Ontology Alignment Revisited: A Bibliometric Narrative

Majid Mohammadi [a,*], Amir Ebrahimi Fard [a]
[a] *Delft University of Technology, The Netherlands*

Abstract. Ontology alignment is an important problem in the Semantic Web with diverse applications in various disciplines. This paper delineates this vital field of study by analyzing a core set of research outputs from the domain. In this regard, the related publication records are extracted for the period of 2001 to 2018 by using a proper inquiry on the well-known database Scopus. The article details the evolution and progress of ontology alignment since its genesis by conducting two classes of analyses, namely, semantic and structural, on the retrieved publication records from Scopus. Semantic analysis entails the overall discovery of concepts, notions, and research lines flowing underneath ontology alignment, while the structural analysis provides a meta-level overview of the field by probing into the collaboration network and citation analysis in author and country levels. In addition to these analyses, the paper discusses the limitations and puts forward lines for the further progress of ontology alignment.

Keywords: ontology alignment, bibliometrics, scientometric

1. Introduction

The Semantic Web is an extension of the World Wide Web which aims to provide metadata for machines so that they can further understand and construe the published information and data on the Web. The consequence of having such metadata is that computers can also make reasonable interpretations, ideally identical to how humans process and understand the information. The main key to providing such information for computers is *ontologies*, with which one can model the underlying objects of a domain along with their interrelations. The design of ontologies is thoroughly subjective and is primarily reliant on the vision of the creator. Since humans potentially consider different aspects of a domain, or they might use different terminology for similar concepts, the ontologies are not uniquely defined and are distinct from each other, even those from one particular domain. If information systems using these ontologies are assumed to work independently, the difference in ontologies does not cause any issue. However, the issue emerges when these systems want to interoperate and exchange data. In this regard, a pre-processing strategy is required to align the ontologies of these heterogeneous information systems, after which they can interact with each other. This strategy is called ontology alignment (also called ontology mapping and ontology matching).

The necessity of having a tool to automatically align two different ontologies was recognized in the late 90s and early 2000 [73–75, 98]. Since the heterogeneity problem was quite epidemic and was a stumbling block in the way of interoperability, it soon found its way in many applications such as agent communication in agent-based modeling [98], ontology merging [74, 75], data integration [80], business-to-business e-commerce [76], ontology development and visioning [16], database evolution [17, 58], web service discovery [23], to name just a few. Due to the diverse applications of ontology alignment, many research studies have been dedicated to resolving the heterogeneity among information systems, and new problems are modeled to be solved by alignment techniques.

In recent two decades, tremendous efforts have been taken to further improve the field of ontology alignment [26, 45, 67–70, 77]. As a result of such efforts, there are several valuable materials available for ontology alignment researchers and others who want to uti-

*Corresponding author. E-mail: m.mohammadi@tudelft.nl.

lize or get acquainted with alignment techniques. For instance, there is a book containing the fundamentals of ontology alignment and recent advances made in this field [26]. Further, there are some useful reviews and surveys [77, 94] and some other stating the current state and future challenges of ontology alignment [89]. Although these materials are essential and help researchers get familiar with the notions of ontology alignment, they do not provide an overview of the field. Such an overview is not only critical to those who want to get familiar with the domain, but it is also vital for researchers in this field to track its progress and find new ways to further improve it. In addition, such studies allow science policymakers to get a comprehensive picture of the field. This enables them to assess the scientific advancement of this subject domain. Assessment is a vital element in decision-making as it gives a realistic picture of the current situation and prevent us from over- or under-estimation. Therefore, to be sure the right decisions are made for this field to proceed in the right path, assessment is essential. In this regard, this article utilizes the bibliometric analysis to provide a bird's-eye view to the interdisciplinary domain of ontology alignment.

Bibliometrics is a quantitative approach to study scientific activities. At its most fundamental level, bibliometrics aims to unveil the latent dynamics of scientific research and analyze its key influential factors. Content, citation, and collaboration analyses are among the commonly-practiced techniques using bibliometrics analysis. In this regard, many research fields use these techniques to delineate the importance of their fields, the impact of the lead researchers, or gauge the impact of a particular research output [61]. In recent years, the bibliometric analysis has drawn a lot of attention and is applied to major bibliometrics data engines. Bibliometrics covers a broad spectrum of application domains [31]. Some of the studies in bibliometrics have more methodological orientations and try to scrutinize the existing bibliometric measures, e.g., citation and impact factor, or to come up with new ones. For instance, Chorus et al. [12] defined a metric for self-citation and studied trends in impact factor biased self-citations of scholarly journals. In the other research, Thelwall et al. [93] made a comparison for 11 altmetrics in Web of Science to understand the relationship between real citations and alternative metrics in social media. In another prominent study, Ke et al. [52] made a large-scale analysis of the sleeping beauty (SB) phenomenon in science and introduced a parameter-free measure that quantifies the extent to which a specific paper can be considered an SB.

The other line of studies in bibliometrics calibrates research activities to provide insights regarding dynamic and vital influential factors behind scientific research. Citation analysis [38, 88], co-authorship analysis [35, 85, 88], and co-occurrence word analysis [15] are prevalent in this application domain of bibliometrics. For instance, in the study carried out by Bromham et al. [8] on The Australian Research Council's grant proposal data, they studied the relationship between research interdisciplinarity and the chance of winning grants and discovered the higher the degree of interdisciplinarity, the lower the probability of being funded. Also, in another research focusing on collaboration in the field of Genomics, Petersen et al. [79] discovered cross-disciplinary research draw more attention and get more citations correspondingly. One type of research that falls into the same line of research is the bibliometric analysis of scientific fields. For instance, Frank et al. [32] studied the bibliometric evolution of AI research and its related fields since 1950.

On the other side of the spectrum, bibliometrics is utilized to address much broader goals. Some studies use bibliometrics data or analysis for answering questions which are not for the purpose of scientific activities evaluation. This is a very recent approach toward bibliometrics, which can provide an opportunity for other disciplines to benefit the tools and techniques developed in this field. For instance, Candia et al. [10] studied the problem of collective memory decay using multiple datasets including American Physical Society (APS) papers and the United States Patent and Trademark Office (USPTO) patents. In the other work, Guimera et al. [36] studied the self-assembly of creative teams in the collaboration network using empirical study over a bibliometric dataset constitutes of 50 years records of recognized journals in social psychology, ecology, economics, and astronomy. In another research, Liu et al. [60] studied the phenomenon of a hot streak for individuals career. They conducted this study by combining over 20,000 researcher profiles in Google Scholar and Web of Science. Ebrahimi Fard et al. [21] also used bibliometric analysis to study the readiness of academia amid a war with the diffusion of fake-news in social media.

This article brings forth a bibliometric approach to analyze the growth and advancement of ontology alignment. In this regard, we searched Scopus to extract the research outputs regarding ontology alignment. We based the bibliometric analysis on the Scopus data since other databases such as Web of Science (WoS) do not index the ontology matching workshop, the primary venue in this field. We retrieve and analyze around 2,975 research outputs from Scopus, including articles, conference papers, book chapters, and reviews.

We carry out two classes of bibliometric analyses on the retrieved articles from Scopus. The first one is *semantic analysis* concerning the overall discovery of concepts, notions, and research lines flowing underneath the scientific disciplines. In this analysis, we use latent Dirichlet allocation (LDA) [5] to model the topics underlying the ontology alignment bibliometric data. To do so, the title, abstract, and keywords of each document were subjected to LDA, and six topics were extracted accordingly. Although the topics are extracted based merely on the words and their frequency in each document, the extracted topics are interestingly meaningful and delineate different applications to which ontology alignment can be applied or the problems it can address. Another analysis in this category is *thematic*, in which we show the shares of ontology alignment to top-cited articles and top percentile journals. Also, the fundamental disciplines contributing to ontology alignment are discussed in this analysis. In addition to semantic analysis, we perform *structural analysis* in order to obtain a meta-level overview of the field. We break the structural analysis into two categories. First, we analyze the collaborations between different authors and countries in ontology alignment based on their co-authorships in the bibliometric data. Second, we gauge the impact of researchers and countries by analyzing their number of published articles and their number of citations and further visualizing their citation networks. The analysis of bibliometric data helped us address some current issues in the field and also provide some solutions for its further improvement.

The remainder of this article is structured as follows. Section 2 is dedicated to the methodology by which we retrieve ontology alignment research outputs and tools that are used to analyze them. Semantic analysis is covered in Sections 3 and 4, where the topic analysis is presented in the former and the thematic analysis is discussed in the latter. Section 5 is devoted to the collaboration analysis, and the impact analysis at the author and country levels are explained in Section 6. We conclude the paper and discuss the lesson learned from the analyses in Section 7.

2. Research Methodology

In this section, we first discuss the research strategy that is employed to extract ontology alignment research outputs from two well-known databases, namely Web of Science (WoS) and Scopus. We then explain the tools and methods that are used for the analysis of extracted bibliometric data in further sections.

2.1. Ontology Alignment Bibliometric Search Approach

Bibliometric approaches aim at the quantitative analysis of the research outputs, such as publications and patents, in order to comprehend and track the scale, direction, and the innovation of a field. The major prerequisite for such an analysis is to find the relative research items according to which the analysis could be performed. In this regard, there are several well-known databases such as Scopus and Thomson Reuters Web of Science, from which the research items can be retrieved by proper queries.

For the bibliometric analysis, there are several standard ways to extract the pertinent research items to a problem/domain. Index-based methods [13] use the categories already defined by the publication database and retrieve the research outputs accordingly. The approach is simple, but the search is restricted to the indices created by the journals. In particular, we observed that there is no particular index for ontology matching in several prestigious publishers such as IEEE and ACM. Another approach is based on citation and co-citation [100], wherein one first needs to find a core corpus of research outputs that everyone agrees upon. The basic corpus of publications then evolves by using its citations and co-citations. The major drawback of this technique is that it is difficult to replication, and there is no consensus on the interpretation of citations and co-citations. For ontology alignment, in particular, finding a core amount of publications which everyone agrees upon is not easy to acquire. One potential way would be to use the papers published in the ontology matching workshop, but the number of articles in the workshop is quite restricted so that the final corpus would not include the exhaustive set of all publications for this problem. Another way to get the bibliometric data is to detect a set of journals dedicated to a domain and analyze their published articles [59]d. For ontology alignment, unfortunately, there is no particular journal to conduct the analysis. On top of that,

Table 1

Four steps for filtering the ontology alignment research outputs. *Total* is the number of items at the beginning of each step, *Relevant* and *Irrelevant* denote the number of items that are flagged as related and unrelated to ontology alignment, respectively, and *Uncertain* is the number of items that could not be flagged as either relevant or irrelevant so that they are passed to the next phase. The search query for retrieving data from Scopus is: *TITLE-ABS-KEY ("ontology matching" OR "ontology Alignment" OR "ontology mapping" OR "OAEI").*

	Total	Relevant	Irrelevant	Uncertain	Description
1	-	-	-	3289	Retrieving bibliometric data from Scopus
2	3289	1820	225	1244	Inspecting the publication items by revising the title only
3	1244	1094	53	97	Inspecting the publication items by revising the abstract only
4	97	61	36	0	Inspecting the whole paper
sum	-	2975	314	-	

ontology alignment is interdisciplinary by nature since it is used as a pre-processing strategy in many circumstances and has thus diverse applications. As a result, the research outputs are not restricted to a specific journal or domain.

One of the most popular yet straightforward methods is to use several expert-defined keywords based on which the research outputs are retrieved [56, 81]. This method is semi-automatic since the results of the search are then reviewed by an expert to exclude the irrelevant items from the analysis. After inspecting the keywords of top 50 cited research outputs in this domain and further discussion with the experts in this domain, we arrived at three main keywords: “ontology alignment", “ontology matching", and “ontology mapping". The keyword "ontology" alone refers to a more general concept in the Semantic Web and add research items that are irrelevant to ontology alignment. Thus, we need to conduct the search based on the keyword “ontology alignment", which is interchangeably referred to as “ontology matching" or “ontology mapping" as well. Thus, these terms should be considered for searching the databases. We further realize that the ontology alignment evaluation initiative (OAEI) is also essential since it might also add some research items. Since the research items that contain “ontology alignment evaluation initiative" are completely covered by articles retrieved solely by the keyword “ontology alignment", this term is redundant. However, “OAEI" must be used as another keyword. Since the use of keywords to get the articles could retrieve a considerable number of articles in ontology alignment, we use this technique to get ontology alignment research outputs.

We used four keywords to retrieve the research outputs related to ontology alignment and conducted the inquiry on Web of Science (WoS). Further, the identified research items need to be processed by an expert to verify if they are relevant to the domain. First, we conducted an inquiry in WoS by searching the identified keywords in the title, abstract, and keywords of the research outputs. The result of the search included 1,536 articles spanning from 1999 until November 2018. The 1,536 articles were processed in three different phases to leave out the articles that are not pertinent to the domain. In the beginning, the title of papers was considered since most of the related works to ontology alignment could be easily detected. In this phase, 1,166 items were identified as relevant or irrelevant, and 370 items were passed to the second phase. In the second phase, the abstract of the papers was regarded in which 316 articles were recognized as relevant, and the remainder 54 were passed to the third phase. In the final stage, the 54 articles were thoroughly inspected and the papers were classified as relevant and irrelevant. In total, 1,420 research items were labeled as relevant to ontology alignment and the rest article were eliminated.

After rigorous examinations of the remaining articles, we realized that the research items regarding the ontology matching workshop are not indexed by WoS. Since this workshop is the essential venue of this domain, we refused to continue the analysis based on WoS data. Therefore, we conducted the same search strategy in Scopus and realized that the items recovered by this database include the articles from the ontology matching workshop as well. The inquiry in Scopus retrieved 3,289 articles from 2001 up until 2018. Although it does not index several papers from the late 90s and early 2000 [73, 75], it includes all the paper from the ontology matching workshop. Since the number of articles that are not indexed by Scopus is not significant, especially compared to WoS, we use Scopus data for further analysis. The retrieved articles from Scopus underwent the same procedure as Web of Science articles in order to discard the irrelevant papers. After conducting the three phases of processing the research items, 2,975 articles are labeled as relevant to

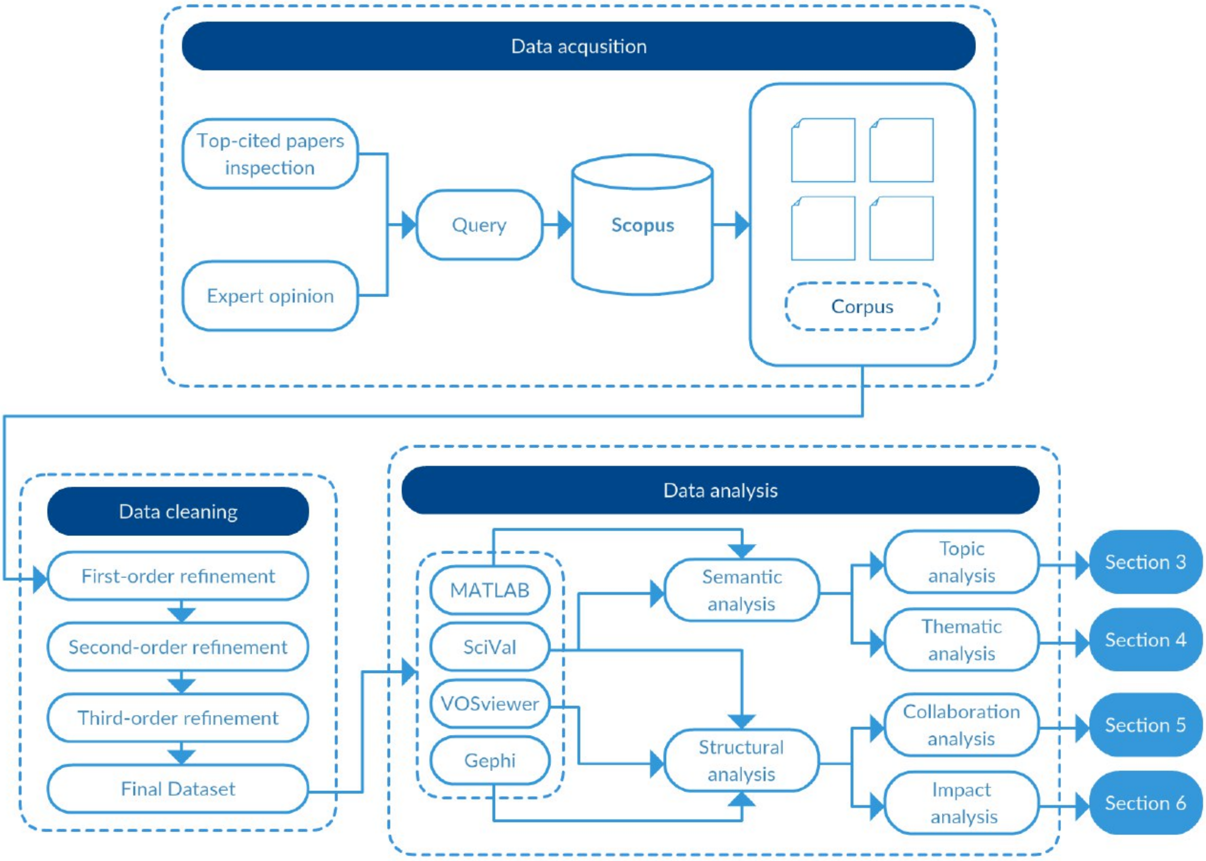


Fig. 1. The workflow of the analyses conducted in this paper.

ontology alignment. Table 1 tabulates the taken steps for obtaining and cleaning the bibliometric data from Scopus. Further analyses are performed on the remaining items after taking these steps.

