



ИПМ им.М.В.Келдыша РАН

Абрау-2019 • Труды конференции



Труды XXI Всероссийской научной конференции

Научный сервис в сети Интернет

З.В. Апанович

Сопоставление авторов и публикаций в разноязычных библиографических базах знаний

Рекомендуемая форма библиографической ссылки

Апанович З.В. Сопоставление авторов и публикаций в разноязычных библиографических базах знаний // Научный сервис в сети Интернет: труды XXI Всероссийской научной конференции (23-28 сентября 2019 г., г. Новороссийск). — М.: ИПМ им. М.В.Келдыша, 2019. — С. 52-62. — URL: <http://keldysh.ru/abrau/2019/theses/42.pdf> doi:[10.20948/abrau-2019-42](https://doi.org/10.20948/abrau-2019-42)

Размещена также [презентация к докладу](#)

Сопоставление авторов и публикаций в разноязычных библиографических базах знаний

З. В. Апанович¹

¹ *Институт систем информатики им. А.П. Ершова СО РАН, Новосибирск
apanovich@iis.nsk.su*

Аннотация. Задача кросс-языкового сопоставления авторов и публикаций является частным случаем задачи присваивания уникального идентификатора одной и той же сущности реального мира в разноязычных источниках данных. В данной работе представлены результаты экспериментов с несколькими версиями системы сопоставления авторов и публикаций в англоязычном источнике на основе русскоязычного источника данных. В этих версиях системы тестировались различные эвристики, поэтому в данной работе рассматриваются те из них, которые давали наилучшие результаты. Важным элементом системы является интерактивная визуализация, представляющая информацию о распределении публикаций по авторам, показывающая распределение каждой группы публикаций по соавторам и годам публикации, а также предоставляющая возможность редактирования результатов анализа. Система визуализации дополнена методами упорядочения матриц смежностей. Эксперименты с реализованной системой показали, что основным источником повышения качества алгоритма устранения неоднозначностей является расширение множества подтвержденных публикаций. Подходы, используемые в данной системе, применимы для решения задачи связывания именованных сущностей в различных разноязычных источниках данных.

Ключевые слова: разноязычные базы знаний, кросс-языковое сопоставление авторов и публикаций, связывание сущностей, интерактивная визуализация, кластеризация

Matching of authors and publications in multilingual bibliographic knowledge bases

Z. V. Apanovich¹

¹ *A.P. Ershov Institute of Informatics Systems, Novosibirsk
apanovich@iis.nsk.su*

Abstract. The problem of cross-lingual matching of authors and publications is a special case of the task of assigning a unique identifier to the same real-world entity in multilingual data sources. This paper presents the results of experiments with

several versions of a cross-lingual system for matching authors of English-language publications based on a Russian-language data source. Since different heuristics have been tested in these versions of the system, here we consider only those that have given the best results. An important element of the system is its interactive visualization tool, which provides information on the distribution of publications by authors, showing the distribution of each group of publications by co-authors and years of publication, as well as providing the ability to edit the results of analysis. The visualization system is supplemented with methods for the ordering of adjacency matrices. Experiments have shown that the main source of improving the quality of disambiguation algorithm is extending the set of confirmed publications. The approaches used in this system are applicable to solving the problem of linking named entities in various multilingual data sources.

Keywords: multilingual knowledge bases, cross-lingual matching of authors and publications, identity resolution, clustering, interactive visualization

1. Введение

В последние годы появляется все больше крупномасштабных баз знаний, при создании которых используются методы интеграции множественных источников данных, в том числе, разноязычных, что приводит к ошибкам как на уровне схемы, так и на уровне отдельных экземпляров. В случае научных баз данных и знаний, одна из проблем связана с тем, что иногда публикации одного автора приписываются нескольким разным персонам, и наоборот, публикации нескольких авторов, приписываются одной персоне [1-2]. Наши прежние работы показали, что дополнительные проблемы возникают с именами собственными, которые транслитерируются или переводятся с одного языка на другой [3]. Это особенно неприятно в ресурсах, используемых для подсчета научного рейтинга, таких как Scopus, Web of Science и др. Заметим, что благодаря таким проектам как ORCID (<http://orcid.org/>), а также взаимодействию научной общественности с этими источниками данных [4], качество решения этой проблемы улучшилось. Тем не менее, даже в области библиографических ресурсов проблему нельзя назвать полностью решенной, а помимо библиографических баз знаний, постоянно возникают новые области приложения многоязычных баз знаний, где нужны аналогичные исследования.

