Analysis and Visualization Algorithm for Cross-Language Author Names Disambiguation

© Zinaida Apanovich © Vladimir Isachenko
A.P. Ershov Institute of Informatics Systems, Novosibirsk State University,
Novosibirsk, Russia
apanovich@iis.nsk.su. vv.isachenko@gmail.com

Abstract. A new algorithm for the cross-language disambiguation of author names is presented. The algorithm uses the matching of Russian and English papers and journal titles. An interactive visualization tool simplifies the analysis and modification of the results obtained.

Keywords: cross-language disambiguation, clustering, interactive visualization.

1 Introduction

The recent years have been marked by a rapid spread of large-scale knowledge bases, such as Google Knowledge Vault, Deep Dive, Microsoft Academic Graph, etc., extracting facts from texts and integrating data from multiple sources automatically which is potentially error-prone. Most errors are related to the differences in data source schemas and identity resolution errors.

Name ambiguity in the context of bibliographic citation records is a difficult problem affecting the quality of content in digital libraries. It has been a subject of intensive research [4, 5, 10, 12]. An important aspect of this problem is multilingualism. Multilingual resources such as DBPedia, VIAF, WorldCat, etc., are becoming increasingly common.

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 disambiguation of the authors of scientific publications. Also, algorithms aimed at cross-language identity resolution have recently gained importance in the field of the Semantic Web [7, 8, 15].

Our previous research demonstrated that Russian names allowing several transliterations represent a challenge. Experiments with several multilingual datasets have shown that Russian names admitting several transliterations are often treated as homonyms, and several persons with identical name variations are treated as synonyms [1, 2]. This is especially unpleasant when errors occur in the resources used to calculate scientific ratings. For example, the family name of a researcher of the A.P. Ershov Institute of Informatics Systems of the Siberian Branch of the Russian Academy of Sciences (IIS SB RAS), Валерий Александрович Непомнящий can be transliterated as Nepomnyashchii, Nepomniaschy, Nepomnyashchiy, etc. Because of these name variations, his papers in the Scopus data base are assigned to four people with distinct Scopus identifiers. Moreover, some of his papers are merged with the papers by Владимир Непомнящий from Moscow and so assigned to yet another "virtual" person.

Recognizing the great importance of this issue, large bibliographic data sources started such projects as ORCID (http://orcid.org/). It provides persistent digital identifiers (Open Researcher and Contributor ID) that distinguish every researcher from any other. Also, ORCID supports automated linkages between a researcher and his or her professional activities, such as publications. Nevertheless, this project has not coped with the problem entirely, and further investigations are needed.

An algorithm for the cross-language identity resolution using the SBRAS Open Archive is presented in [2]. The algorithm relies heavily on the information about Siberian researchers and their affiliations and for this reason has a very limited application. Another problem is the absence of a trustworthy data source [4, 9] since our experiments have shown erroneous authorship attributions in all important international data source such as DBLP, Scopus, etc. A possible solution to this problem would be using as a ground truth source a national data base such as eLIBRARY.ru (https://elibrary.ru).

The authors have committed themselves to answering the following question: To what extent a data source such as eLIBRARY.ru can be used to refine the quality of the identity resolution of English data sources? 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 [2]; 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 and their English translations. Due to this extended matching step, the new clustering algorithm for disambiguation of authors have proven to be more efficient that the previous one. Finally, an

Proceedings of the XX International Conference "Data Analytics and Management in Data Intensive Domains" (DAMDID/RCDL'2018), Moscow, Russia, October 9-12, 2018

interactive visualization algorithm provides comprehensible disambiguation results and enables their modification. As for the visualization issue, only a few works directly relate to our program [6, 11, 13, 14]. None of them is related to the cross-language disambiguation issue.

The paper is organized as follow s: first, the datasets and metadata essential for our algorithm are presented. After that, the matching and clustering algorithm and implementation details are described. Finally, we demonstrate an interactive visualization, which facilitates the comprehension of the disambiguation results and allows users to improve them.