2.2. *Tools and Methods for Bibliometric Analysis*

In this section, we explain the methods and tools that are used to analyze the 2,975 ontology alignment research outputs extracted from Scopus. We conduct two levels of analysis regarding the structural and semantic of the extract articles. For the semantic analysis, we use latent Dirichlet allocation (LDA) [5], which is a method for topic modeling. This analysis aims to find the underlying topics in ontology alignment based on the articles published in this domain in the period 2001-2018. We use the LDA implementation in MATLAB to analyze the articles. We also conduct a thematic analysis where we discuss the number of publications overall and in top journals along with the disciplines contribute to the interdisciplinary ontology alignment.

The second type of analysis is structural analysis, that is divided into two subcategories. We initially analyze the collaborations between authors and countries worldwide and scrutinize the level of international and academic-corporate collaborations over the last few years. We then probe into the contributions and impacts of authors and countries in ontology alignment. For these analyses, we use VOSviewer [95], SciVal[1], and also Gephi for network visualization [3]. Some of the analyses of this section are limited to recent six years due to the fact that more bibliometric metadata are only available in the recent years. Figure 1 displays the workflow of this article, from data acquisition to their processing using different tools.

[1] https://www.scival.com

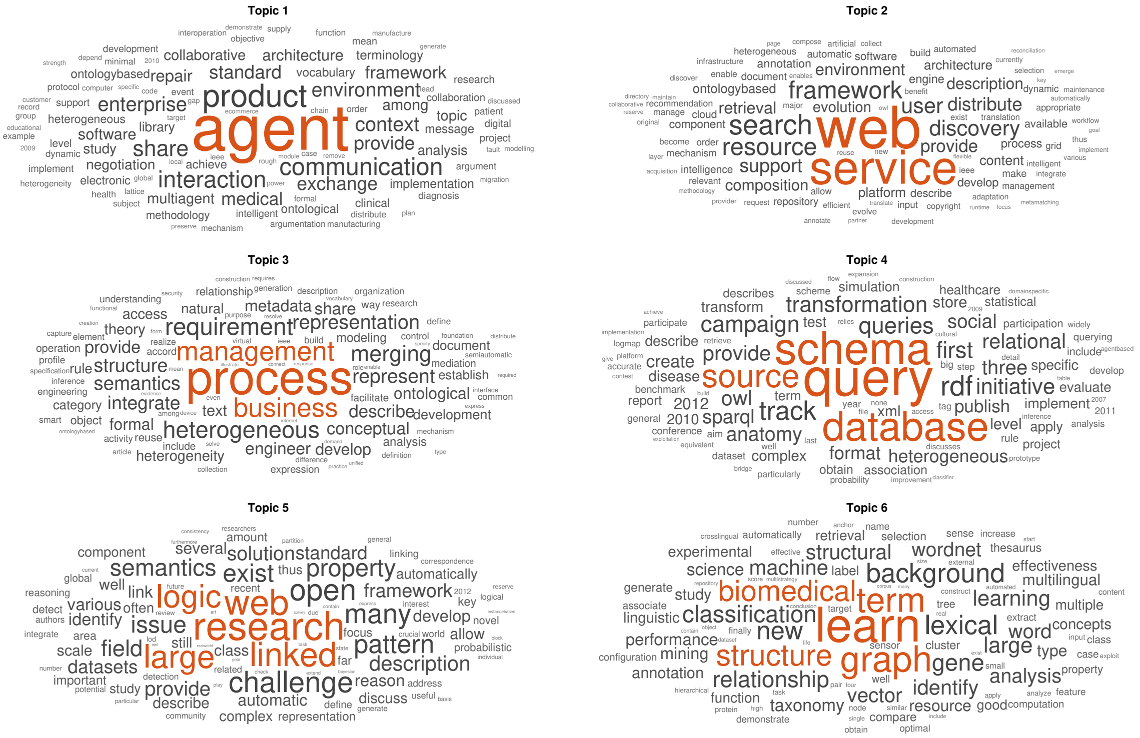


Fig. 2. Six topics of ontology alignment based on the bibliometric data 2001-2018. The topics are identified by LDA [5] and visualized in MATLAB.

3. Topic Analysis of Ontology Alignment

Topic modeling is a statistical approach to discover the underlying topics of a set of documents based on the frequency of words that appeared in the documents. Topic modeling is generally used to find the underlying hidden semantic structure of a text body. One of the most well-known algorithms for modeling the topic is latent Dirichlet allocation (LDA) [5], which can find the hidden topics of a set of texts with a given number of topics.

In this experiment, we aim at analyzing different topics in order to provide a broader picture of domains and problems to which the ontology alignment is an essential contribution. In this regard, the title, abstract, and keywords of the obtained research outputs were subjected to LDA, and the identified topics were visualized using word clouds. It is typically necessary to conduct several pre-processing procedures on the given data in order for the final topics to be more meaningful. The following pre-processing strategies are used before applying LDA:

- **Tokenization:** Tokenization means that sentences are broken into their constituent words. Tokenization transforms the set of documents into the set of bags of words, which are useful for further topic analysis.
- **Meaningless Word Removal** For meaningful analysis, stop words and the words whose length is less than three were removed since they do not have any further impact on the analysis. At the same time, we removed the terms used for the query since they exist in all the retrieved documents, while they do not convey any useful information for ontology alignment topics.
- **Lemmatization** Words were also lemmatized, which means that the verbs in the past or future tense are changed into their present tense and other third person words are replaced by their corresponding first person. We further used the Porter stemming algorithm [82], which replaces the words with their associated roots.

The processed documents were then subjected to LDA for topic modeling. LDA also needs to have the number of topics to be identified. We tried a variety of numbers between four up until 14, and we found out that more meaningful topics are obtained by six topics. Figure 2 displays the topics identified by LDA. In this figure, the size of each term indicates its importance, and the most vital terms are also identified by the orange color. Note that the order of topics is arbitrary, and hence, the topics that appeared at the beginning do not have any privilege over those presented later on. In the following, we analyze the topics discovered by LDA.

- **Agent-based Modeling** Agent-based modeling is used to simulate the actions and interactions of independent agents in order to estimate their impacts on the overall system. Agents communicate together by exchanging messages composed in different languages such as Agent Communication Language developed in Foundation for Intelligent Physical Agents (FIPA-ACL) and Knowledge Query and Manipulation Language (KQML). These languages determine only the overall structure of the messages and not their contents. The actual content of a message expressed by an agent is typically modeled by an ontology. As a result, when two independent agents communicate with each other, it is unlikely that they can understand each other if they do not use the same ontology for communication. In this regard, ontology alignment has been extensively used in order for the agents to understand various messages in different formats. One of the first problems to which ontology alignment was applied is agent communication, where the paper was published in 2002 [98]. There are also several systems that employ agent-based modeling for automatic ontology alignment, where they call such systems as *agent-based ontology alignment* [25, 48, 57]. In this view, the mappings between two ontologies are deemed as a product of communications between two intelligent agents [86]. Topic 1 in Figure 2 illustrates the topic related to articles of ontology alignment that used the notions of agent-based modeling. In this topic, as expected, the term “agent" is identified as the most important word. There are also several other terms related to agent-based modeling. For instance, “communication" and “interaction" as are usually paired with “agent", and the terms “collaborative", “negotiation", and “exchange" that refer to collaborations, negotiation, and data exchange between agents, are the features that agents can be equipped with by using ontology alignment. There are also some general terms of agent-based modeling such as “environment", “software" (as in software agent), “multi-agent", that are visible in Topic 1 of Figure 2.
- **Web Service Discovery:** Web Services contain the services provided by some providers who expose their particular services to a broad audience by using Web technologies. Semantic Web Services (SWS) are strong tools to describe the services more richly so that their discovery by requesters become even easier. Web Service discovery is the process of finding a service which is able to deliver a particular service to the person who has requested it. Sometimes a request cannot be responded merely by a single service, but by a number of services. In this circumstance, it is required to have a composition process, which integrates several services in order to meet a particular need of requesters. SWS can be modeled by different standards such as Web Services Description Language (WSDL) [1] and OWL-S [63], and different terminologies might be used by different providers/requesters. As a result, the discovery of services requires the use of ontology alignment techniques so that the discrepancy between different services are reconciled and the rate of discovery success increases significantly. SWS discovery was also one of the first applications that ontology alignment could address. The first paper employing ontology alignment for SWS discovery dates back to 2003 [9], and it has been since used in other studies [28, 87, 91].

  Topic 2 in Figure 2 is devoted to this important application of the ontology alignment. As is readily observable, the terms “web" and “services" are detected as the main terms, and there are also the terms “discovery" and “composition" as well that are the general terms in the Semantic Web Service domain.
- **Process Model Matching:** Process models comprise a set of related activities or tasks which need to be done in a specific order to finally produce a service or product. The matching of these processes is of the essence for several tasks such as system validation and process harmonization [4, 33, 97]. In this regard, ontology alignment systems or techniques can be used. A more specific application is matching the business processes [14, 29, 41, 47], where the process models are typically related to e-commerce [90].

  In the OAEI 2016 and 2017, there was a track about matching different process models of the university admission systems. As a result, the problem is completely well-known for the ontology alignment community as well. As Topic 3 in Figure 2 shows, LDA has been able to detect the importance of this problem for ontology alignment. The terms “process" and “business" are at the heart of this topic, which accentuates the importance of process model matching in ontology alignment. The term “management" also has a significant weight. Interestingly, “business process management" refers to a domain where matching of processes has become a major research area [54].

- **Query Answering:** Information provided by different sources is not described by a unified schema on the Web. At the same time, users do not utilize the same terminology in their search queries. Thus, a semantic query answering is required to rewrite the query in order to provide sensible results. Since both the information on the web and the queries are the reasons for the discrepancy, ontology alignment can be helpful to address this challenge and to improve the relevance of the retrieved information. Thus, ontology alignment has been used extensively in this regard. As Topic 4 in Figure 2 illustrates, this problem is quite important in ontology alignment. The term “query" is the most accented terms, and there are some other related terms such as “relational", “database", “schema". In the OAEI 2014 and 2015, there was a track for answering queries by the aid of ontology alignment systems, which indicates that this problem is also quite well-known to the community.
- **Linked Data and Logic:** One of the primary objectives of the Semantic Web is to link the data on the Web to other available resources so that useful information can be provided from the available data. Since the published data on the Web are designed by many people, interlinking of these data is not straightforward due to their heterogeneous nature. As a result, ontology alignment is a potential solution to fulfill this essential objective of the Semantic Web [42, 53]. This is the reason that “linked" in Topic 5 of Figure 2 has been centralized. Another vital term in this topic is “logic". Logic has been widely used to align two different ontologies [45, 55, 71, 72]. One of the well-known systems is LogMap [45], which is based on logic and is one of the top systems at the OAEI in the recent decade. Also, logic has been used to repair the alignment automatically obtained from the alignment systems [64, 65, 78].
- **Machine Learning and Biomedical Ontology:** Topic 6 is a mixture of a well-known approach for ontology alignment and one essential domain to which ontology alignment has been applied. There are several ontology alignment systems which use machine learning techniques for alignment. In fact, machine learning techniques are one of the first approaches that are used for aligning ontologies and dates back to 2002 [18, 19]. There are also many machine learning-based systems that require to have a gold standard for training [40, 62, 84]. These systems are sometimes called pre-trained systems [26] and need to have [a part of] the reference alignment for training. The system is then ready to map the rest of the ontologies or other ontologies in the same domain. The terms “learn", “machine", “learning", and “classification" in Topic 6 are the indicators of these alignment systems.

  Another term in this topic is “biomedical", which is one of the most important domains to which the ontology alignment has been applied. The anatomy track, which is about matching the adult mouse anatomy and a part of NCI thesaurus comprising the human anatomy, is one of the first tracks of the OAEI [20]. There are several recent tracks such as disease and phenotype [37] and large biomedical [43, 44, 46] tracks which have been recently added to the OAEI tracks. Therefore, there is no surprise to see this term as a major topic of ontology alignment. The terms “large" and “background" are also related to this theme since there is one large biomedical track in the OAEI and it is the common practice to use of background knowledge such as UMLS [7] for matching ontologies in the biomedical domain.

4. Thematic Analysis

In this section, the thematic analysis of the ontology alignment publications is presented based on the collected bibliometric data between 2001 and 2018. We first study the number and types of the research outputs, and it is then followed by the contributions of ontology alignment publications to the top-cited and top journal percentiles. Afterward, the disciplines contributed to ontology alignment are discussed.

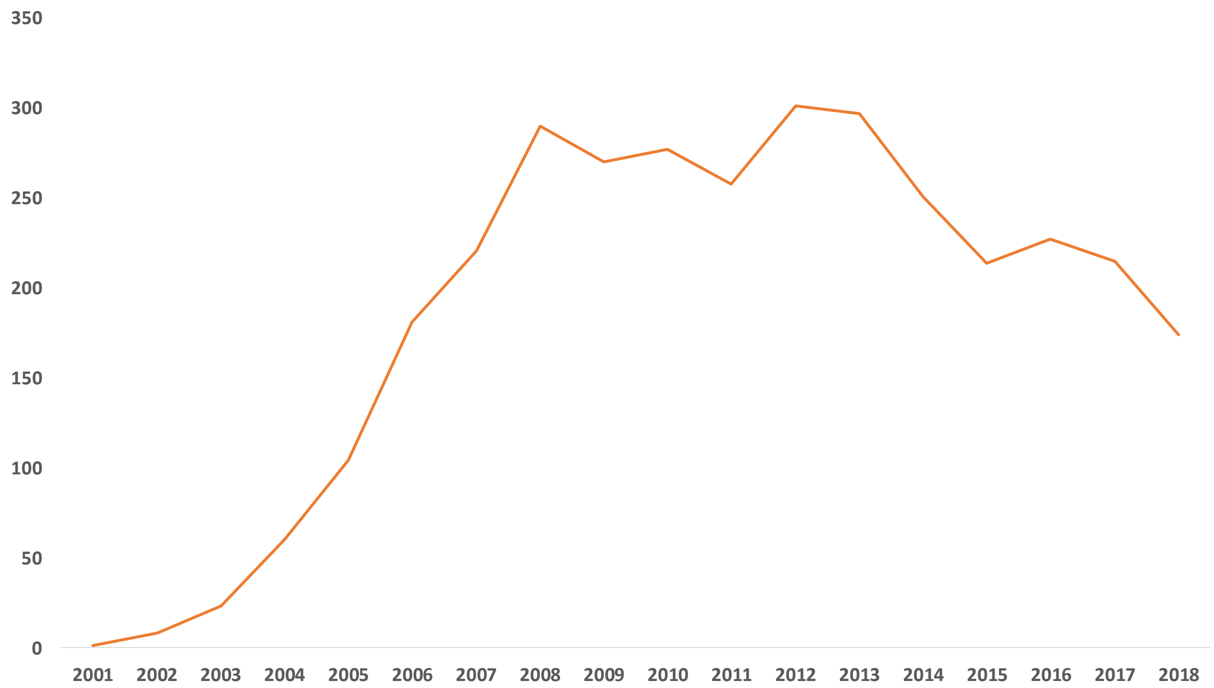


Fig. 3. The number of documents published about ontology alignment on Scopus between 2001 to 2018.