В работе [3] нами был представлен алгоритм установления идентичности сущностей на основе данных открытого архива Сибирского отделения. Алгоритм сильно опирался на информацию о местах работы отдельных сибирских ученых, и в силу этого имел весьма ограниченную область применения. Поэтому нас заинтересовала возможность использования большого источника данных национального масштаба, такого как eLIBRARY.ru (<https://elibrary.ru>) для идентификации авторов публикаций в разноязычных источниках данных. В ходе работы над алгоритмом была разработана новая интерактивная визуализация, которая позволяет легко понимать результаты алгоритма и их модифицировать. Следует отметить, что алгоритмы визуализации, направленные на устранение неоднозначностей, появились только недавно [7].

При этом вообще нет работ, направленных на визуализацию кросс-языкового устранения неоднозначностей.

2. Наборы данных и метаданные

В качестве англоязычного источника данных в данной работе использовалась электронная библиотека SpringerLink (<https://link.springer.com/>). Она содержит более 12 миллионов документов в различных областях исследования, а ее список метаданных постоянно расширяется. Важной причиной выбора этого источника является открытость доступа к его данным, а также наличие платформы связанных данных SciGraph (<https://www.springernature.com/gp/researchers/scigraph>), демонстрирующей упомянутые проблемы. В ней, так же как и в других англоязычных базах знаний, многим русскоязычным авторам соответствует несколько «виртуальных» персон, а также имеются публикации, авторы которых вообще не имеют уникального идентификатора. Еще одной проблемой, характерной для англоязычных баз знаний является неполнота и некорректность данных о российских научных организациях. В частности, платформа SciGraph использует для идентификации организаций-мест работы авторов базу данных GRID (<https://www.grid.ac/stats>), содержащую в настоящий момент информацию о 2033 российских организациях. Сибирское Отделение РАН, описанное в этой базе данных (<https://www.grid.ac/institutes/grid.415877.8>), имеет очень неполный список из тридцати подчиненных организаций, в котором совсем нет ИСИ СО РАН, зато в этом списке указаны образовательные организации, не имеющие никакого отношения к СО РАН.

В качестве русскоязычного источника информации использовалась научная библиотека eLIBRARY.ru, которая хранит данные о более чем 30 миллионах публикаций в различных областях науки и техники, более чем о 900 000 научных сотрудниках и 13, 000 организациях, из которых свыше 3200 являются зарегистрированными организациями. ИСИ СО РАН также является организацией, зарегистрированной в eLIBRARY.ru, и информация о публикациях сотрудников ИСИ СО РАН регулярно обновляется сотрудниками ИСИ СО РАН. Более того, усилиями Анны Андреевны Бульонковой в eLIBRARY.ru введен полный список публикаций академика Андрея Петровича Ершова.

Существенным отличием двух библиотек является то, что SpringerLink - это англоязычный источник данных, а eLIBRARY.ru - русскоязычный источник. Наборы метаданных, предоставляемые eLIBRARY.ru сходны с теми, что предоставляются библиотекой SpringerLink, но доступ к этим данным ограничен. По этим причинам алгоритм устранения неоднозначностей несколько раз модифицировался так, чтобы использовать только свободно доступные данные.

3. Схема алгоритма идентификации авторов публикаций

Алгоритм включает следующие этапы:

- Для заданного полного русскоязычного имени автора публикации генерируется так называемая *расширенная транслитерация*, включающая все возможные формы имени и транслитерации [3].
- Найденные варианты расширенной транслитерации используются для поиска публикаций по ключевым словам в англоязычном источнике данных.
- Заданное полное русскоязычное имя используется для выявления полных или частичных однофамильцев указанной персоны в русскоязычном источнике данных (eLIBRARY.ru). Множество полных и частичных однофамильцев, заданных русскоязычным именем, используется для извлечения групп публикаций из eLIBRARY.ru.
- Публикации, извлеченные из англоязычного источника (SpringerLink), сопоставляются с группами публикаций, извлеченными из русскоязычного источника (eLIBRARY.ru). При этом создаются группы *подтвержденных публикаций*.
- Статьи из англоязычного источника, которым не удалось найти соответствия в русскоязычном источнике данных, анализируются и объединяются в группы на основе сравнения с подтвержденными группами публикаций с применением различных метрик сходства.
- Результаты кластеризации предоставляются пользователю в виде интерактивной визуализации, которая позволяет пользователю проанализировать и улучшить результаты кластеризации.