2 Datasets and their metadata

Data sparsity is the decisive factor leading to author disambiguation errors in all the existing data sources. Since metadata provide an important evidence for name disambiguation tasks, the lack of key metadata can result in a poor disambiguation outcome. With publishers providing more metadata with more frequent updating, recent citations have higher metadata availability and the lists of the available metadata are growing continuously.

We have chosen the SpringerLink digital library (https://link.springer.com/) as an English-language bibliographic data source. The main reason for it is a permanently expanding set of metadata. SpringerLink is currently one of the largest digital libraries with over 10 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 pdf format, etc. One of the recent innovations is the "translated from" label for the papers written in foreign languages. This additional data makes possible the improvement of the disambiguation quality by matching the data of the original and translated paper versions.

eLIBRARY.ru stores data in the field of science, technology, medicine and education on more than 28 million publications, more than 500 000 researchers and over 3, 000 registered organizations. 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 employee's publications. The sets of metadata provided by eLIBRARY.ru are similar to those of SpringerLink, but access to these metadata is restricted. To be more specific, the list of publications of an author is freely available, but detailed metadata on his/her papers is not free. Therefore, our disambiguation algorithm is based on the *freely available data* 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.

3 The algorithm description

The main steps of our algorithm are as follows.

1. Given a full Russian name, all possible forms and English transliterations are generated.
2. English forms of the name are used for the keyword search of publications in the 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 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 makes it possible to analyze and refine the clustering result.

The general scheme of our algorithm is shown in Fig. 1. Next, we describe each step in more detail.

3.1 Extended transliteration

eLIBRARY.ru identifies the researchers by their normalized name, affiliation and location of the employing organization. Since eLIBRARY.ru is a Russian-language data source, all the three attributes are written in Cyrillic. The format of the normalized name is <LastName First Name Middle Name>.

However, several English name variations can correspond to a normalized Russian name. It can be < First Name Middle Name Last Name>, <First Name Last Name>, <First Name First letter of the Middle Name Last Name >, etc. All these forms should be first generated in Russian and then transliterated in English. Again, every Russian name can be transliterated in many ways. For example, the Russian family name Ершов can be spelt as Ershov, Yershov, Jerszow, and the first name Андрей can be written as Andrei, Andrey, Andrew. Therefore, in order to identify in an English knowledge base all the possible synonyms of a person from eLIBRARY.ru, our program generates the most complete list of English spellings for each Russian name. This procedure is applied in the character by character manner.

Given a full normalized Russian name, the program generates a set of all possible English transliterations and form variations *E_strings,* as explained earlier [2]. This step should allow extracting the most complete set of synonyms for a given person.