4.1. Number and Types of Published Documents

In this subsection, we discuss the number and the types of research outputs in the ontology matching data from 2001 to 2018. The essence of having the automatic mapping between two ontologies was discussed for in the late 90s and early 2000 for different problems such as ontology merging [74, 75] and further in business-to-business (B2B) electronic commerce [76], where the mappings between ontologies, taxonomies, and the classification system are required. In the preceding years, the existence and importance of such an automatic mapping were discussed in several other problems such as agent communication in multi-agent systems [98]. Ever since, ontology alignment has been the topics of numerous research studies, by which various problems have been addressed. Figure 3 shows the number of research outputs from 2001 to 2018. According to this figure, the number of outputs has been steadily increased until 2008, when around 290 research articles are published. From 2008 up until 2013, the number of publications has been approximately the same, where the maximum number of outputs is in 2013 with 300 publications. From 2013, the number of documents has experienced a steady decrease, where its minimum number reached in 2018 with 175 research outputs. Interestingly, in 2013, Shvaiko and Euzenat [89] showed the improvement of the field based on their analysis on the state-of-the-art ontology matching systems and the results of evaluations, while they observed that the speed of the ontology alignment progress was slowing down. The slow progress in the field has shown itself in the number of publications in the field as one important criterion.

We also analyze the types of research outputs in the ontology alignment field. Figure 4 displays the percentage of different types of papers published in the ontology matching domain between 2001 up until and including 2018. According to this figure, the vast majority of research outputs, i.e., around 65%, are published in conference venues. It is no surprise since the main venue for this field is the ontology matching workshop held in international semantic web conference (ISWC), where there has been several papers and posters along with the alignment contests. Aside from conference papers, journal articles comprise 25% of the publications and are ranked as the following types of publication in ontology matching. Conference reviews and book chapters are the other major types of articles in this domain.

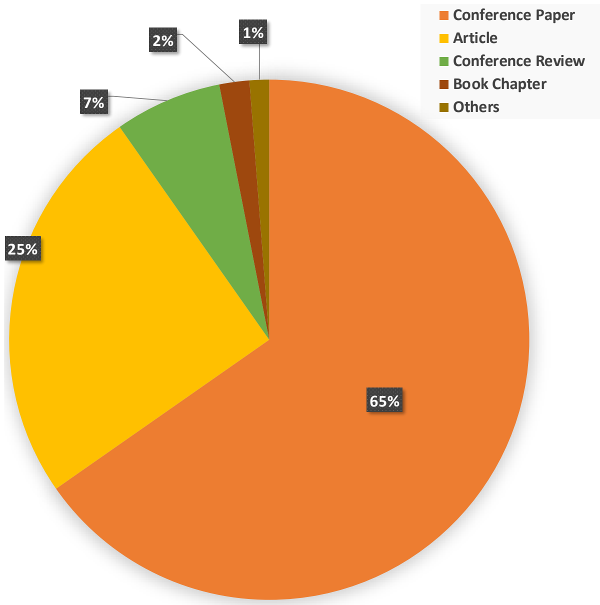


Fig. 4. The types of documents published about ontology alignment on Scopus between 2001 to 2018.

4.2. Outputs in Top Percentiles Worldwide

In this subsection, the appearance of ontology alignment research outputs in top-cited and journal percentiles in the recent six years is explored.

We first look into the ratio of ontology alignment publications in the top-cited percentiles. In recent six years, there are 60 publications in the top 10% most cited publications worldwide. The distribution of these top-cited articles in the recent six years is shown in Figure 5, where the light color shows the percentile in top 10% most cited and the dark color denotes the percentile in top 1% most cited articles worldwide. According to this figure, ontology alignment research outputs constitute one percent, 0.9%, and 0.5% of the top 1% most cited articles worldwide in years 2013, 2015, and 2017, respectively. At the same time, there is no ontology alignment output in the top 1% most cited for 2014 and 2016. It is also readily seen that ontology alignment outputs form 6.1%, 5.3%, and 5.8% of the top 10% most cited article worldwide in years 2013, 2016, and 2018. Interestingly, although the outputs from 2016 and 2018, which have a considerable amount of papers in the top 10% most cited articles, they do not have any in the top 1% most cited research outputs worldwide. The top-cited articles in the recent six years are tabulated in Table 2. As expected, four of these articles are published in 2013 so that this year is considered the best year in the recent six years in terms of the number of research outputs in the top 1% and the top 10% most cited articles.

We also analyze the share of ontology alignment in the top journals by CiteScore. Figure 6 illustrates the ratio of ontology alignment outputs in top journals from 2013 to 2018. According to this figure, the ontology alignment share in the top 1% journals is 2.1%, 1.7%, 1.9%, and 1.5% for 2013, 2015, 2017, and 2018, respectively. Similarly, the share in the top 10% journals is 13.8% as the highest, followed by 11.8% in 2014, 9.9% in 2015, and 9.2% in 2018. There are a steady decrease and increase in publishing in the top 10% journals from 2013 to 2016 and from 2016 to 2018, respectively.

4.3. Fundamental Disciplines to Ontology Alignment

In this section, we consider the disciplines that contribute to ontology alignment. In this regard, we take advantage of all science journal classification (ASJC) used in Scopus and visualize the main areas along with their subcategories that contribute to the growth and evolution of the ontology alignment field.

Figure 7-(a) displays the main subject areas contributed to ontology alignment based on the publication data between 2001 and 2018. As expected, computer science is the area with the maximum number of publications and constitutes 55% of the overall research articles. More in details, Figure 7-(b) dis-

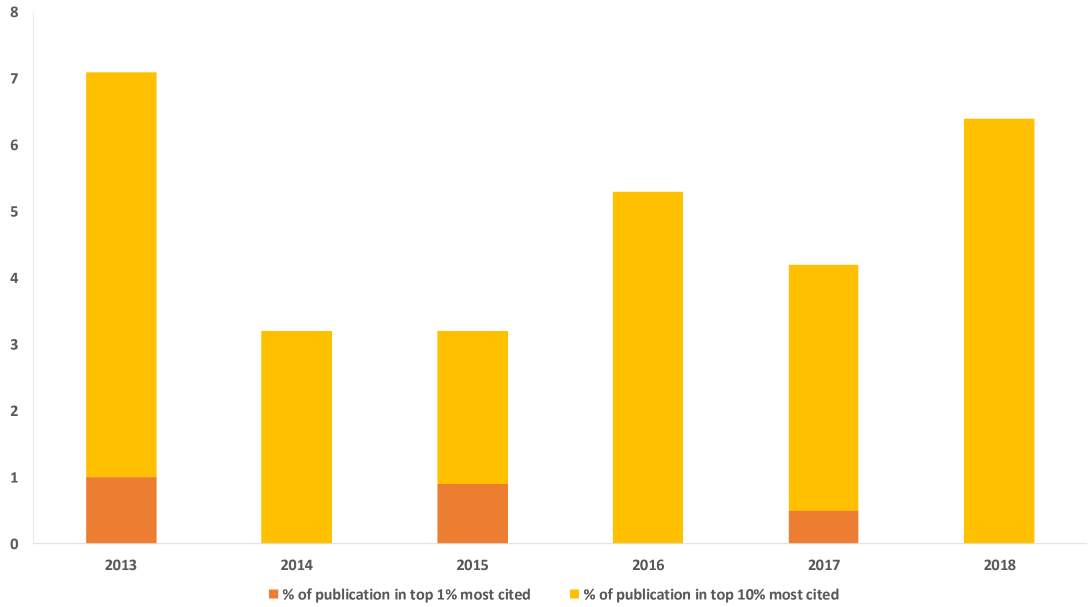


Fig. 5. The share of ontology alignment research outputs to the top 1% and the top 10% most cited articles published in all disciplines.

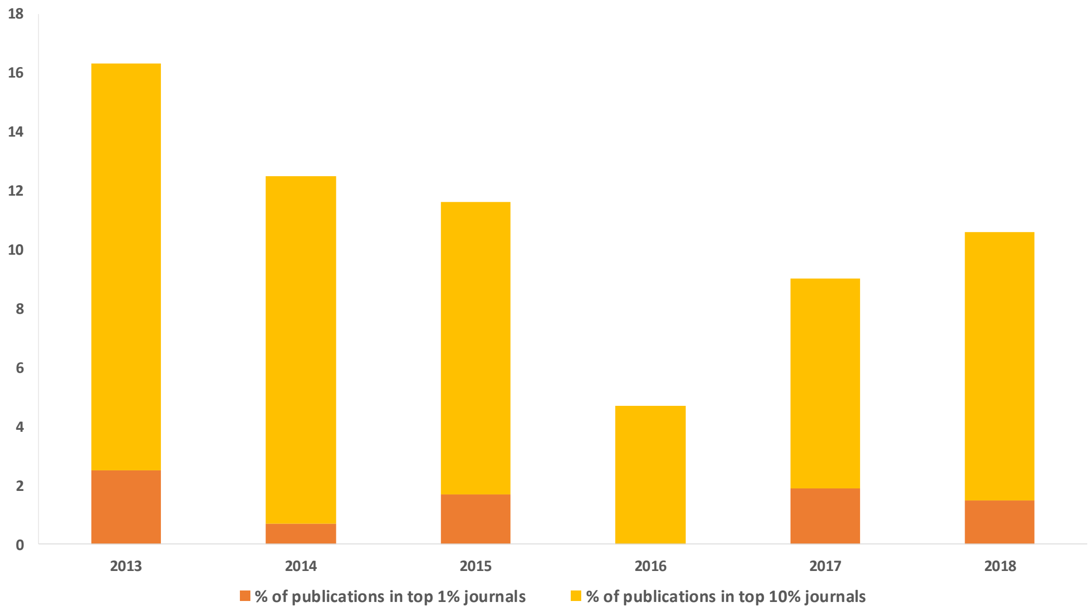


Fig. 6. The share of ontology alignment research outputs to the top 1% and the top 10% journals of all disciplines.

Table 2

Five top-cited publications in ontology alignment in the recent six years.

	Title	Main Authors	Year
1	Ontology matching: State of the art and future challenges	Shvaiko, P., Euzenat, J.	2013
2	Ontology matching: Second edition	Shvaiko, P., Euzenat, J.	2013
3	Ontology matching: A literature review	Otero-Cerdeira, L., RodrÃ■guez-MartÃ■nez, F.J.	2015
4	Scaling semantic parsers with on-the-fly ontology matching	Kwiatkowski, T., Choi, E., Artzi, Y.	2013
5	The AgreementMakerLight ontology matching system	Faria, D., Pesquita, C., Santos, E.	2013

plays the subcategories of computer science contributing to ontology alignment. According to this figure, *general computer science* has the most of published articles, followed by *software* (12.2%), *computer networks and communication* (11.8%), and *information systems* (9.7%).

The second major discipline to the ontology alignment development is *mathematics*, which forms 14.8% of the overall research outputs. In particular, Figure 7-(c) illustrates the subcategories of mathematics, which indicates that *theoretical computer science* makes up 60.9% of the overall articles related to this category, and it is followed by *general mathematics* (9.3%) and *modeling and simulation* (9.0%).

The third main area is *engineering*, which forms 10.8% of all publications in this field. More in details, Figure 7-(d) displays the subcategories of engineering contributing to ontology alignment. According to this figure, *control and system engineering* has 29.7% of publications and is followed by *general engineering* (26.4%) and *electrical and electronic engineering* (20.4%).

Social science is another important area contributing to ontology alignment, which constitutes 3.8% of all publications. Figure 7-(e) shows the subcategories of social science which have the most share in ontology alignment research outputs. According to this figure, *library and information system* has the largest part of publications, i.e., 39.2%, and *education* (19.6%) and *linguistic and language* (16.5%) follow it.

Decision science has 3.3% of overall research outputs in ontology alignment. Figure 7-(f) shows that *information system and management* (83.3%), *general decision science* (11.1%), and *management science and operations research* (5.6%) are the subcategories in this category with contributions to the development of ontology alignment.

As the last subject area contributing to ontology alignment, *medicine* makes up 2.5% of publications. More in details, Figure 7-(g) displays the subcategories of this area with the shares in research outputs. According to this figure, *health informatics* (70.5%), *general medicine* (8.2%), and *medicine (miscellaneous)* (8.2%) are the subcategories with contributions to ontology alignment.

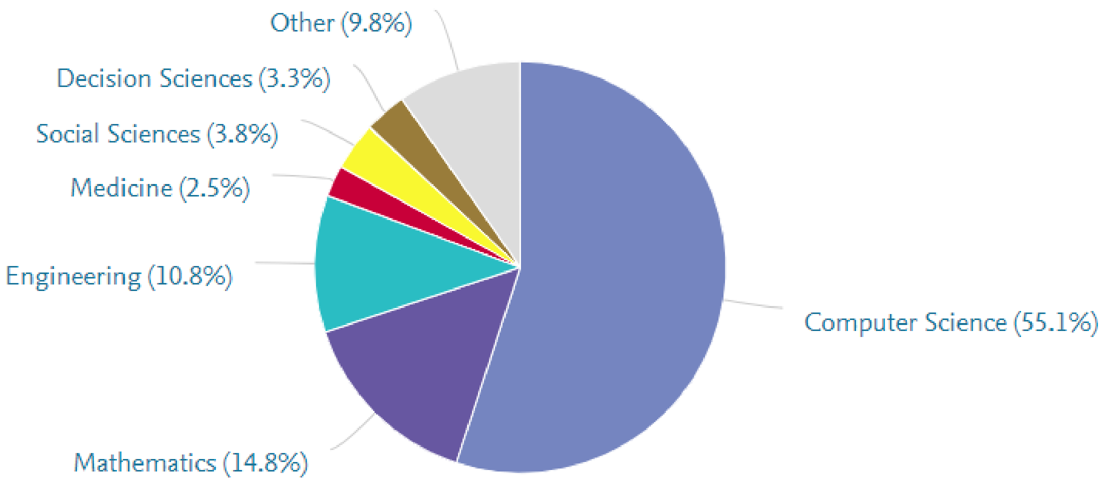
5. Research Collaboration in Ontology Alignment

The collaboration of researchers within an area widens the impact and scope of the corresponding scientific field. As a result, it is of the essence to monitor, and even encourage, the collaborations between different researchers and institutes all over the world. In this section, the research collaboration in ontology alignment is investigated. In this regard, we consider the collaboration between authors and countries and identify their most collaborative elements. We further analyze the trend of collaborations in recent years and academic-corporate collaborations in ontology alignment.

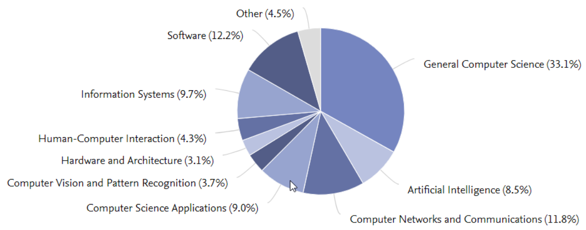
5.1. Author Collaboration

The collaborations among authors of ontology alignment are visualized in Figure 8 based on the bibliometric data 2001-2018. The authors in this graph are represented by nodes, where their size is proportionate to the number of collaborative articles, and their color denotes the average number of citations per publication according to the collected data from 2001-2018. Also, the thickness of each edge between two authors is commensurate with the number of collaborative publications of the authors at the two ends.