Такие шаги алгоритма, как расширенная транслитерация и извлечение данных из англоязычного источника SpringerLink подробно описаны в работе [3]. Поэтому мы остановимся на этапах создания групп подтвержденных публикаций, алгоритме кластеризации и интерактивной визуализации.

Статьи, извлеченные из англоязычного источника, могут принадлежать как одному автору, заданному своим полным русскоязычным именем, так и его полным или частичным однофамильцам. Поэтому статьи, извлеченные из англоязычного источника, надо разбить на группы S_1, S_2, \dots, S_n так чтобы публикации в каждой группе принадлежали одной реальной персоне при этом все вариации его или ее имени были синонимами. Подмножество S_1 должно содержать статьи, принадлежащие исследуемой персоне. С этой целью публикации, извлеченные из SpringerLink, сравниваются с публикациями, извлеченными из eLIBRARY.ru. Важно заметить, что в eLIBRARY.ru публикации уж разбиты на группы E_1, E_2, \dots, E_m , соответствующие отдельным авторам. Поэтому, если статья $s_i \in S$ совпадает со статьей e_j из группы E_m библиотеки eLIBRARY.ru, она присваивается группе S_m .

eLIBRARY.ru идентифицирует персону по ее русскоязычному имени в формате <Фамилия Имя Отчество>, месту работы и городу, в котором

находится место работы. Простейшим вариантом было бы извлечь из eLIBRARY.ru список публикаций заданной персоны по ее или его полному имени. Но у каждой персоны может быть несколько частичных или полных однофамильцев, например, Альберт Петрович Ершов и Алексей Петрович Ершов имеют одинаковые сокращенные формы имени «А.П. Ершов», и, стало быть, их публикации могут быть неправильно атрибутированы. Поэтому алгоритм создает группы *подтвержденных* публикаций для каждого потенциального частичного или полного однофамильца исследуемой персоны. Группы подтвержденных публикаций создаются установлением соответствия между публикациями из англоязычного источника (в данном случае SpringerLink) и русскоязычного источника (в данном случае eLIBRARY.ru).

Поэтому при сравнении публикаций из двух источников выделяются два основных случая.

- 1) Статьи в обоих источниках имеют англоязычное описание. В эту же категорию попадает ситуация, когда публикация в eLIBRARY.ru содержит описание русскоязычной версии, а в поле «Версии» описана англоязычная версия публикации.
- 2) В англоязычном источнике имеется описание англоязычной версии публикации, а в русскоязычном источнике имеется только описание русскоязычной версии.

В первом случае, когда описания обеих статей дано на английском языке, осуществляется сравнение названий публикаций и списков авторов. Название не может однозначно идентифицировать публикацию, так как зачастую существует несколько публикаций с одним и тем же названием. Тем не менее, точное совпадение названий и имен авторов можно рассматривать как подтверждение того, что у данных статей один и тот же автор. Эксперименты показали, что периодически возникает частичное совпадение названий публикаций, в частности, из-за ошибок сканирования текстов статей. Например, статья из eLIBRARY.ru под названием *SCHEMATOLOGY IN A MULTI-LANGUAGE OPTIMIZER* имеет название *Schematology in a MJ I/T I-language OPT imizer* в библиотеке SpringerLink. В случае отсутствия точного совпадения между названиями публикаций, осуществляется стемминг названий и подсчитывается доля совпадающих слов в названии. Если доля совпадений превышает пороговое значение, названия считаются похожими, а найденные соответствия записываются в специальный файл для дальнейшего экспертного контроля.

Многие русскоязычные журналы сначала публикуются на русском языке, а затем переводятся на английский язык. Например, журнал *Программирование* издается на английском языке под названием *Programming and Computer Software*. Иногда для таких двуязычных изданий в eLIBRARY.ru размещается описание обеих версий публикации. В этом случае программа использует информацию об англоязычной версии для сравнения публикаций в англоязычном и русскоязычном источнике.