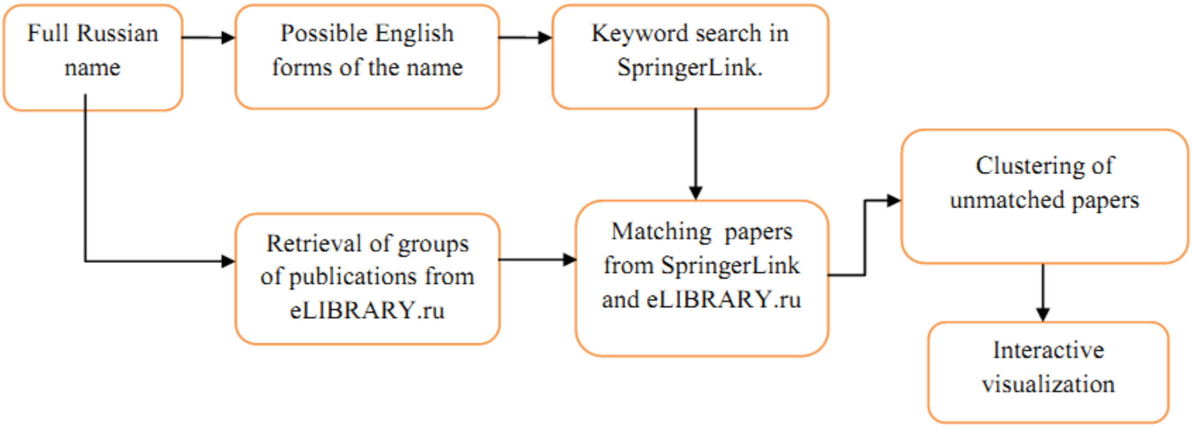

Full Russian name
Possible English forms of the name
Keyword search in SpringerLink.
Retrieval of groups of publications from eLIBRARY.ru
Matching papers from SpringerLink and eLIBRARY.ru
Clustering of unmatched papers
Interactive visualization

Figure 1 The general scheme of our algorithm

3.2 Extraction of papers from SpringerLink

Each generated string $s \in E\_strings$ is used for key word search in SpringerLink. Publications having one of the key words as the author are retained. Sets of meta-data such as the title, list of authors, list of author's affiliations, publication date, venue (title of journal or conference proceedings), keywords, pdf_url are extracted from SpringerLink. If a publication extracted is a translation of a Russian original paper, SpringerLink provides a special label "translated from." For example, the paper by A. P. Ershov Design characteristics of a multilanguage programming system
has the label "Translated from Kibernetika, No. 4, pp. 11–27, July–August, 1975." Also, such attributes as DOI (https://doi.org/10.1007/BF01070432) and bibliographic data (Cybernetics July 1975, Volume 11, Issue 4, pp. 526–541) are provided. Moreover, the SpringerLink database gives the ISSN of the translated version. This information can be used for cross-language matching of the original paper in Russian and its translation in English.

3.3 Extraction of paper lists from eLIBRARY.ru

eLIBRARY.ru specifies persons by their full Russian name in <LastName First Name Middle Name> format, affiliation and location of the employing organization. eLIBRARY.ru lists of publications are used to create confirmed groups. The simplest solution would be to extract from eLIBRARY.ru the publication list of a person specified by his or her full name. The problem is, however, that a person under consideration can have several homonyms and "partial" homonyms, when a short form of the person's name coincides with the short form of another person's name. For example, a full Russian name "Andrei Petrovich Ershov" has a short form "A. P. Ershov." However, Alexander Petrovich Ershov from the Lavrentiev Institute of Hydrodynamics and Alexei Petrovich Ershov from Moscow State University have the same short form of their names and used to 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. The groups of confirmed papers of SpringerLink are created by matching the papers from SpringerLink and eLIBRARY.ru.

3.4 Matching the papers from SpringerLink and eLIBRARY.ru

The authors of the articles extracted from SpringerLink can be both homonyms and synonyms. The disambiguation algorithm should process the list of articles and determine which of their authors are synonyms and which of them are homonyms. In other words, the list of articles should be clustered into the subsets S_1, S_2,…, 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,…, 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.

A paper $s_i \in S$ is considered to be identical to a paper $e_j \in E$ in the following cases:

Case 1 Title(s_i) = Title (e_i) AND Authors(s_i) = Authors(e_i)

The unique identifier of a paper is its DOI; regrettably, only 56% of our sample papers have DOI specified in ELIBRARY.ru. 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 has the title *Schematology in a MJ I/T I-language OPT imizer* in SpringerLink. In the absence of the paper titles exact match, 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.

Case 2 Cross-language identification of paper and journal titles.

Many Russian journals are first published in Russian and then translated in English. A typical example is the *Программирование* 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 their English counterpart. These publications, however, are very important for making confirmed paper groups as large as possible.

There are several problems involved in this situation. First, it is impossible to compare papers by 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 the translated papers have disjoint sets of attributes such as venue, ISSN, publication data, page numbers, etc.