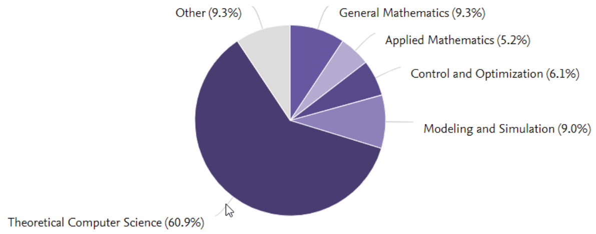
According to Figure 8, it is evident that the organizers of the OAEI have great partnerships, and the most collaborative authors are also coming from this community. In particular, J. Euzenat, P. Shvaiko, and E. Jiminez-Ruiz have 82, 63, and 63 co-publications, respectively, and lead the list of the author collaborations. Table 3 tabulates the top collaborative authors in ontology alignment. In this table, the number of ar-



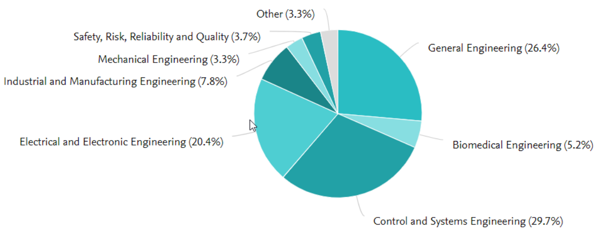
(a) The main subject areas of ontology alignment.



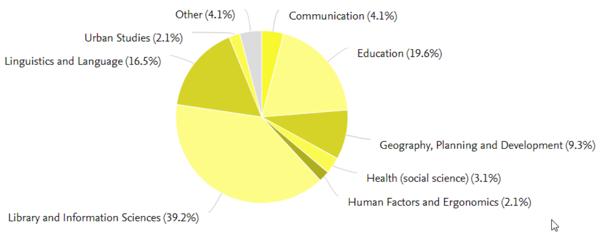
(b) Subcategories of computer science



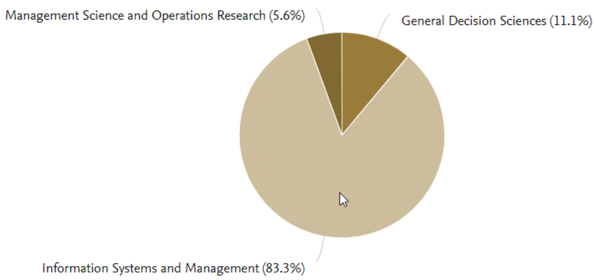
(c) Subcategories of mathematics



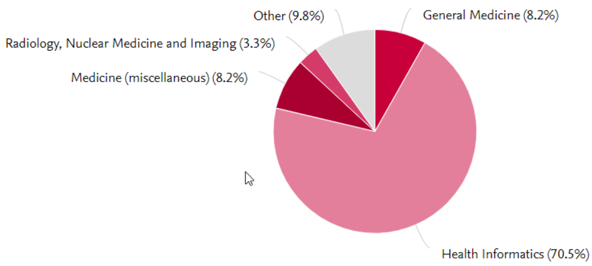
(d) Subcategories of engineering



(e) Subcategories of social science



(f) Subcategories of decision science



(g) Subcategories of medicine

Fig. 7. The Disciplines and their associated subcategories contributed to ontology alignment.

ticles co-authored with others, the total number of all co-authors, and the average citation per paper are displayed in the first to third columns, respectively. According to this table, J. Euzanat, P. Shvaiko, and C. Trojahn have the maximum number of collaborations in terms of the total number of co-authors.

Aside from the OAEI organizers, the developers of AgreementMaker and AgreementMakerLight (AML) [27], i.e., D. Faria, C. Pesquita, and I. Cruz, that are positioned at the top of Figure 8, have significant collaborations with each other. Also, D. Faria is one of the OAEI organizers, and along with C. Pesquita, have several collaborations with the OAEI community as well. Other members in this cluster have partnerships solely with each other.

Another dense cluster is placed at the bottom left of Figure 8. A closer look at the authors indicates that they are mainly from China, and have few collaborations with researchers outside of their country. The exceptions are S. Wang with 46, Juanzi Li with 48, and Yingjie Li with 40 co-authorship, including collaboration with other groups around the world. The main reason of such collaborations is that they are mainly studied outside of China, and had a better opportunity for international collaborations. Other authors from this community who had mainly collaborated intranationally are Y. Wang with 44, J. Wang with 37, and S. Zhang with 32 co-authorships.

The color of nodes in Figure 8 is proportionate to the average number of citations, where nodes closer to the yellow color have a higher number of citations per paper. It is interesting to see that the OAEI organizers, who vastly collaborate with other researchers, have the maximum citation average compared to others. Based on Table 3, Shvaiko with 116.58 citations per publication leads the author list in terms of average citations, followed by J. Euzenat with 70.92, and A. Ferrara with 31.76 citations per paper. Another important point is about the Chinese community. Among these researchers, the authors who have collaborations with international communities have higher average citations. This confirms previous studies that multinational research collaboration are associated with increased citations [39, 96]. In general, international research collaboration is recognized as a means of cultivating research quality, enhanced resource utilization, and high impact [39, 83]. It also has indirect strategic, economic, or political benefits [34]. In fact enlarging team sizes, increasing interdisciplinarity, and intensifying ties across institutional and geographic borders is a signal of field evolution from a solitary enterprise to an expanding social movement [39].

To further detect the communities of collaborations in ontology alignment, the co-authorship network in Figure 8 was subjected to a community detection algorithm, and the major collaborative communities were detected accordingly. There are quite a few community detection algorithms [30], and we choose Louvain algorithm [6], due to its speed, scalability, and simplicity [11, 24]. It is also one of the most popular community detection algorithms and has been implemented in many software and programming packages such as Gephi. Figure 9 plots the identified communities, where each community is depicted in a particular color. Six communities were identified by the algorithm, two of which are quite significant and include 75% of all researchers in this domain, and are shown in yellow and purple colors. The former is the OAEI organizers who have created a very large community. The thickness of edges of this network also indicates that researchers in the OAEI community greatly collaborate with each other. The other dominant community is the Chinese, in which the collaborations, though not as significant as the OAEI community, are remarkable. Also, this figure indicates that the collaborations between Chinese and OAEI communities are not of significance, and researchers primarily cooperate with other researchers from the same community.

5.2. *Country Collaboration*

In this section, the collaborations between countries are discussed and visualized. Figure 10 displays the co-authorship between different countries using a graph. The nodes of this graph are the countries with the size of node being proportionate to the number of publications collaborating with other countries, and the strength of the edge between each pair of countries is commensurate to the number of publications written jointly by authors of the corresponding countries. According to this figure, France, UK, Germany, the US, and Italy are respectively the top five countries having the most collaborative papers worldwide.

The top five countries have the most collaborative research outputs together. For French authors of ontology alignment publications, the international collaboration is mostly with Germany (with 13% share of all collaborative research outputs), Italy (12%) and the US (8%). For the UK, the co-authorship is most frequently with Italy (14%), Germany 13%, and the US (11%). German international co-authorships are mostly dom-

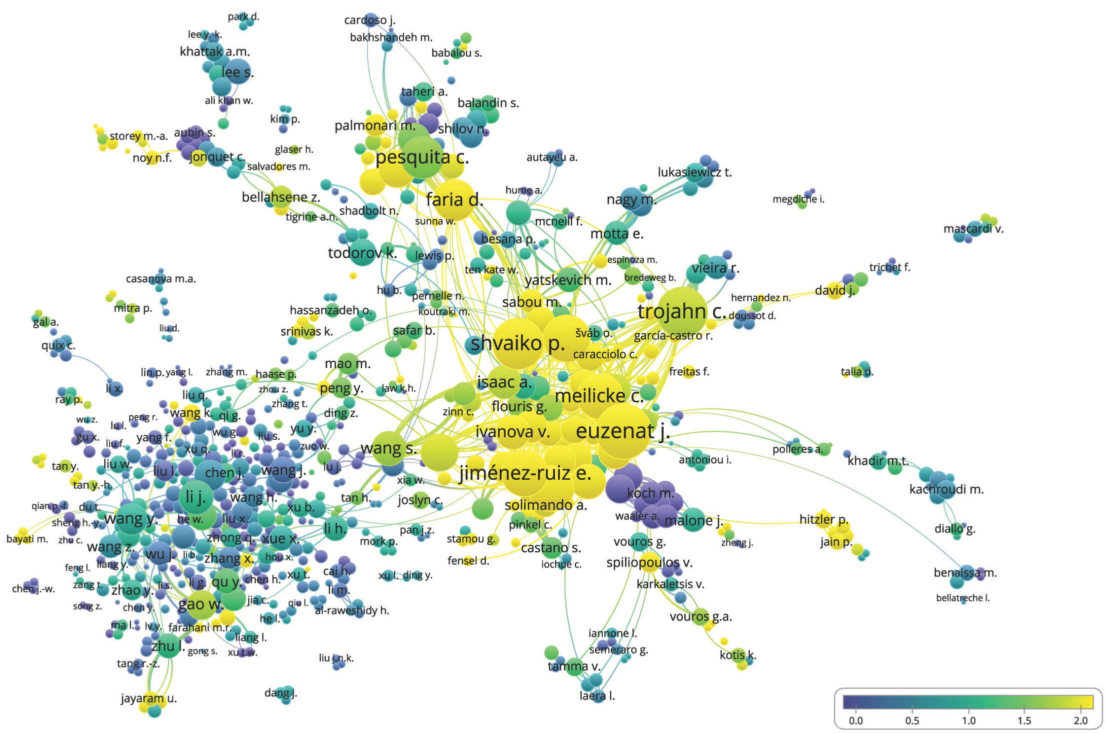


Fig. 8. Collaborations of authors in ontology alignment based on the bibliometric data 2001-2018. The size of nodes is proportionate to the number of collaborative publications by the associated author, the color of nodes represent the average citations of authors per paper, and the thickness of each edge is commensurate with the number of collaborative articles of the authors at the two ends.

Table 3

The ontology alignment researchers with the maximum number of collaborative publications. The first column tabulates the number of collaborative articles, the second column denotes the number of all collaborative researchers, and the third column is the average citation per publication.

Author	# Co-authored Articles	# All Co-authors	Average Citations
Euzenat J.	82	240	70.92
Shvaiko P.	63	205	116.58
Jiminez-Ruiz E.	63	168	16.86
Trojahn C.	59	204	13.98
StuckenSchmidt H.	59	177	18.95
Ferrara A.	56	163	31.76
Meilicke C.	54	160	22.72

inated by collaborations with France (14%), the UK (13%), and Italy (12%). US international ontology alignment collaborations are also most commonly with the UK (12%), then with researchers from France (10%) and Italy (9%). Italian ontology researchers most frequently collaborate with researchers from the UK (16%), then with colleagues from France (14.3%) and Germany (14%). In sum, these five countries are the most collaborative countries worldwide that mainly cooperate with each other.

We further analyze the collaboration between countries along with the number of their publications and citations they have received. In this regard, Figure 11 plots the collaboration between countries with the size of nodes being proportional to the number of the publications of the country, while Figure 12 shows the same graph with the size of nodes being commensurate to the average number of citations per publication. According to Figure 11, China has the maximum publications among other countries with 578 research outputs forming around 20% of all ontology alignment re-

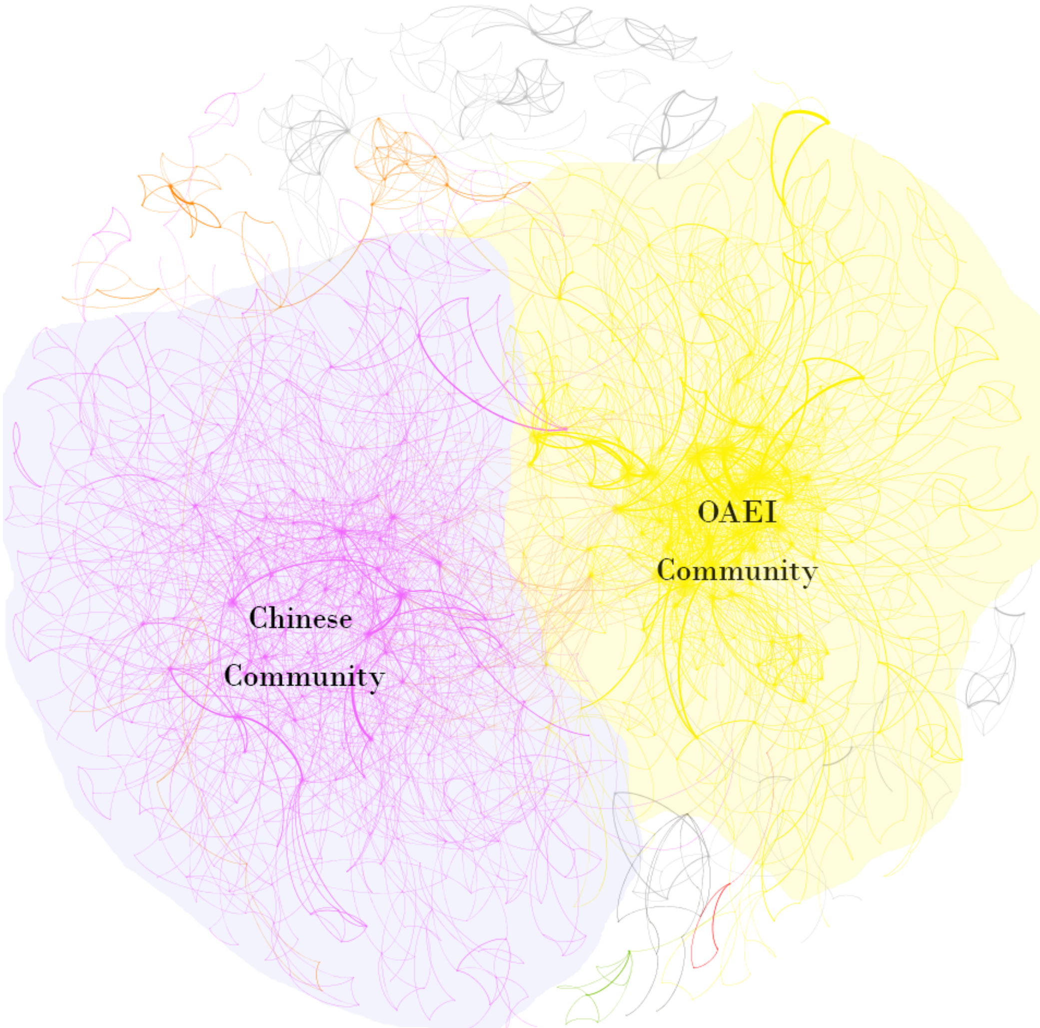


Fig. 9. Communities of collaborations in ontology alignment based on the bibliometric data 2001-2018.

search outputs. At the same time, the number of citations of this country is not commensurate to the number of publications. That is to say, the research outputs with at least one Chinese author have not gained enough attention. One of the primary reasons is the lack of international collaborations, as Figure 10 simply explains. The top five collaborative countries are readily seen to have more publications and citations among other countries. This corroborates the importance of collaborations since collaborations increase the visibility of research outputs to a wide audience, and the research studies thus get the attention they deserve. Another important point is the positive correlation between co-authorships and the size of publication outputs. It also makes sense since cooperation between researchers helps the use of the common wisdom, which is then emerged as more publications for the whole group.