В то же время, в eLIBRARY.ru имеется значительное количество описаний публикаций, особенно касающихся давних публикаций, не содержащих англоязычного описания. Но эта информация очень важна для того, чтобы сделать группы подтвержденных публикаций как можно более многочисленными. Очевидный вариант решения состоит в том, чтобы перевести название публикации на английский язык, а затем сравнить полученный перевод с названием из англоязычного источника. Такая опция имеется в программе и реализована при помощи переводчика Яндекс.Переводчик. Но эксперименты показали, что этот способ выявляет описания идентичных публикаций не более, чем в 50 процентах случаев. С другой стороны, каждая переводная статья в базе данных SpringerLink содержит упоминание о русском оригинале. Например, статья под названием *Parallel programming with types of control*, имеющая в качестве автора V.E. Kotov, имеет в базе данных SpringerLink метку “ *Translated from Kibernetika, No. 3, pp. 1–13, May–June, 1979*”. Этой информации достаточно для того, чтобы найти русскоязычную версию статьи Котов В.Е. *Параллельное программирование с типами управления. Кибернетика. 1979. № 3. С. 1-13* в eLIBRARY.ru. Поиск соответствия между англоязычной и русскоязычной версией осуществляется на основе сравнения таких атрибутов как номер и год издания места публикации, а также номера страниц. В этом случае соответствующая англоязычная статья помечается как связанная с русскоязычной версией и подтвержденная публикация, а пара, состоящая из англоязычной и русскоязычной версии публикации, сохраняется для дальнейшей обработки.

Среднее количество статей, присвоенных подтвержденным группам в процессе установления соответствия, было около 69 процентов, в то время как количество неправильно сопоставленных публикаций было близко к нулю. Основной причиной, почему система не может правильно сопоставить некоторые статьи является неполнота данных. Для того чтобы расширить множество правильно идентифицированных авторов публикаций, все неподтвержденные статьи, извлеченные из SpringerLink, подвергались процессу кластеризации.

4. Кластеризация неподтвержденных статей

В основе алгоритма кластеризации неподтвержденных публикаций находится сравнение каждой неподтвержденной публикации с группами подтвержденных публикаций на основе различных метрик сходства. Неподтвержденная публикация присоединяется к группе подтвержденных публикаций, если ее значение сходства с заданной группой превышает пороговое значение.

Сравнение неподтвержденной публикации с группами подтвержденных публикаций осуществляются по следующим атрибутам:

- названия статей,

- списки авторов публикаций,
- темы и ключевые слова,
- даты публикаций,
- место публикации (журнал или конференция),
- сходство текстов публикаций (TF-IDF).

При подсчете программа придерживается следующих правил:

Если обнаружено полное совпадение названий двух публикаций, коэффициент сходства устанавливается в 1.0. В противном случае производится стемминг названий, и коэффициент сходства устанавливается равным доле совпадающих слов.

Коэффициент сходства между местами публикации (название журнала, конференции) устанавливается равным 1.0 при полном совпадении названий.

Коэффициент сходства между датами публикации устанавливается равным 0.1, если разница в датах публикаций не превышает 5 лет. Если разница в датах публикаций превышает 25 лет, коэффициент сходства устанавливается равным -0.1.

Коэффициенты сходства тем и ключевых слов статей принимаются равными доле совпадающих терминов.

Для каждой подтвержденной группы публикаций создается список всех соавторов, и чем чаще некоторый соавтор встречается в списке соавторов исследуемого автора, тем больше его вес.

Также для каждой подтвержденной группы публикаций создается список мест работы автора, характеризующего каждую группу, и при сравнении авторов публикации сопоставляются не только имена авторов, но и их места работы.

Текстовое сходство двух публикаций оценивается при помощи TF-IDF и косинусной близости соответствующих векторов.

Все неподтвержденные статьи имеют номер группы равный -1. Если обнаружилось сходство, превышающее пороговое, между группой подтвержденных статей, и некоторой неподтвержденной статьей, то неподтвержденная статья присоединяется к подтвержденной группе.

Если обнаружилось сходство, превышающее пороговое, между двумя неподтвержденными статьями, то эти публикации объединяются в одну группу.