Although SpringerLink provides information about journal titles in the Latin alphabet only, every translated paper in the database mentions its Russian original. For example, the paper by A. P. Ershov *Design characteristics of a multilanguage programming system* has the label "Translated from Kibernetika, No. 4, pp. 11–27, July–August, 1975" 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 matched 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 author is data sparsity. To extend the set of the identified authors of papers, a clustering algorithm was applied to the unmatched papers.

3.5 Clustering unmatched papers

The two important aspects of our algorithm is the cross-language generation of the confirmed groups and similarity evaluation between unmatched papers. The papers of SpringerLink, which were not grouped at the previous step, are now grouped together if their similarity exceeds a specified threshold. (The threshold and all the attributes score values can be adjusted during the interactive visualization step). The schema of the algorithm is as follows.

Let $A = \cup(A_{gi})$ be a set of papers obtained after the matching step of SpringerLink and eLIBRARY.ru papers, where g_i is a group number. For the group of unmatched papers g_i = -1.

Then the following algorithm is applied:

```
For each paper $s \in A$
    For each paper  $t \in A$
        $d :=$similarity_score$(s,t)$
        If  $(d > \text{threshold})$
            if  $(\text{Group}(s) = -1 \text{ and } \text{Group}(t) = -1)$
                $\text{NewGroup}(s, t)$
            Else
                $\text{MergeGroups}(s, t)$
```

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 articles by distinct authors.

3.6 Papers similarity scores calculation

To calculate the *assignment likelihood* of an ambiguous author *A* to a confirmed group *G*, we consider similarity between p_A and $p_i \in p_G$. Given the attributes collected by the SpringerLinkExtractor, all the attributes are pairwise compared, which results in a number of scores that are summarized in the final step.

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 *Co-author_similarity_score* uses Jaccard Index to evaluate the overlap ratio of their co-author lists.

Subjects and keywords similarity The *subject_similarity_score* and *keyword_similarity_score* are calculated in the same way as the *co-author_similarity_score.*

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.

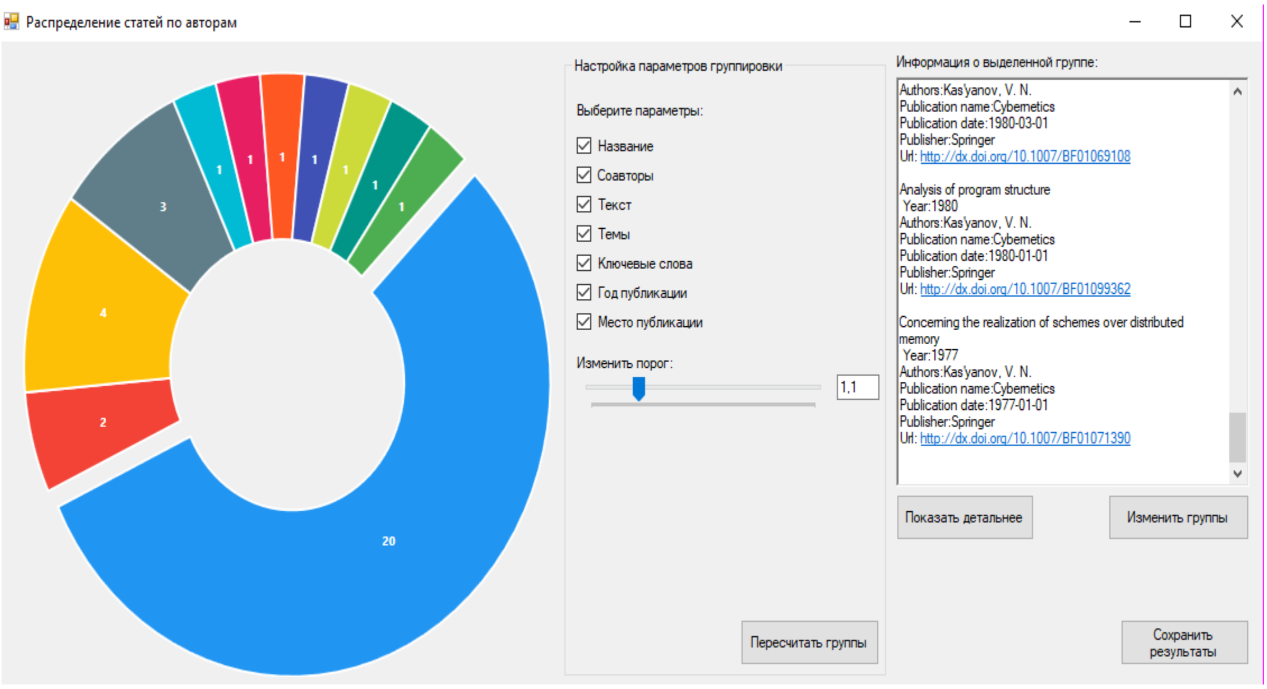
Venue similarity The *publication_venue_score* (i.e., conference/journal title) is set to 0.1 if there is an exact match between their titles.