5.3. International Collaboration

In this subsection, we take a closer look at the levels of collaborations in the recent six years. In this regard, we count the number of research outputs published by authors from different countries (international collaboration), people from different institutes within a country (only national collaboration), different researchers of an institute (only institutional collaboration), and single-authored ones (no collaboration). Figure 13 plots the share of ontology alignment published articles in each of these categories. According to this figure, as expected, the maximum collabo-

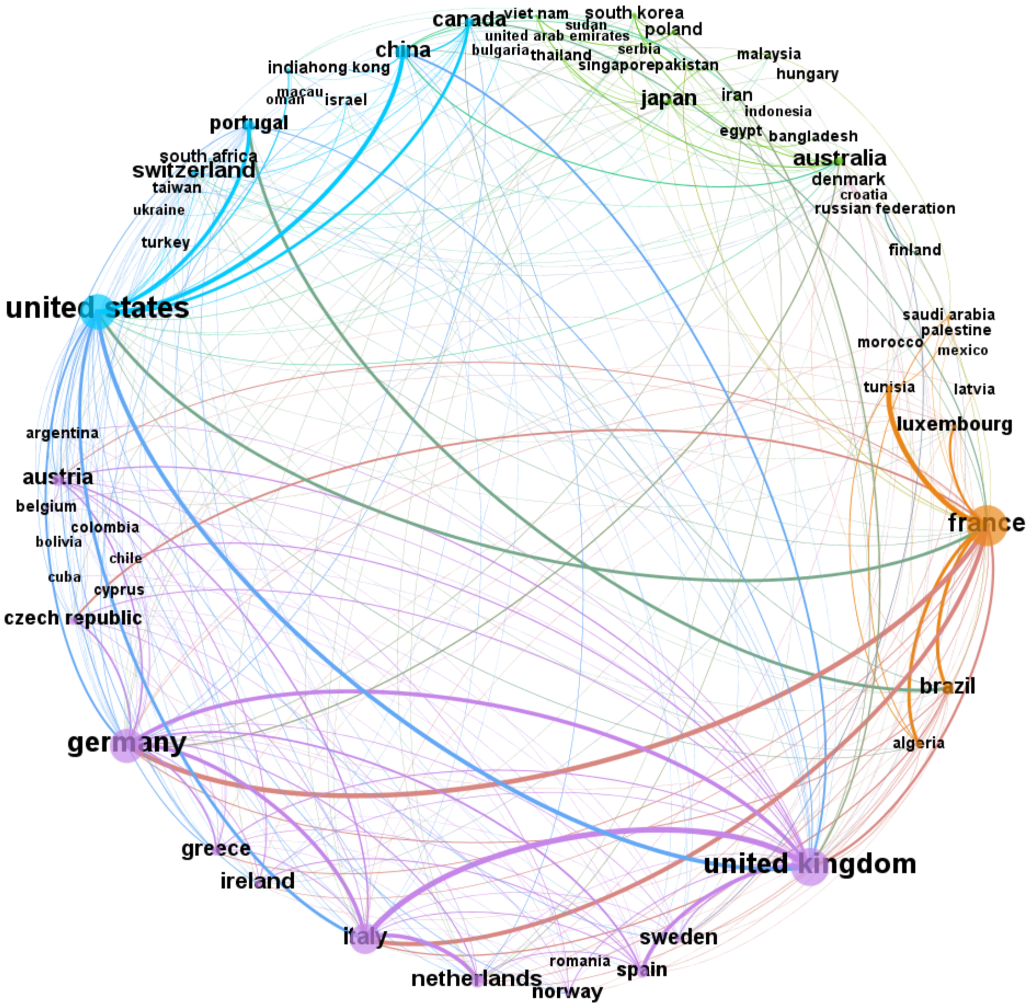


Fig. 10. Collaborations of countries in ontology alignment based on the bibliometric data 2001-2018. The size of nodes represents the number of all collaborative papers of researchers from the associated country.

rations are among the researchers from the same institutes, and the only national and only institutional collaborations have approximately similar shares of ontology alignment research outputs. Single-authored articles are at the bottom of this list and from around 10% of overall publications.

It is also seen from Figure 13 that international collaboration has increased in the recent six years, from 18.9% in 2013 to 27% in 2018. The maximum international collaborations are also in the year 2017, which constitute 28.9% of all ontology alignment research outputs. The collaborations inside the institutions have experienced a significant decrease, from 54.6% in 2013 to 45.5% in 2018. Similarly, the collaborations between institutions within a country have declined, while the single-authored articles have been growing. Ironically, the share of international collaborations (the most desired) and the share of no collaboration have been increased together over the last few years.

5.4. *Academic-Corporate Collaboration*

In this section, we consider the relationship between academic institutions and corporates based on ontology alignment published articles in the recent six years.

In an academic-corporate relationship, collaboration is of the essence for both sides. It helps the academia to ensure industrial relevance in its research [99], and on the other hand, it provides the opportunity of knowl-

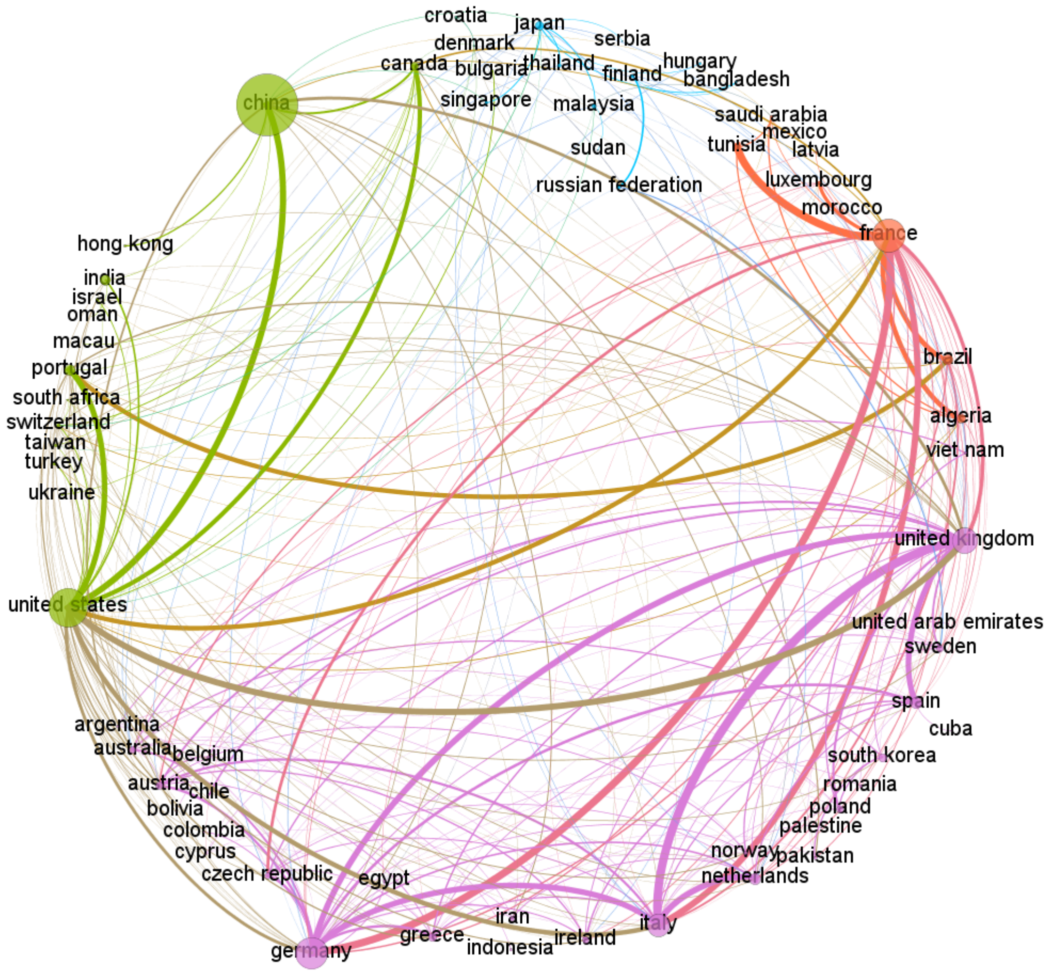


Fig. 11. Collaborations of countries in ontology alignment based on the bibliometric data 2001-2018. The size of nodes is proportionate to the number of published documents with at least one author from the associated country.

edge complementary and risk sharing with the corporates [2]. Figure 14 displays the share of published articles by the academic-corporate collaborations. According to this figure, a tiny portion of papers is published jointly by academia and corporates, with a maximum of 3% in 2014. The minimum portion is also from the years 2013 and 2018 with 1.7% of all ontology alignment published articles in the corresponding years. This figure elaborates that the relation between industry and academic institutions is not strong enough sine vast majority of collaborations is inside either academia or industry.

6. Contribution and Impact in Ontology Alignment

In this section, the impact of authors and countries that are active in ontology alignment is investigated. In this regard, we count the number of documents published per authors or countries and the number of citations of their documents to analyze their influence on ontology alignment.

6.1. Contribution and Impact of Authors in Ontology Alignment

In this subsection, the influence of authors on ontology alignment is investigated by counting the publications of top authors and their number of citations. Figure 15 plots the publication count of the top 10 au-

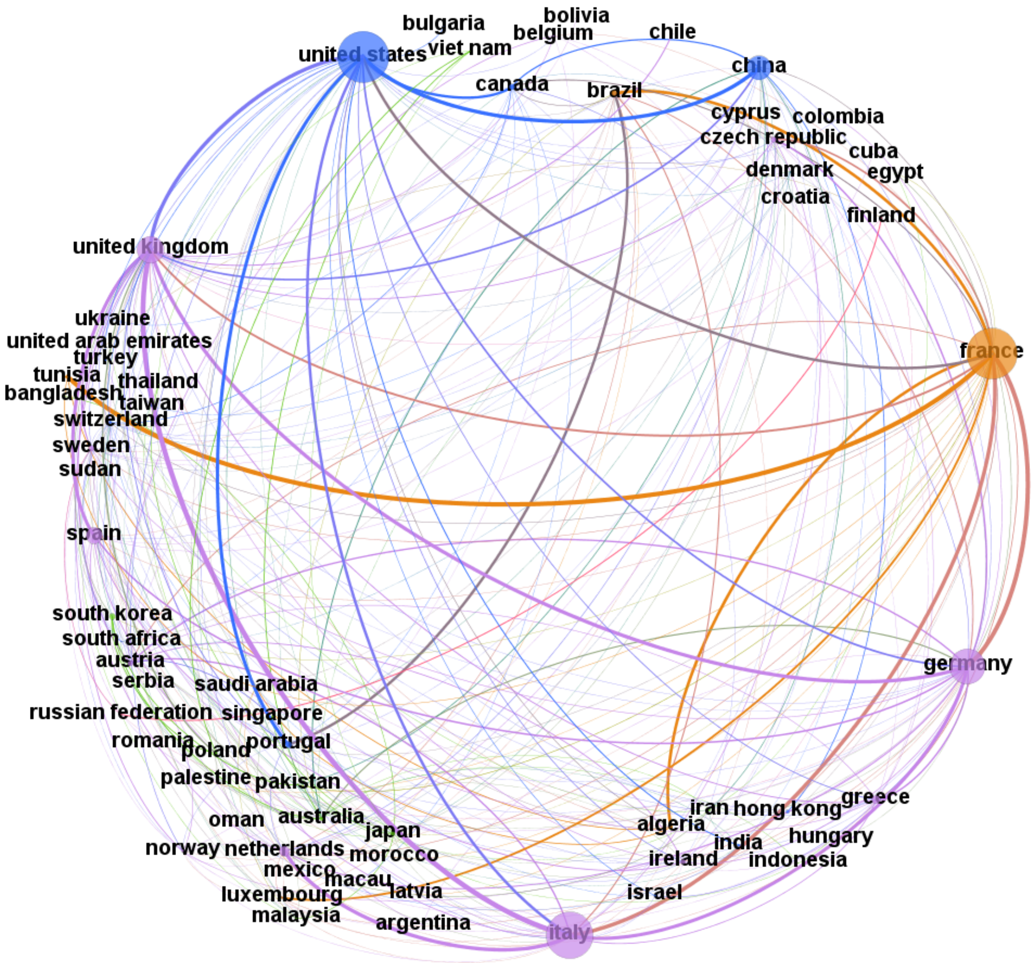


Fig. 12. Collaborations of countries in ontology alignment based on the bibliometric data 2001-2018. The size of nodes is proportionate to the number of citations of the corresponding country.

thors in ontology alignment. The top 10 authors have published around 308 articles spanning from 2001 to 2018, which constitute 10.3% of all publications in ontology alignment. According to Figure 15, J. Euzenat tops the list of authors with 70 publications in ontology alignment, and C. Trojahn with 50, E. Jimenez-Ruiz and H. Stuckenschmidt with 45 and P. Shvaiko with 38 research outputs follow.

We further analyze the ontology alignment authors in terms of their number of citations. In this regard, Figure 16 plots the top 10 authors with the maximum number of citations. According to this figure, J. Euzenat leads the list with 4,610 citations, and P. Shvaiko with 3,847, M. Scholemmer with 1,040, and Stuckenschmidt with 834 citations follow.

To provide a broader view of the author impacts on ontology alignment, we visualize the citation map of the author using VOSviewer. Figure 17 visualizes the citation map of ontology alignment authors, where the size of nodes and their colors are proportionate to the citations and average citations per paper, respectively. Also, the thickness of edges is proportional to the number of times the authors at the two ends have cited each other. Aside from the top authors shown in Figure 16, another critical observation from Figure 17 is that the researchers who are active in the OAEI have received more attention in comparison to others. The bottom-right region of this figure shows that the OAEI organizers and participants have higher average citations than other researchers.

Aside from the OAEI organizers and participants, several other researchers have been cited well. Y. Kalfoglou and M. Schorlemmer have a well-cited review paper published in 2003, and two methods for

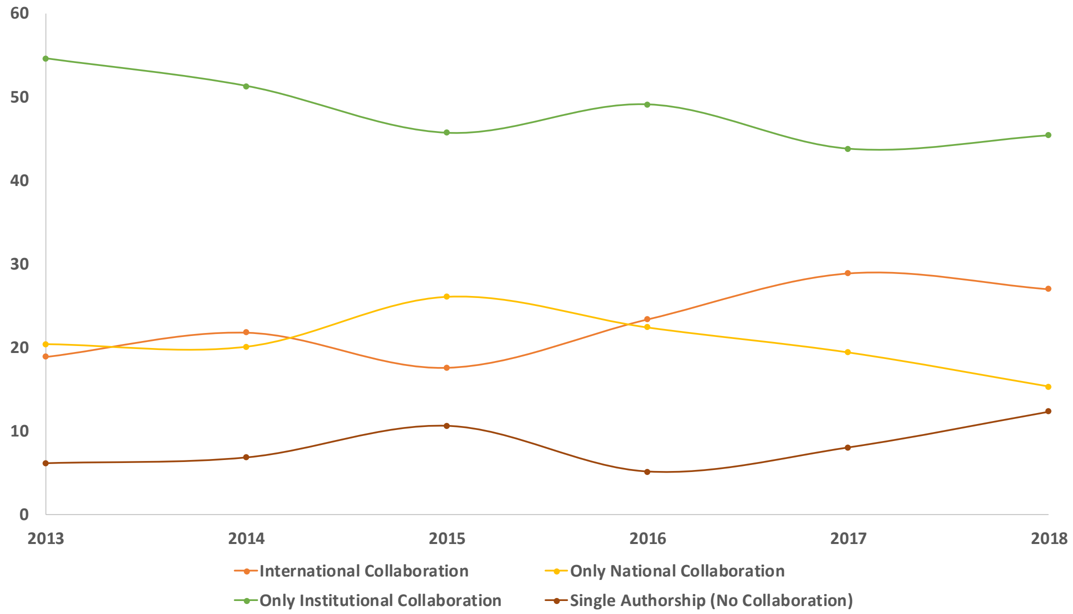


Fig. 13. The share of different types of Collaborations in ontology alignment in the recent six years.

