorithms for the cross-language identification of entities and on implementing a full-fledged framework including tools for accessing various data sources, methods for establishing correspondence between different schemes of data sources, etc.

The author thanks Avramenko M.Yu., Paramoshin A.A., Isachenko V.V., and Eliseev E.S. for participating in the implementation of the various versions of the matching and clustering algorithm.

## References

- [1] Reijnhoudt L., Costas R., Noyons E., Boerner K., Scharnhorst A. "Seed+ expand": A validated methodology for creating high quality publication oeuvres of individual researchers // Proc. ISSI . –Vienna. – 2013 .
- [2] Lawrie D., Mayfield J., McNamee P., Oard D. W. Cross-Language Person-Entity Linking from Twenty Languages. 2015.
- [3] Apanovich Z., Marchuk A. Experiments on Russian-English identity resolution // Proc. ICADL 2015 .– Lecture Notes in Comp. Sci. – 2015. – Vol. 9469
- [4] Apanovich Z., Isachenko V. Analysis and visualization algorithm for cross-language author names disambiguation. // Proc. XX International Conference "Data Analytics and Management in Data Intensive Domains" (DAMDID/RCDL'2018). – Moscow, Russia, October 9-12, 2018. – P. 277 – 283.

- 
- [5] Scopus Homepage. - <https://www.scopus.com>, last accessed 2019/11/20.
- [6] eLIBRARY.ru Homepage. - <https://elibrary.ru>, last accessed 2019/11/20.
- [7] Izaak A.D., Znamenskaia E.A., Chebukov D.E. O poteriannykh tsitirovaniiax v Web of Science i ikh vliiaii na impakt-factory zhurnalov // Proc. Nauchnyi servis v seti Internet . –Novorossiisk. – 2018. – P. 238 – 243.
- [8] ORCID Homepage. – <http://orcid.org>, last accessed 2019/11/20.
- [9] SN SciGraph Homepage. – <https://www.springernature.com/gp/researchers/scigraph>, last accessed 2019/11/20.
- [10] GRID Homepage. – <https://www.grid.ac>, last accessed 2019/11/20.
- [11] Nentwig, M., Hartung, M., Ngomo, A.C.N., Rahm, E. A survey of current link discovery frameworks // Proc. Semant. Web. – 2017. –Vol. 8 -- P. 419 – 436 .
- [12] Collarana, D., Lange, C., Auer, S.: FuhSen. A platform for federated, RDF-based hybrid search // Proc. 25th International Conference on World Wide Web. – 2016. – P. 171 – 174.
- [13] Collarana D., Galkin M., Lange C., Scerri S., Auer S., Vidal M. Synthesizing knowledge graphs from Web sources with the MINTe+framework // The Semantic Web ISWC 2018. –2018. – ( Lect. Notes in Comp. Sci.; 11137).
- [14] Saeedi A., Peukert E., Rahm E. Using link features for entity clustering in knowledge graphs. // The Semantic Web. ESWC 2018. –2018. – (Lect. Notes in Comp. Sci. ; 10843).
- [15] Frey J., Hofer M., Obraczka D., Lehmann J., Hellmann S. DBpedia FlexiFusion the Best of Wikipedia > Wikidata > Your Data. // The Semantic Web ISWC 2019. – 2019. – ( Lect. Notes in Comp. Sci. ; 11779).

- [16] Chen M., Tian Y., Yang M. and Zaniolo C. Multilingual knowledge graph embeddings for cross-lingual knowledge alignment. [ Electronic resource] –2017. – Mode of access: arXiv preprint arXiv:1611.03954.
- [17] Sun Z., Hu W., and Li C. Cross-lingual entity alignment via joint attribute-preserving embedding // International Semantic Web Conference. Springer . – 2017. – P. 628–644.
- [18] Wang Z., Lv Q., Lan X. and Zhang Y. Cross-lingual knowledge graph alignment via graph convolutional networks // Proc. Conference on Empirical Methods in Natural Language Processing. – 2018. –P. 349 – 353.
- [19] Xu K., Wang L., Yu M., Feng Y., Song Y., Wang Z., Yu D. Cross-lingual knowledge graph alignment via graph matching neural network. – [ Electronic resource] –2019.—Mode of access: arxiv.org/pdf/1905.11605v3.pdf; last accessed 2019/11/20.
- [20] Yaman B., Pasin M., Freudenberg M. Interlinking SciGraph and DBpedia Datasets using link discovery and named entity recognition techniques // Proc. 2nd Conference on Language, Data and Knowledge (LDK 2019). (OpenAccess Series in Informatics). – 2019, No. 15. — P. 15:1 – 15:8
- [21] SpringerLink digital library Homepage. - <https://link.springer.com/>, last accessed 2019/11/20.  
Fekete J.-D. Reorder.js. A JavaScript library to reorder tables and networks // Proc. IEEE VIS 2015. – October 2015, Chicago, United States.