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.

4 Interactive visualization for understanding and editing the matching and clustering results

Several interrelated visualizations seek to simplify the understanding and editing of the matching and clustering results. A global view of the obtained groups of publications is shown in Fig. 2 as a pie chart. Each segment of the pie chart corresponds to a separate group of publications attributed to a single author. The size of a segment in the pie chart is proportional to the number



Распределение статей по авторам

Настройка параметров группировки

Выберите параметры:

☑ Название

☑ Соавторы

☑ Текст

☑ Темы

☑ Ключевые слова

☑ Год публикации

☑ Место публикации

Изменить порог: 1,1

Информация о выделенной группе:

Authors:Kas'yanov, V. N.
Publication name:Cybernetics
Publication date:1980-03-01
Publisher:Springer
Url: http://dx.doi.org/10.1007/BF01069108

Analysis of program structure
Year:1980
Authors:Kas'yanov, V. N.
Publication name:Cybernetics
Publication date:1980-01-01
Publisher:Springer
Url: http://dx.doi.org/10.1007/BF01099362

Concerning the realization of schemes over distributed memory
Year:1977
Authors:Kas'yanov, V. N.
Publication name:Cybernetics
Publication date:1977-01-01
Publisher:Springer
Url: http://dx.doi.org/10.1007/BF01071390

Figure 2 A global view of the matching and clustering results

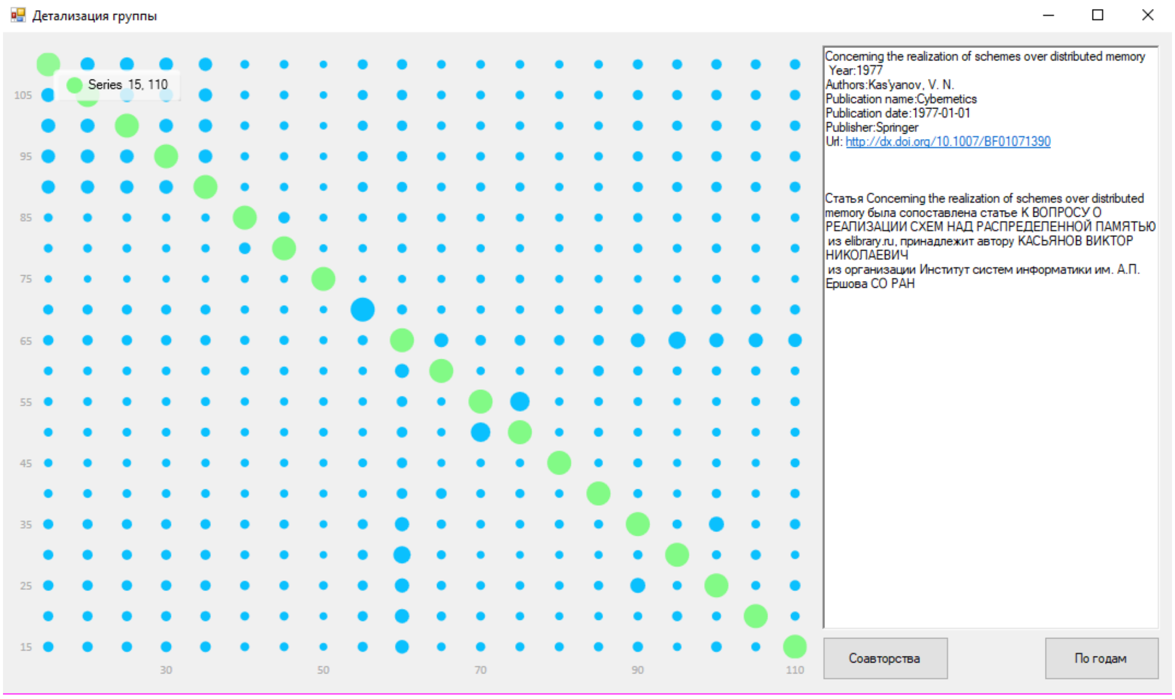
of documents assigned to this group. A short textual description of a chosen documents group appears after the mouse click on a segment of the pie chart in the right panel. A set of checkboxes in the center of the global view enables the interactive adjustment of the clustering results. Users can change the list of parameters taken into account by the clustering cost function as well as the group similarity threshold value. When the "Recalculate groups" button is pressed, the system automatically recalculates the clustering results, which makes the clustering algorithm interactively adjustable through the visualization.