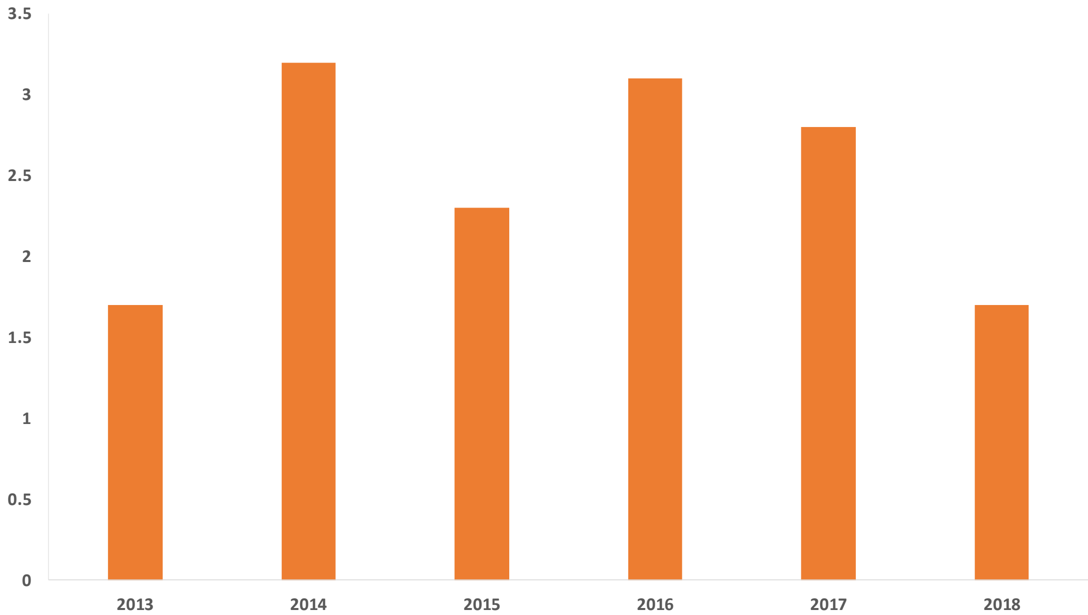


Fig. 14. The share of academic-corporate collaborations in ontology alignment.

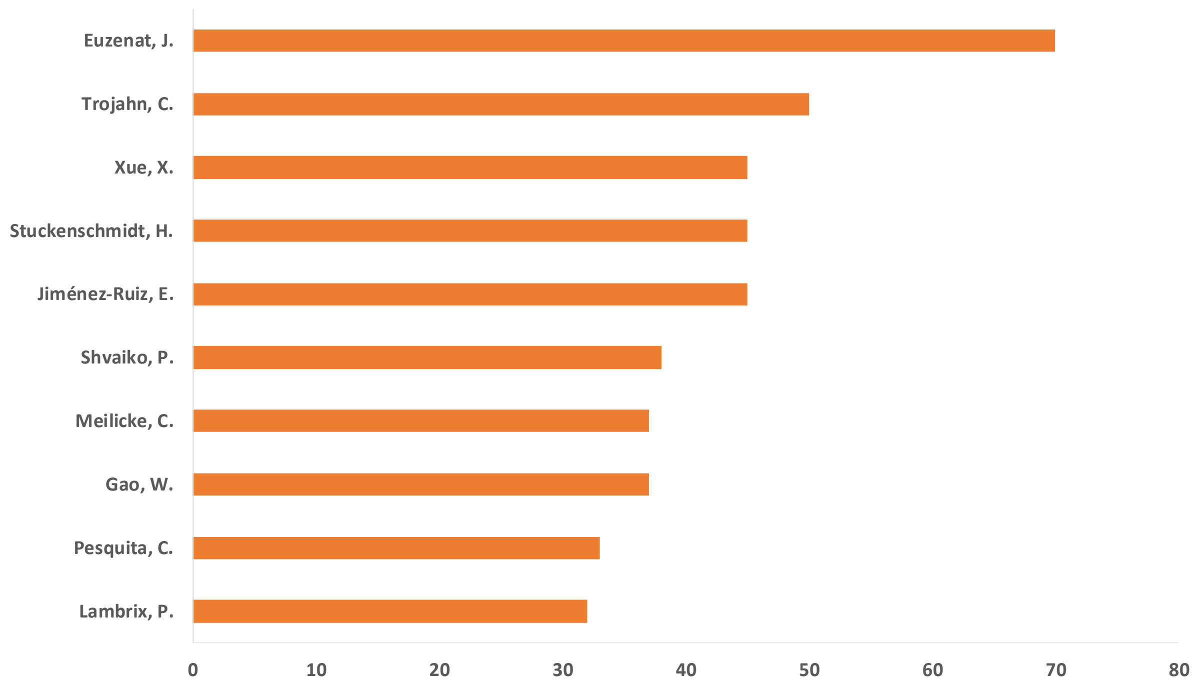


Fig. 15. The top 10 authors of ontology alignment in terms of their number of published documents based on the bibliometric data 2001-2018.

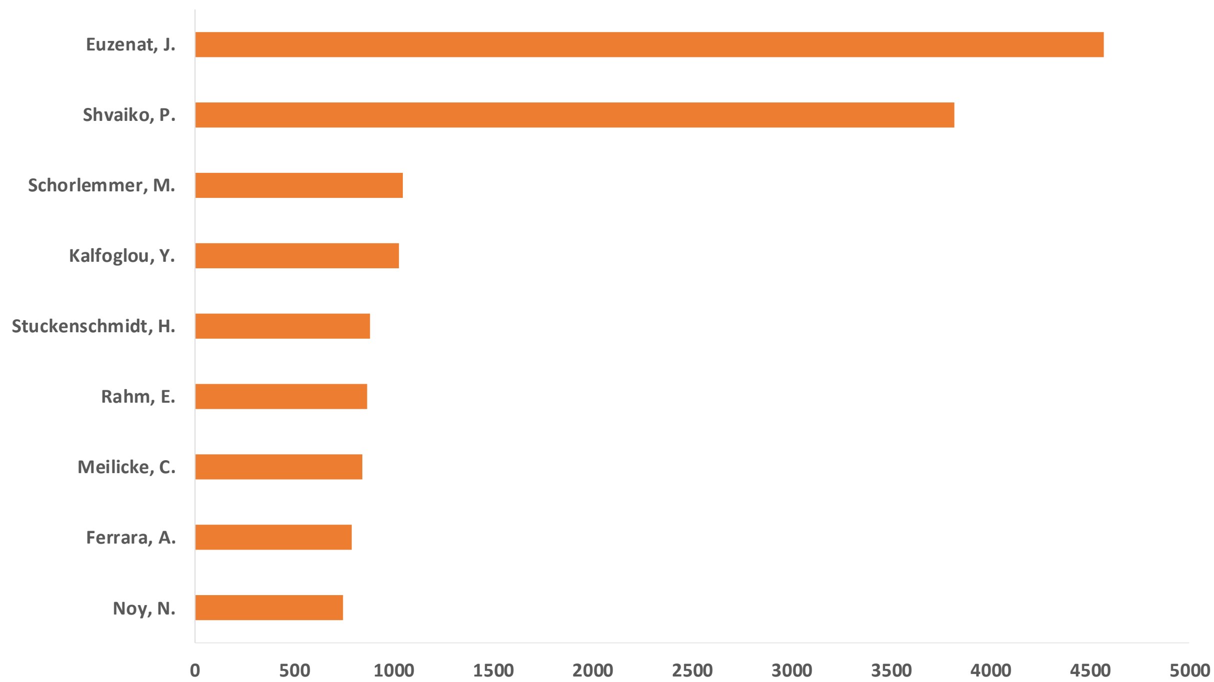


Fig. 16. The top 10 authors of ontology alignment in terms of their number of citations based on the bibliometric data 2001-2018.

ontology alignment published in 2002 and 2003 [49–51]. N. Noy has also conducted several fundamental research in ontology alignment in the first years of this century [66, 73–75]. Y. Li, J. Tang, and J. Li have developed (along with colleagues) the ontology alignment system, called RiMOM, where their seminal work was published in 2008 [92**?**]. M. Ehrig has published the book entitled *ontology alignment: bridging the semantic gap* in 2006, which has got more than 600 citations to date [22]. As one can readily realize, these researchers have received a great amount of attention because of their research in the first decade of this century. However, there are several well-cited researchers among the OAEI organizers, whose research studies are more recent. The examples are E. Jiminez-Ruiz who developed LogMap [45], D. Faria and C. Pesquita, who developed AML [27]. As a result, the attention of ontology alignment is more focused on the OAEI in recent years, and other researchers in this field have not drawn significant attention.

6.2. Contribution and Impact of Countries in Ontology Alignment

In this section, we analyze the impact and contributions of different countries on the evolution of ontology alignment. First, we visualize the number of ontology alignment publications for each country. Figure 18 displays the top 10 countries in terms of their publication count. According to this figure, China leads the list with 578 publications, which has a quite well difference with the second country, the US, with 360 publications in ontology alignment. France with 306, Germany with 295, and the UK with 234 publications in ontology alignment are the following countries in this list.

We further analyze the countries concerning their number of citations. Figure 19 shows the top 10 countries regarding the number of citations. Based on this figure, the US tops the list with 6,662 citations, and it is followed by France with 6563 and Italy with 6027 citations. China, while has the maximum number of ontology alignment citations, is the sixth country in this ranking with 3080 citations overall.

Besides, we visualize the citation network of countries using VOSviewer. Figure 20 illustrates the citation network where the nodes are the countries whose size are proportionate to the number of citations based on the bibliometric data 2001-2018. According to this figure, the number of citations of Chinese researchers is not commensurate with their number of publications: While they publish more than any country, they are ranked sixth concerning the number of citations. Italy has the highest average citations per publication with 29.54 for 204 documents and is followed by France with 21.45, Spain 20.04, the US with 18.51, Germany with 15.33, and the UK with 14.65 citations per publication. Moreover, based on Figure 20, the US has the maximum number of citations, followed by France, Italy, Germany, and the UK.

In addition, the community algorithm has been applied to the network in Figure 20. A community from this figure is a set of countries whose researchers refer to each others' studies more often. Basically, two major communities were detected that are shown in purple and green colors. One community comprises the US, UK, China, and Spain along with some other countries with less significant contribution to ontology alignment. Another community consists of France, Italy, and Germany as the major countries along with some less considerable ones such as the Netherlands and Portugal.

7. Conclusion and Discussion

In this paper, we revisited ontology alignment by using bibliometric analyses. In this regard, we made an inquiry in Scopus and extracted articles pertinent to ontology alignment. After inspecting the articles and excluding irrelevant items to ontology alignment, 2,975 articles remained based on which the quantitative analyses were carried out.

We conducted two classes of analyses, namely, semantic and structural. Semantic analysis entails the overall discovery of concepts, notions, and research lines flowing underneath ontology alignment, while the structural analysis provides a meta-level overview of the field by probing into the collaboration network and citation analysis in author and country levels. Each of these analyses was divided into two subcategories. In the semantic analysis, we first conducted a topic modeling on the extracted bibliometric data by subjecting title, keywords, and abstracts of articles to the latent Dirichlet allocation (LDA), a well-established statistical method for modeling topics. Although the topics were detected based merely on the frequency of the words in the articles, the identified topics readily referred to a problem that ontology alignment can address or a domain to which it has been applied. The other semantic analysis was thematic, wherein the number of annual publications, the types of research

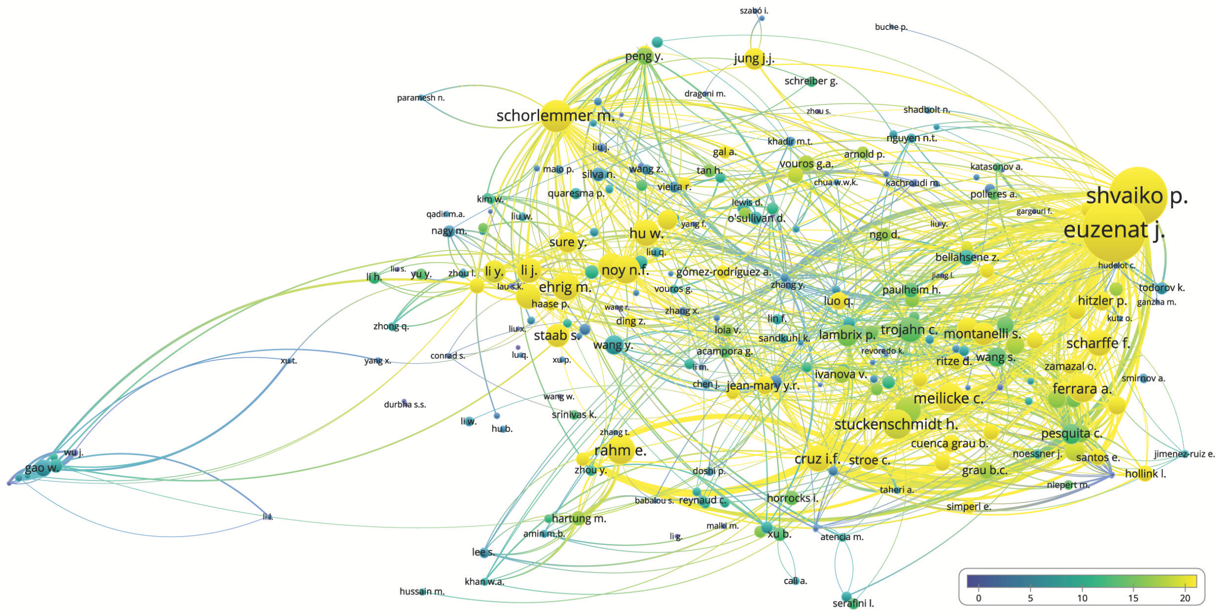


Fig. 17. The citation map of ontology alignment researchers, where the size of nodes is proportionate with the number of citations and its color represents the average citations per paper for the associated researcher. The thickness of edges in this network is proportional to the number of times the authors at the ends have cited each other.

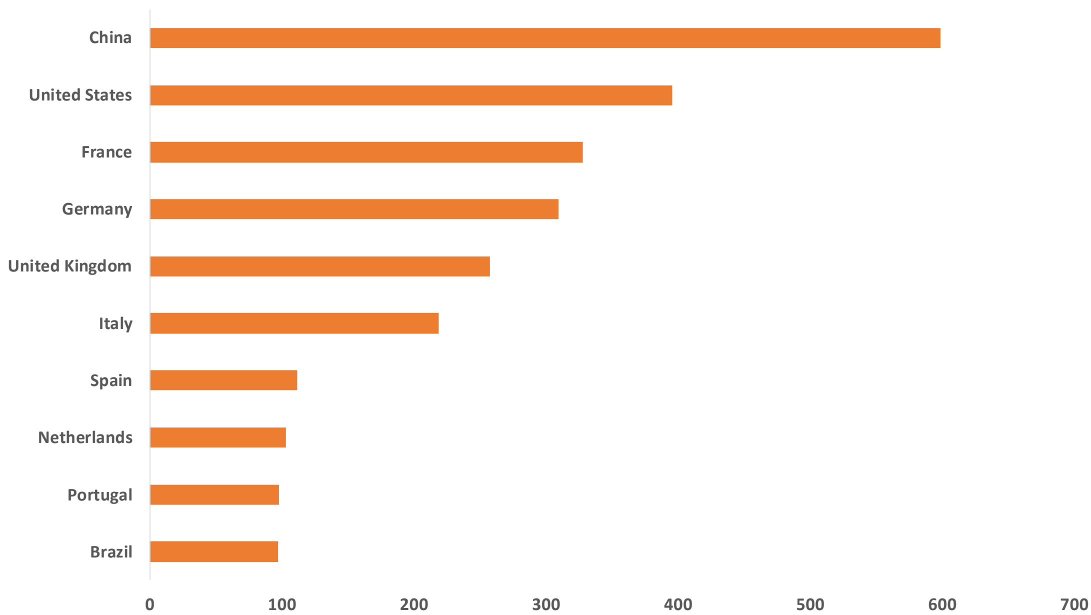


Fig. 18. The top 10 countries of ontology alignment in terms of their number of published documents based on the bibliometric data 2001-2018.