При объединении двух групп алгоритм следит за тем, чтобы обе группы не принадлежали множеству подтвержденных групп. Если такая ситуация возникает, то объединения групп не происходит, поскольку подтвержденные группы публикаций заведомо принадлежат разным авторам.

Заметим, что весовые коэффициенты всех атрибутов, участвующих в сравнении, а также пороговое значение могут настраиваться на этапе интерактивной визуализации.

5. Интерактивная визуализация для анализа и модификации результатов установления соответствия и кластеризации

Для упрощения понимания и модификации результатов установления соответствия между публикациями, описанными на разных языках, было разработано несколько взаимосвязанных визуализаций, подробно описанных в работах [5-6].

В частности, результаты разбиения по группам всех публикаций, извлеченных из англоязычного источника, представлялись при помощи круговой диаграммы, а результат установления соответствия между англоязычными и русскоязычными описаниями публикаций представлялся при помощи матрицы смежностей, где диагональные элементы показывали, удалось или нет установить соответствие. Каждый сегмент круговой диаграммы соответствует отдельной группе публикаций, принадлежащей одному реальному автору. Размер сегмента пропорционален количеству документов, попавших в эту группу. Короткое текстовое описание выбранной группы документов появляется при выборе любого сегмента при помощи мыши. На рис. 1. показан результат распределения публикаций автора, заданного полным именем Валерий Александрович Непомнящий, по различным полным или частичным однофамильцам.

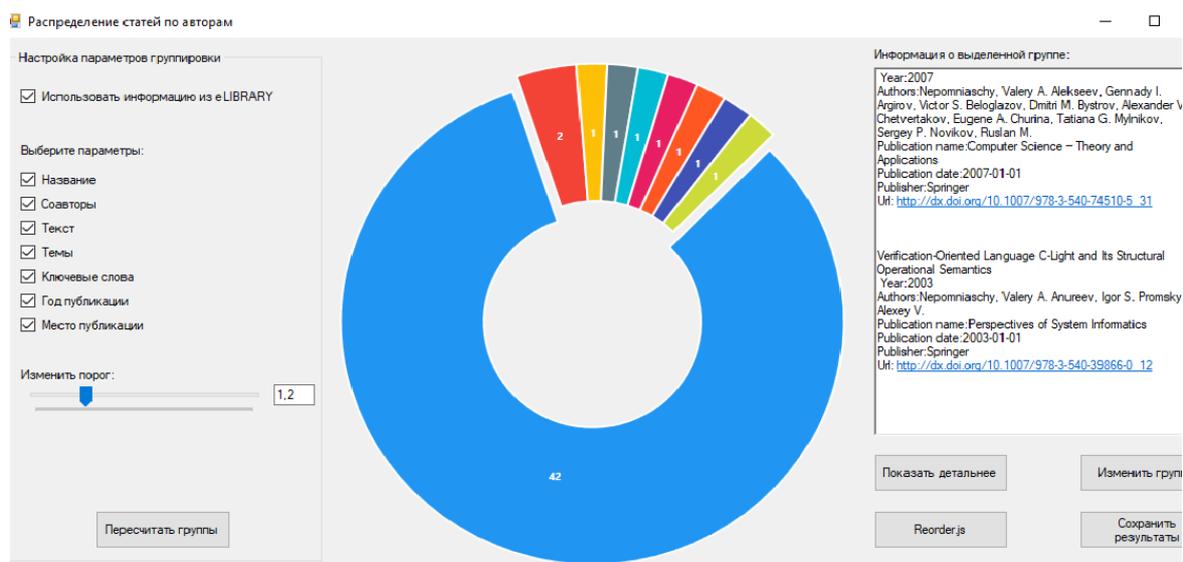


Рис. 1. Распределение групп публикаций по разным авторам

Эксперименты показали, что для настройки системы анализа пользователю удобно пользоваться глобальной визуализацией всех извлеченных публикаций, чтобы иметь возможность сравнивать публикации, распределяемые алгоритмом по разным группам. Поэтому в программу визуализации был добавлен новый модуль на основе библиотеки georder.js [8]. Визуализация, создаваемая этой программой, показана на рис. 2.