The "Change groups" button displays another window which allows interactive modification of the clustering results by dragging a paper from one group to another.

The "Save results" button allows saving the updated clustering results.

The "Show details" button provides access to the visualization of individual group parameters. For example, a group of papers can be represented as an adjacency matrix A shown in Fig. 3. Each entry $a_{ij} \in A$ is shown as a colored circle with its radius proportional to the similarity value between the papers p_i and p_j.

If a paper is assigned to the group by matching with eLIBRARY.ru procedure the corresponding diagonal circle is green, otherwise it is blue. For example, a group of papers assigned by the matching and clustering procedure to the employee of the IIS SBRAS Kas'yanov V.N. is shown in Fig. 3. It is easy to see that all but one paper by Kas'yanov V.N. were found in eLIBRARY.ru, and many of them have descriptions in Russian only. When an entry a_{ij} is chosen by a mouse click, it is highlighted in red, and the description of the corresponding document pair appears, as well as a detailed explanation of the coefficient obtained.

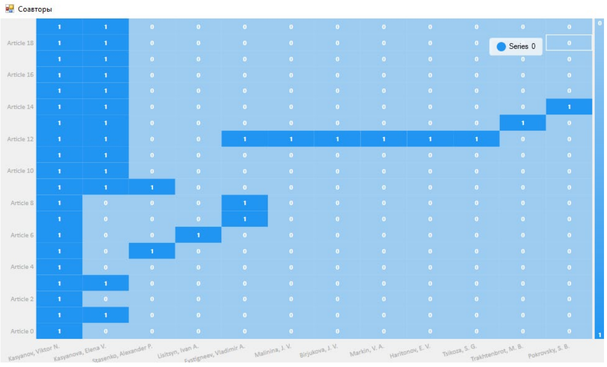


Concerning the realization of schemes over distributed memory
 Year:1977
Authors:Kas'yanov, V. N.
Publication name:Cybernetics
Publication date:1977-01-01
Publisher:Springer
Url: http://dx.doi.org/10.1007/BF01071390

Статья Concerning the realization of schemes over distributed memory была сопоставлена статье К ВОПРОСУ О РЕАЛИЗАЦИИ СХЕМ НАД РАСПРЕДЕЛЕННОЙ ПАМЯТЬЮ
 из elibrary.ru, принадлежит автору КАСЬЯНОВ ВИКТОР НИКОЛАЕВИЧ
 из организации Институт систем информатики им. А.П. Ершова СО РАН