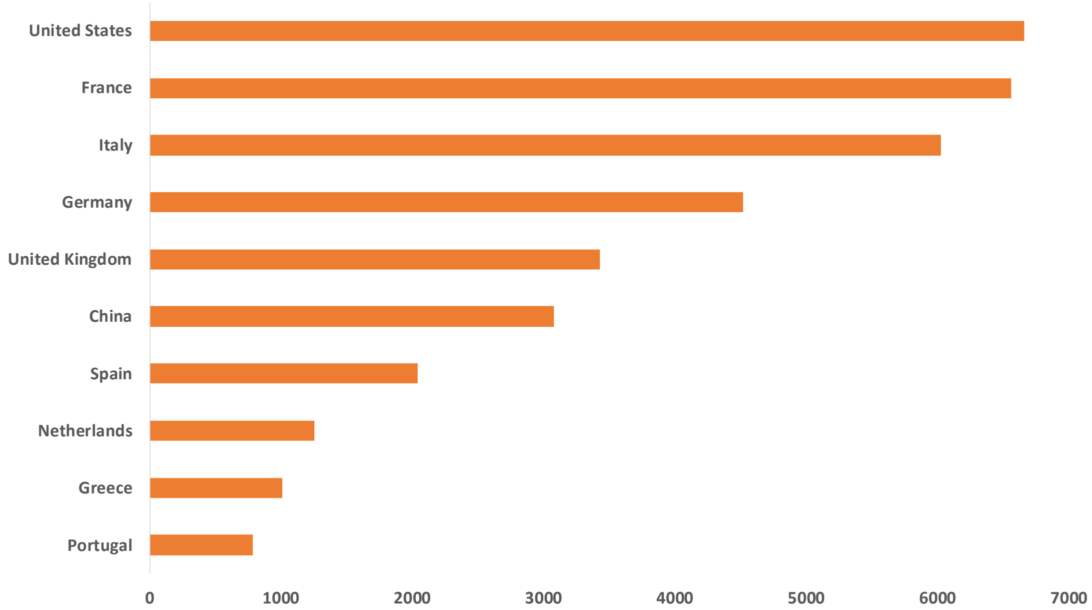


Fig. 19. The top 10 authors of ontology alignment in terms of their number of citations based on the bibliometric data 2001-2018.

outputs, the share of ontology alignment in top-cited articles and top journals, and contributing disciplines to ontology alignment were explained and discussed in detail.

The second class of analyses was structural, wherein we carried out two classes of probings. First, we studied the collaborations of ontology alignment researchers. We observed that international collaboration has been improved over the last few years. Also, the collaboration between academia and corporate were gauged, which was not significant enough. We also observed that ontology alignment researchers fall into two major communities. The first community comprises the OAEI organizers, and the second consists of the Chinese researchers. Although the researchers in these communities cooperate closely with each other, the communities, especially the Chinese, seem like isolated islands that do not interact with the other community. We also observed that, although the number of publications with at least one Chinese researcher is considerable, they do not get enough attention, possibly the attention that they deserve. Aside from authors, we also analyzed the collaborations between countries and realized that the top five most collaborative countries mostly work together more than any other country in the world. Another structural analysis was regarding the impact and contribution of researchers and countries for the field of ontology alignment. In these analyses, we identified the authors and countries with maximal influence on the field by counting their number of publications and citations. We also visualized their citation network, where we identified two hosts of countries that mostly cite each other.

We observed that the articles from the ontology matching workshop are not indexed by Web of Science (WoS). This workshop is particularly important for the ontology alignment community since highly-cited researchers from the field actively participate in the venue. Also, the articles from the workshop represent the state-of-the-art challenges and novelties in this domain. In addition, the OAEI contest is also a part of this workshop, wherein new challenges are introduced to be tackled and, new alignment systems participate to overcome those challenges.

Based on the contributions of ontology alignment to the top-cited articles and top journal percentiles, we perceived that ontology alignment is indeed a very essential field of study, the research outputs of the domain receive significant attention, and fundamental research studies are conducted and published in top journals. However, this field has only one venue, the ontology matching workshop, which, ironically, is not indexed by WoS. This calls for having more serious venues such as a dedicated journal for ontology alignment. The pertinent articles to ontology alignment are dispersed in various journals, while the conference pa-

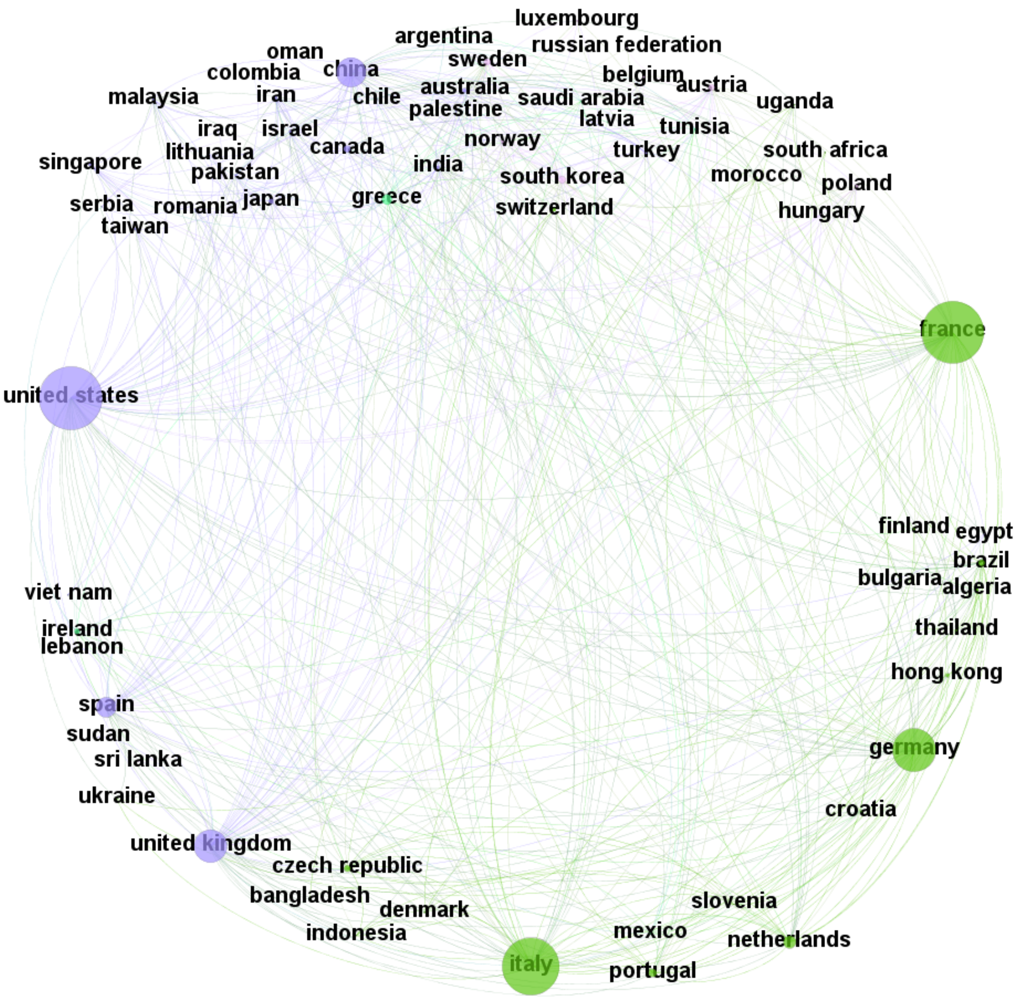


Fig. 20. The citation map of countries contributing to ontology alignment, where the size of nodes is proportionate to the number of citations.

pers are mainly published in the ontology matching workshop. Having a journal devoted to ontology alignment helps the research in this field to be well-focused in one venue. Also, there have been several special issues on ontology alignment over the last few years that are also hosted by different journals. A dedicated journal can be solely the host of these special issues.

The topics identified in the semantic analysis were directly related to the problems and applications related to ontology alignment. Some of these topics are well-established in ontology alignment, and the researchers of the field are well aware of it. However, there are several other topics that are totally neglected by researchers and the OAEI organizers in particular. By viewing the OAEI benchmarks, one readily gets the impression that ontology alignment is particularly useful for the biomedical domain since most of the tracks are from this domain, i.e., anatomy, disease and phenotype, and large biomedical. While it is indisputable that ontology alignment has been successfully applied to various biomedical ontologies, its use and applicability are not restricted to this domain. We encourage the OAEI organizers and other researchers in the field to prepare several other standard benchmarks for evaluating ontology alignment systems. The benchmarks can be from Semantic Web Services, agent-based modeling, knowledge graphs, and business processes.

We also identified two primary communities for ontology alignment, in both of which around 75% of ontology alignment researchers fall. The collaborations between these two communities are not significant, and researchers of these communities collaborate mainly with other researchers in the same community. Based on the citation analysis, we realized that the research of the Chinese community does not get considerable attention, while the number of research outputs of this community is significant. One reason for not getting enough attention is probably the lack of international collaboration in the Chinese community. Another reason might be that the contributions of this community do not address the current state-of-the-art challenges so that they do not draw enough attention. In any case, ontology alignment should take advantage of the potentials in the Chinese community. Thus, the collaborations between these two communities are strongly recommended so that the Chinese researcher would be familiar with the current challenges in the domain, they can disseminate their research to a wider audience, and can play a more vital role for progressing the field of ontology alignment. A potential direction for future research is to inspect the research of these two communities to see precisely what they are working on, and then find the potential topics on which these communities can cooperate with each other.

We discussed and visualized the impact of researchers and countries on ontology alignment as well. We observed that the OAEI organizers and participants get a considerable amount of attention. There are several other researchers with significant attention, but their research studies were mainly carried out in the first decade of this century. Thus, if researchers want to get attention, they must be involved in the OAEI community and participate actively in the contest.

Another critical observation from the analyses in this article is the insignificant collaborations between academia and corporate. We observed that the academic-corporate relations constitute on average, around 2% of all publications in ontology alignment. One possible reason is that the teams inside the enterprises have the ability to resolve the problems. However, that does not explain the inconsiderable relation between academia and industry. The more realistic reason is that the companies have not realized that ontology alignment can enhance their business functions. This is also coming from the fact that most of the standard benchmarks are restricted to particular problems and domain, while ontology alignment can be widely used to address disparate issues. One way to increase the impact and use of ontology alignment is to conduct several qualitative case studies to show how ontology alignment can automate the manual procedures and consequently, increase the profits of the companies. Another avenue for the progress of ontology matching is to dedicate more research funding to the applicability of ontology alignment and find untapped domains and problems to which ontology alignment is a potential solution. Such projects create new benchmarks for the OAEI and extend the ontology alignment applicability so that a broader audience can understand the importance of the field. This also can help overcome the slow progress we have observed from 2013.

References

[1] Rama Akkiraju, Joel Farrell, John A Miller, Meenakshi Nagarajan, Amit P Sheth, and Kunal Verma. 2005. Web service semantics-wsdl-s. (2005).

[2] Omar Al-Tabbaa and Samuel Ankrah. 2018. âĂŸEngineeredâĂŹUniversity-Industry Collaboration: A Social Capital Perspective. *European Management Review* (2018).

[3] Mathieu Bastian, Sebastien Heymann, and Mathieu Jacomy. 2009. Gephi: an open source software for exploring and manipulating networks. In *Third international AAAI conference on weblogs and social media*.

[4] Yacine Belhoul, Mohammed Haddad, Ahmed Gater, Daniela Grigori, Hamamache Kheddouci, and Mokrane Bouzeghoub. 2013. Spectral graph approach for process model matchmaking. In *2013 IEEE International Conference on Services Computing*. IEEE, 408–415.

[5] David M Blei, Andrew Y Ng, and Michael I Jordan. 2003. Latent dirichlet allocation. *Journal of machine Learning research* 3, Jan (2003), 993–1022.

[6] Vincent D Blondel, Jean-Loup Guillaume, Renaud Lambiotte, and Etienne Lefebvre. 2008. Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment* 2008, 10 (10 2008), P10008.

[7] Olivier Bodenreider. 2004. The unified medical language system (UMLS): integrating biomedical terminology. *Nucleic acids research* 32, suppl_1 (2004), D267–D270.

[8] Lindell Bromham, Russell Dinnage, and Xia Hua. 2016. Interdisciplinary research has consistently lower funding success. *Nature* 534, 7609 (2016), 684.

[9] Mark H Burstein. 2003. The many faces of mapping and translation for semantic web services. In *Proceedings of the Fourth International Conference on Web Information Systems Engineering, 2003. WISE 2003.* IEEE, 261–268.

[10] Cristian Candia, C Jara-Figueroa, Carlos Rodriguez-Sickert, Albert-László Barabási, and César A Hidalgo. 2019. The universal decay of collective memory and attention. *Nature Human Behaviour* 3, 1 (2019), 82.

[11] Pravin Chopade and Justin Zhan. 2016. A framework for community detection in large networks using game-theoretic modeling. *IEEE Transactions on Big Data* 3, 3 (2016), 276–288.

[12] Caspar Chorus and Ludo Waltman. 2016. A large-scale analysis of impact factor biased journal self-citations. *PLoS One* 11, 8 (2016), e0161021.

[13] Susan Cozzens, Sonia Gatchair, Jongseok Kang, Kyung-Sup Kim, Hyuck Jai Lee, Gonzalo Ordóñez, and Alan Porter. 2010. Emerging technologies: quantitative identification and measurement. *Technology Analysis & Strategic Management* 22, 3 (2010), 361–376.

[14] Remco Dijkman, Marlon Dumas, Luciano Garcia-Banuelos, and Reina Kaarik. 2009. Aligning business process models. In *2009 IEEE International Enterprise Distributed Object Computing Conference*. IEEE, 45–53.

[15] Ying Ding, Gobinda G Chowdhury, and Schubert Foo. 2001. Bibliometric cartography of information retrieval research by using co-word analysis. *Information processing & management* 37, 6 (2001), 817–842.

[16] Ying Ding and Schubert Foo. 2002. Ontology research and development. Part 1-a review of ontology generation. *Journal of information science* 28, 2 (2002), 123–136.

[17] Ying Ding and Schubert Foo. 2002. Ontology research and development. part 2-a review of ontology mapping and evolving. *Journal of information science* 28, 5 (2002), 375–388.

[18] AnHai Doan, Jayant Madhavan, Pedro Domingos, and Alon Halevy. 2002. Learning to map between ontologies on the semantic web. In *Proceedings of the 11th international conference on World Wide Web*. AcM, 662–673.

[19] AnHai Doan, Jayant Madhavan, Pedro Domingos, and Alon Halevy. 2004. Ontology matching: A machine learning approach. In *Handbook on ontologies*. Springer, 385–403.

[20] Zlatan Dragisic, Valentina Ivanova, Huanyu Li, and Patrick Lambrix. 2017. Experiences from the anatomy track in the ontology alignment evaluation initiative. *Journal of biomedical semantics* 8, 1 (2017), 56.

[21] Amir Ebrahimi Fard and Cunningham Scott. [n. d.]. Assessing the Readiness of the Academia in the Topic of False and Unverified Information. *To be appeared in ACM Journal of Data and Information Quality (JDIQ)* ([n. d.]).

[22] Marc Ehrig. 2006. *Ontology alignment: bridging the semantic gap*. Vol. 4. Springer Science & Business Media.

[23] Islam Elgedawy. 2003. A conceptual framework for web services semantic discovery. In *OTM Confederated International Conferences" On the Move to Meaningful Internet Systems"*. Springer, 1004–1016.

[24] Scott Emmons, Stephen Kobourov, Mike Gallant, and Katy Börner. 2016. Analysis of network clustering algorithms and cluster quality metrics at scale. *PloS one* 11, 7 (2016), e0159161.