Блоки вдоль диагонали матрицы смежностей соответствуют группам публикаций, наиболее сходных между собой (то есть, принадлежащих одному и тому же автору). Программа использует те же самые значения сходства, что и программа кластеризации, реализованная в системе идентификации сущностей, поэтому легко видеть, что исходная программа устранения неоднозначностей кластеризует группы публикаций заведомо не хуже, чем алгоритмы кластеризации, встроенные в reorder.js. При этом визуализация на основе матриц смежностей позволяет пользователю находить в интерактивном режиме атрибуты, которые могут быть настроены для повышения качества программы устранения неоднозначностей.

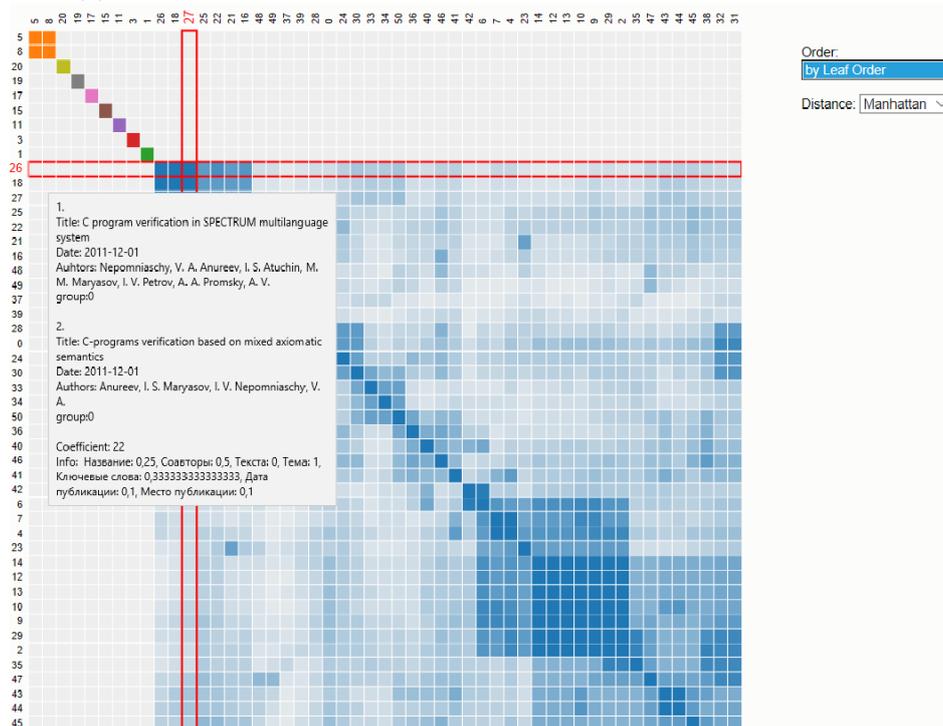


Рис. 2. Группы, сгенерированные методом упорядочения матриц смежности

6. Заключение

В данной работе описан новый алгоритм устранения неоднозначностей авторов научных публикаций на основе установления соответствия между публикациями из англоязычного и русскоязычного источников данных. Алгоритм не только корректно кластеризует публикации, но и позволяет определить личность автора, включая его место работы и местоположение организации, в которой он работает.

Программа, реализующая данный алгоритм тестировалась на 100 сотрудниках ИСИ СО РАН, работавших в разные периоды. Также набор данных содержит публикации академика А.П. Ершова. Всего для этих персон было найдено 3,175 публикаций. Все результаты, полученные программой, проверялись вручную. Для каждой персоны подсчитывались следующие значения:

- Общее количество статей, найденных в SpringerLink для каждого полного русскоязычного имени в тестовом наборе данных;
- Количество статей, реально принадлежащих сотруднику, описанному в тестовом наборе данных;
- Количество статей, правильно распознанных процедурой установления соответствия между русскоязычным и англоязычным источником;
- Количество статей, правильно распознанных процедурой установления соответствия + алгоритм кластеризации;

Эксперименты показали, что основным ресурсом для повышения качества устранения неоднозначностей является расширение множества подтвержденных публикаций за счет установления соответствия между публикациями в русскоязычном и англоязычном источниках данных. Сочетание алгоритма установления соответствия и алгоритма кластеризации позволяет корректно распознавать от 92 до 93 процентов публикаций.