Figure 3 Paper similarity adjacency matrix

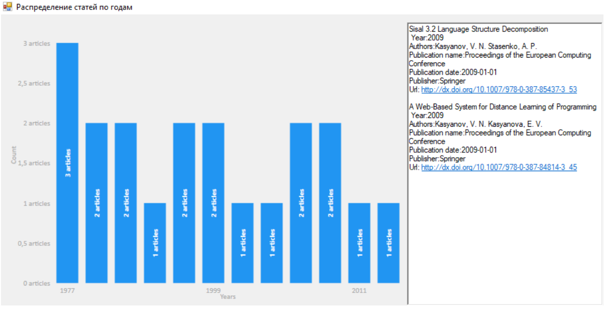
The "Co-authorship" button opens another window representing co-authors of scientific publications in the form of a matrix (see Fig. 4).



	Kasyanov, Viktor N.	Kasyanova, Elena V.	Stasenko, Alexander P.	Lisitsyn, Ivan A.	Fvstigneev, Vladimir A.	Malinina, J. V.	Birjukova, J. V.	Markin, V. A.	Haritonov, E. V.	Tsikoza, S. G.	Trakhtenbrot, M. B.	Pokrovsky, S. B.
	1	1	0	0	0	0	0	0	0	0	0	0
Article 18	1	1	0	0	0	0	0	0	0	0	[illegible]	0
	1	1	0	0	0	0	0	0	0	0	0	0
Article 16	1	1	0	0	0	0	0	0	0	0	0	0
	1	1	0	0	0	0	0	0	0	0	0	0
Article 14	1	1	0	0	0	0	0	0	0	0	0	1
	1	1	0	0	0	0	0	0	0	0	1	0
Article 12	1	1	0	0	1	1	1	1	1	1	0	0
	1	1	0	0	0	0	0	0	0	0	0	0
Article 10	1	1	0	0	0	0	0	0	0	0	0	0
	1	1	1	0	0	0	0	0	0	0	0	0
Article 8	1	0	0	0	1	0	0	0	0	0	0	0
	1	0	0	0	1	0	0	0	0	0	0	0
Article 6	1	0	0	1	0	0	0	0	0	0	0	0
	1	0	1	0	0	0	0	0	0	0	0	0
Article 4	1	0	0	0	0	0	0	0	0	0	0	0
	1	1	0	0	0	0	0	0	0	0	0	0
Article 2	1	0	0	0	0	0	0	0	0	0	0	0
	1	1	0	0	0	0	0	0	0	0	0	0
Article 0	1	0	0	0	0	0	0	0	0	0	0	0

Figure 4 Co-authorship table for a group of papers

One more window can be opened by the "By year" button. This view is shown in Figure 5. It represents distribution of papers by year.



Sisal 3.2 Language Structure Decomposition
 Year:2009
Authors:Kasyanov, V. N. Stasenko, A. P.
Publication name:Proceedings of the European Computing Conference
Publication date:2009-01-01
Publisher:Springer
Url: http://dx.doi.org/10.1007/978-0-387-85437-3_53

A Web-Based System for Distance Learning of Programming
 Year:2009
Authors:Kasyanov, V. N. Kasyanova, E. V.
Publication name:Proceedings of the European Computing Conference
Publication date:2009-01-01
Publisher:Springer
Url: http://dx.doi.org/10.1007/978-0-387-84814-3_45

Figure 5 Distribution of papers by year

These two views allow for a visual search of the so-called "group outliers" that do not really belong to the same author. By choosing a paper of interest with a mouse click and pressing the "Remove from the group" button, the user can change the paper allocation. The clustering algorithm will either automatically move this paper to another group, or create a new group containing this paper.

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 persons 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 IIS SB RAS. The total number of papers found in SpringerLink for all Russian names in this dataset was equal to 3,175. All the results obtained 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 have been correctly recognized by the matching algorithm;
- number of papers that have been correctly recognized by the matching + clustering algorithm;

These experiments have shown that 69.4 percent of papers have been correctly recognized by the matching algorithm; 86.6 percent is the share of papers that have been correctly recognized by the clustering algorithm; and 95 percent of papers have been correctly recognized by the matching + clustering algorithm.