Также результаты работы алгоритма сравнивались с данными Scopus, что позволило обнаружить значительное количество ошибок, которые легко могут быть устранены сравнением с русскоязычным источником. Например, сравнивая данные Scopus с данными eLibrary.ru, легко устанавливается, что Viktor Sabelfeld из Карлсруе и V.K. Sabel'fel'd вместе с Victor Sabelfeld из Новосибирска являются одной и той же персоной. Данный алгоритм может быть использован при интеграции любых разноязычных баз знаний.

Дальнейшее развитие этих исследований предполагается вести как в направлении реализации новых алгоритмов кросс-языковой идентификации сущностей, так и в направлении реализации полноценного фреймворка, включающего инструменты доступа к различным источникам данных, методы установления соответствия между различными схемами источников данных и др.

Автор выражает благодарность Авраменко М.Ю., Парамошину А.А., Исаченко В.В., Елисееву Е. С. за участие в реализации различных версий данной программы.

Литература

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. // Proceedings of ISSI 2013 Vienna, arXiv:1301.5177 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. // Proceedings of the ICADL-2015 Conference Seul, South Korea, LNCS 9469. — 2015. — pp. 12-21. — Springer International Publishing Switzerland.

4. Изаак А.Д., Знаменская Е.А., Чебуков Д.Е. О потерянных цитированиях в Web of Science и их влиянии на импакт-факторы журналов// Научный сервис в сети Интернет. 2018. — № 20 (20) . — С. 238-243.
5. Apanovich Z., Isachenko V., Analysis and visualization algorithm for cross-language author names disambiguation// Proceedings of the XX International Conference “Data Analytics and Management in Data Intensive Domains” (DAMDID/RCDL’2018), Moscow, Russia, October 9-12, 2018. — pp. 277-283
6. Исаченко В. В., Апанович З. В. Система анализа и визуализации для кросс-языковой идентификации авторов научных публикаций // Вестн. НГУ. Серия: Информационные технологии. 2018. — Т. 16, № 2. — С. 49–61. doi: 10.25205/1818-7900-2018-16-2-49-61
7. Shen Q., Wu T., Yang H., Wu Y., Qu H., Cui W.: NameClarifier: A Visual Analytics System for Author Name Disambiguation. // IEEE Transactions on Visualization and Computer Graphics, 2017. — vol. 23, no. 1. — pp. 141-150.
8. Fekete J.-D. Reorder.js: A JavaScript Library to Reorder Tables and Networks// IEEE VIS 2015, Oct. 2015. Poster. URL: <https://hal.inria.fr/hal-01214274>. 5

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. // Proceedings of ISSI 2013 Vienna, arXiv:1301.5177 . — 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. //Proceedings of the ICADL-2015 Conference Seul, South Korea, LNCS 9469. — 2015 . — pp. 12-21. — Springer International Publishing Switzerland.
4. Izaak A.D., Znamenskaia E.A., Chebukov D.E. O poteriannykh tsitirovaniikh v Web of Science i ikh vliiaii na impakt-factory zhurnalov// Nauchnyi servis v seti Internet. 2018. — № 20 (20) . — S. 238-243.
5. Apanovich Z., Isachenko V., Analysis and visualization algorithm for cross-language author names disambiguation// Proceedings of the XX International Conference “Data Analytics and Management in Data Intensive Domains” (DAMDID/RCDL’2018), Moscow, Russia, October 9-12, 2018. — pp. 277-283
6. Isachenko V. V., Apanovich Z. V. Sistema analiza i vizualizatsii dlia kross-iazikovoi identifikatsii avtorov nauchnykh publikatsii // Vestn. NGU. Seria: Informatsionnye tekhnologii. 2018. — Т. 16, № 2. — S. 49–61. doi: 10.25205/1818-7900-2018-16-2-49-61
7. Shen Q., Wu T., Yang H., Wu Y., Qu H., Cui W. NameClarifier: A Visual Analytics System for Author Name Disambiguation. // IEEE Transactions on Visualization and Computer Graphics, 2017. — vol. 23, no. 1. — pp. 141-150.
8. Fekete J.-D. Reorder.js: A JavaScript Library to Reorder Tables and Networks// IEEE VIS 2015, Oct. 2015. Poster. URL: <https://hal.inria.fr/hal-01214274>. 5