References

[1] Apanovich Z.V., Marchuk A.G.: Experiments on using the LOD cloud datasets to enrich the content of a scientific knowledge base. In:KESW 2013, CCIS 394, pp. 1-14. Springer Verlag, Berlin Heidelberg (2013).

[2] Apanovich Z., Marchuk A.: Experiments on Russian-English identity resolution. In: Proceedings of the ICADL-2015 Conference Seul, South Korea, LNCS 9469, pp. 12-21. Springer International Publishing Switzerland (2015).

[3] D'Angelo, C.A., Giu_rida, C., Abramo, G.: A heuristic approach to author name disambiguation in bibliometrics databases for large-scale research assessments. In: Journal of the American Society for Information Science and Technology 62(2), pp. 257-269 (2011)

[4] Ferreira A. A., Gonçalves M. A., Laender A. H. F.: Disambiguating Author Names in Large Bibliographic Repositories. In: Internat. Conf. on Digital Libraries, New Delhi, India (2013)

[5] Hickey, T. B., Toves J. A.: Managing Ambiguity in VIAF. In: D-Lib Magazine 20 (July/August). (July/August). (2014). doi: 10.1045/july2014-hickey.http://www.dlib.org/dlib/july14/hickey/07hickey.html.

[6] Kang H., Getoor L., Shneiderman B., Bilgic M., Licamele L.: Interactive entity resolution in relational data. In: A visual analytic tool and its evaluation. Visualization and Computer Graphics, IEEE Transactions on, 14(5), pp. 999–1014, (2008).

[7] Lawrie D., Mayfield J., McName P., Oard D. W.: Creating and curating a Cross-language Person-entity linking collection. (2012)

[8] Lawrie D., Mayfield J., McNamee P., Oard D. W.: Cross-Language Person-Entity Linking from Twenty Languages (2015)

[9] Reijnhoudt, L., Costas, R., Noyons, E., Boerner, K., Scharnhorst, A.: "Seed+ expand": A validated methodology for creating high quality publication oeuvres of individual researchers. In: Proceedings of ISSI 2013 Vienna, arXiv:1301.5177 (2013)

[10] Schulz, Chr., Mazloumian A., Petersen A. M., Penner O., Helbing D.: Exploiting citation networks for large-scale author name disambiguation. In: EPJ Data Science, 3 (11). pp. 1-14. (2014)

[11] Shen Q., Wu T., Yang H., Wu Y., Qu H., Cui W.: NameClarifier: A Visual Analytics System for Author Name Disambiguation. In: IEEE Transactions on Visualization and Computer Graphics. vol. 23, no. 1. pp. 141-150. (2017).

[12] Song Y., Huang J., Councill I.G, Jia Li C., Giles L.: Efficient Topic-based Unsupervised Name Disambiguation. In: Proc. of the 7th ACM/IEEE-CS Joint Conf. on Digital Libraries, pp. 342–351 (2007)

[13] Stoffel F., Jentner W., Behrisch M., Fuchs J., Keim D.: Interactive Ambiguity Resolution of Named Entities in Fictional Literature.In: Computer Graphics Forum, v.36 n.3, pp.189-200, (2017).

[14] Strotmann A., Zhao D.: Bubela T.: Author name disambiguation for collaboration network analysis and visualization.In: Proceedings of the American Society for Information Science and Technology, 46(1). pp. 1–20. (2009).

[15] Sun Z., Hu W., Li C.: Cross-Lingual Entity Alignment via Joint Attribute-Preserving Embedding. In: d'Amato C. et al. (eds) ISWC 2017, Part I, LNCS 10587, pp. 628–644,(2017). DOI: 10.1007/978-3-319-68288-4_37