[25] Vadim Ermolayev and Maxim Davidovsky. 2012. Agent-based ontology alignment: basics, applications, theoretical foundations, and demonstration. In *Proceedings of the 2nd International Conference on Web Intelligence, Mining and Semantics*. ACM, 3.

[26] Jérôme Euzenat, Pavel Shvaiko, et al. 2007. *Ontology matching*. Vol. 18. Springer.

[27] Daniel Faria, Catia Pesquita, Emanuel Santos, Matteo Palmonari, Isabel F Cruz, and Francisco M Couto. 2013. The agreementmakerlight ontology matching system. In *OTM Confederated International Conferences" On the Move to Meaningful Internet Systems"*. Springer, 527–541.

[28] Aissa Fellah, Mimoun Malki, and Atilla Elçi. 2016. Web services matchmaking based on a partial ontology alignment. *Int. Journal of Information Technology and Computer Science (IJITCS)* 8, 6 (2016).

[29] Janina Fengel. 2014. Semantic technologies for aligning heterogeneous business process models. *Business Process Management Journal* 20, 4 (2014), 549–570.

[30] Santo Fortunato. 2010. Community detection in graphs. *Physics Reports* 486, 3-5 (2 2010), 75–174.

[31] Santo Fortunato, Carl T Bergstrom, Katy Börner, James A Evans, Dirk Helbing, Staša Milojević, Alexander M Petersen, Filippo Radicchi, Roberta Sinatra, Brian Uzzi, et al. 2018. Science of science. *Science* 359, 6379 (2018), eaao0185.

[32] Morgan R Frank, Dashun Wang, Manuel Cebrian, and Iyad Rahwan. 2019. The evolution of citation graphs in artificial intelligence research. *Nature Machine Intelligence* 1, 2 (2019), 79.

[33] Ahmed Gater, Daniela Grigori, and Mokrane Bouzeghoub. 2010. Complex mapping discovery for semantic process model alignment. In *Proceedings of the 12th International Conference on Information Integration and Web-based Applications & Services*. ACM, 317–324.

[34] Luke Georghiou. 1998. Global cooperation in research. *Research policy* 27, 6 (1998), 611–626.

[35] Wolfgang Glänzel. 2000. Science in Scandinavia: A bibliometric approach. *Scientometrics* 48, 2 (2000), 121–150.

[36] Roger Guimera, Brian Uzzi, Jarrett Spiro, and Luis A Nunes Amaral. 2005. Team assembly mechanisms determine collaboration network structure and team performance. *Science* 308, 5722 (2005), 697–702.

[37] Ian Harrow, Ernesto Jiménez-Ruiz, Andrea Splendiani, Martin Romacker, Peter Woollard, Scott Markel, Yasmin Alam-Faruque, Martin Koch, James Malone, and Arild Waaler. 2017. Matching disease and phenotype ontologies in the ontology alignment evaluation initiative. *Journal of biomedical semantics* 8, 1 (2017), 55.

[38] Yulan He and Siu Cheung Hui. 2002. Mining a web citation database for author co-citation analysis. *Information processing & management* 38, 4 (2002), 491–508.

[39] David Hsiehchen, Magdalena Espinoza, and Antony Hsieh. 2015. Multinational teams and diseconomies of scale in collaborative research. *Science advances* 1, 8 (2015), e1500211.

[40] Jingshan Huang, Jiangbo Dang, José M Vidal, and Michael N Huhns. 2007. Ontology matching using an artificial neural network to learn weights. In *IJCAI workshop on semantic Web for collaborative knowledge acquisition*, Vol. 106.

[41] Bernhard G Humm and Janina Fengel. 2012. Semantics-based business process model similarity. In *International Conference on Business Information Systems*. Springer, 36–47.

[42] Prateek Jain, Pascal Hitzler, Amit P Sheth, Kunal Verma, and Peter Z Yeh. 2010. Ontology alignment for linked open data. In *International semantic web conference*. Springer, 402–417.

[43] Ernesto Jiménez-Ruiz, Asan Agibetov, Matthias Samwald, and Valerie Cross. 2018. Breaking-down the Ontology Alignment Task with a Lexical Index and Neural Embeddings. *arXiv preprint arXiv:1805.12402* (2018).

[44] Ernesto Jiménez-Ruiz, Asan Agibetov, Matthias Samwald, and Valerie Cross. 2018. Breaking-down the Ontology Alignment Task with a Lexical Index and Neural Embeddings. *arXiv preprint arXiv:1805.12402* (2018).

[45] Ernesto Jiménez-Ruiz and Bernardo Cuenca Grau. 2011. Logmap: Logic-based and scalable ontology matching. In *International Semantic Web Conference*. Springer, 273–288.

[46] Ernesto Jiménez-Ruiz, Christian Meilicke, Bernardo Cuenca Grau, and Ian Horrocks. 2013. Evaluating Mapping Repair Systems with Large Biomedical Ontologies. *Description Logics* 13 (2013), 246–257.

[47] Jason J Jung. 2009. Semantic business process integration based on ontology alignment. *Expert Systems with Applications* 36, 8 (2009), 11013–11020.

[48] Yannis Kalfoglou and Marco Schorlemmer. 2003. IF-Map: An ontology-mapping method based on information-flow theory. In *Journal on data semantics I*. Springer, 98–127.

[49] Yannis Kalfoglou and Marco Schorlemmer. 2003. IF-Map: An ontology-mapping method based on information-flow theory. In *Journal on data semantics I*. Springer, 98–127.

[50] Yannis Kalfoglou and Marco Schorlemmer. 2003. IF-Map: An ontology-mapping method based on information-flow theory. In *Journal on data semantics I*. Springer, 98–127.

[51] Yannis Kalfoglou and Marco Schorlemmer. 2003. Ontology mapping: the state of the art. *The knowledge engineering review* 18, 1 (2003), 1–31.

[52] Qing Ke, Emilio Ferrara, Filippo Radicchi, and Alessandro Flammini. 2015. Defining and identifying sleeping beauties in science. *Proceedings of the National Academy of Sciences* 112, 24 (2015), 7426–7431.

[53] Hak Lae Kim, Alexandre Passant, John G Breslin, Simon Scerri, and Stefan Decker. 2008. Review and alignment of tag ontologies for semantically-linked data in collaborative tagging spaces. In *2008 IEEE International Conference on Semantic Computing*. IEEE, 315–322.

[54] Christopher Klinkmüller, Ingo Weber, Jan Mendling, Henrik Leopold, and André Ludwig. 2013. Increasing recall of process model matching by improved activity label matching. In *Business process management*. Springer, 211–218.

[55] Manel Kolli and Zizette Boufaida. 2011. A description logics formalization for the ontology matching. *Procedia Computer Science* 3 (2011), 29–35.

[56] Aliya Kuzhabekova and Jennifer Kuzma. 2014. Mapping the emerging field of genome editing. *Technology Analysis & Strategic Management* 26, 3 (2014), 321–352.

[57] Loredana Laera, Valentina Tamma, Jérôme Euzenat, Trevor Bench-Capon, and Terry Payne. 2006. Reaching agreement over ontology alignments. In *International Semantic Web Conference*. Springer, 371–384.

[58] Nadira Lammari, Jacky Akoka, and Isabelle Comyn-Wattiau. 2003. Supporting database evolution: Using ontologies matching. In *International Conference on Object-Oriented Information Systems*. Springer, 284–288.

[59] Loet Leydesdorff and Ping Zhou. 2007. Nanotechnology as a field of science: Its delineation in terms of journals and patents. *Scientometrics* 70, 3 (2007), 693–713.

[60] Lu Liu, Yang Wang, Roberta Sinatra, C Lee Giles, Chaoming Song, and Dashun Wang. 2018. Hot streaks in artistic, cultural, and scientific careers. *Nature* 559, 7714 (2018), 396.

[61] Ajoy Mallik and Nripendranath Mandal. 2014. Bibliometric analysis of global publication output and collaboration structure study in microRNA research. *Scientometrics* 98, 3 (2014), 2011–2037.

[62] Ming Mao, Yefei Peng, and Michael Spring. 2010. An adaptive ontology mapping approach with neural network based constraint satisfaction. *Web Semantics: Science, Services and Agents on the World Wide Web* 8, 1 (2010), 14–25.

[63] David Martin, Mark Burstein, Jerry Hobbs, Ora Lassila, Drew McDermott, Sheila McIlraith, Srini Narayanan, Massimo Paolucci, Bijan Parsia, Terry Payne, et al. 2004. OWL-S: Semantic markup for web services. *W3C member submission* 22, 4 (2004).

[64] Christian Meilicke. 2011. *Alignment incoherence in ontology matching*. Ph.D. Dissertation. Universität Mannheim.

[65] Christian Meilicke, Heiner Stuckenschmidt, and Andrei Tamilin. 2007. Repairing ontology mappings. In *AAAI*, Vol. 3. 6.

[66] Prasenjit Mitra, Natasha F Noy, and Anuj Rattan Jaiswal. 2005. Omen: A probabilistic ontology mapping tool. In *International Semantic Web Conference*. Springer, 537–547.

[67] Majid Mohammadi. 2019. Bayesian Evaluation and Comparison of Ontology Alignment Systems. *IEEE Access* (2019).

[68] Majid Mohammadi, Amir Ahooye Atashin, Wout Hofman, and Yaohua Tan. 2018. Comparison of ontology alignment systems across single matching task via the McNemarâĂŹs test. *ACM Transactions on Knowledge Discovery from Data (TKDD)* 12, 4 (2018), 51.

[69] Majid Mohammadi, Wout Hofman, and Yao-Hua Tan. 2019. A comparative study of ontology matching systems via inferential statistics. *IEEE Transactions on Knowledge and Data Engineering* 31, 4 (2019), 615–628.

[70] Majid Mohammadi, Wout Hofman, and Yao-Hua Tan. 2019. Simulated annealing-based ontology matching. *ACM Transactions on Management Information Systems (TMIS)* 10, 1 (2019), 3.

[71] Mathias Niepert, Christian Meilicke, and Heiner Stuckenschmidt. 2010. A probabilistic-logical framework for ontology matching. In *Twenty-Fourth AAAI Conference on Artificial Intelligence*.

[72] Henrik Nottelmann and Umberto Straccia. 2006. A probabilistic, logic-based framework for automated web directory alignment. In *Soft Computing in Ontologies and Semantic Web*. Springer, 47–77.

[73] Natalya Fridman Noy and Mark A Musen. 1999. SMART: Automated support for ontology merging and alignment. In *Proc. of the 12th Workshop on Knowledge Acquisition, Modelling, and Management (KAW'99), Banf, Canada*. Citeseer.

[74] Natalya F Noy and Mark A Musen. 2003. The PROMPT suite: interactive tools for ontology merging and mapping. *International journal of human-computer studies* 59, 6 (2003), 983–1024.

[75] Natalya Fridman Noy, Mark A Musen, et al. 2000. Algorithm and tool for automated ontology merging and alignment. In *Proceedings of the 17th National Conference on Artificial Intelligence (AAAI-00). Available as SMI technical report SMI-2000-0831*, Vol. 115. sn.

[76] Leo Obrst, Robert E Wray, and Howard Liu. 2001. Ontological engineering for B2B E-commerce. In *Proceedings of the international conference on Formal Ontology in Information Systems-Volume 2001*. ACM, 117–126.

[77] Lorena Otero-Cerdeira, Francisco J Rodríguez-Martínez, and Alma Gómez-Rodríguez. 2015. Ontology matching: A literature review. *Expert Systems with Applications* 42, 2 (2015), 949–971.

[78] Catia Pesquita, Daniel Faria, Emanuel Santos, and Francisco M Couto. 2013. To repair or not to repair: reconciling correctness and coherence in ontology reference alignments. In *Proc. 8th*

ISWC ontology matching workshop (OM), Sydney (AU), page this volume.

[79] Alexander M Petersen, Dinesh Majeti, Kyeongan Kwon, Mohammed E Ahmed, and Ioannis Pavlidis. 2018. Cross-disciplinary evolution of the genomics revolution. *Science advances* 4, 8 (2018), eaat4211.

[80] Guy Pierra. 2004. The PLIB ontology-based approach to data integration. In *Building the Information Society*. Springer, 13–18.

[81] Alan L Porter, Jan Youtie, Philip Shapira, and David J Schoeneck. 2008. Refining search terms for nanotechnology. *Journal of nanoparticle research* 10, 5 (2008), 715–728.

[82] Martin F Porter. 1980. An algorithm for suffix stripping. *Program* 14, 3 (1980), 130–137.

[83] George Radda. 2002. Biomedical research and international collaboration. *Science* 295, 5554 (2002), 445–446.

[84] Mariano Rubiolo, María Laura Caliusco, Georgina Stegmayer, Mauricio Coronel, and M Gareli Fabrizi. 2012. Knowledge discovery through ontology matching: An approach based on an Artificial Neural Network model. *Information Sciences* 194 (2012), 107–119.

[85] Ulrich Schmoch and Torben Schubert. 2008. Are international co-publications an indicator for quality of scientific research? *Scientometrics* 74, 3 (2008), 361–377.

[86] Marco Schorlemmer, Yannis Kalfoglou, and Manuel Atencia. 2007. A formal foundation for ontology-alignment interaction models. *International Journal on Semantic Web and Information Systems (IJSWIS)* 3, 2 (2007), 50–68.

[87] Aviv Segev and Eran Toch. 2009. Context-based matching and ranking of web services for composition. *IEEE Transactions on Services Computing* 3 (2009), 210–222.

[88] Per O Seglen and Dag W Aksnes. 2000. Scientific productivity and group size: A bibliometric analysis of Norwegian microbiological research. *Scientometrics* 49, 1 (2000), 125–143.

[89] Pavel Shvaiko and Jérôme Euzenat. 2013. Ontology matching: state of the art and future challenges. *IEEE Transactions on knowledge and data engineering* 25, 1 (2013), 158–176.

[90] Nuno Silva, João Rocha, and José Cardoso. 2003. E-Business interoperability through ontology semantic mapping. In *Working Conference on Virtual Enterprises*. Springer, 315–322.

[91] Evren Sirin, Bijan Parsia, and James Hendler. 2004. Filtering and selecting semantic web services with interactive composition techniques. *IEEE Intelligent Systems* 19, 4 (2004), 42–49.

[92] Jie Tang, Juanzi Li, Bangyong Liang, Xiaotong Huang, Yi Li, and Kehong Wang. 2006. Using Bayesian decision for ontology mapping. *Web Semantics: Science, Services and Agents on the World Wide Web* 4, 4 (2006), 243–262.

[93] Mike Thelwall, Stefanie Haustein, Vincent Larivière, and Cassidy R Sugimoto. 2013. Do altmetrics work? Twitter and ten other social web services. *PloS one* 8, 5 (2013), e64841.

[94] Elodie Thiéblin, Ollivier Haemmerlé, Nathalie Hernandez, and Cassia Trojahn. 2019. Survey on complex ontology matching. *Semantic Web* Preprint (2019), 1–39.

[95] Nees van Eck and Ludo Waltman. 2009. Software survey: VOSviewer, a computer program for bibliometric mapping. *Scientometrics* 84, 2 (2009), 523–538.

[96] A Van Raan. 1998. The influence of international collaboration on the impact of research results: Some simple mathematical considerations concerning the role of self-citations. *Scientometrics* 42, 3 (1998), 423–428.

[97] Matthias Weidlich, Tomer Sagi, Henrik Leopold, Avigdor Gal, and Jan Mendling. 2013. Predicting the quality of process model matching. In *Business Process Management*. Springer, 203–210.

[98] Floris Wiesman, Nico Roos, and Paul Vogt. 2002. Automatic ontology mapping for agent communication. In *Proceedings of the first international joint conference on Autonomous agents and multiagent systems: part 2*. ACM, 563–564.

[99] Claes Wohlin, Aybuke Aurum, Lefteris Angelis, Laura Phillips, Yvonne Dittrich, Tony Gorschek, Hakan Grahn, Kennet Henningsson, Simon Kagstrom, Graham Low, et al. 2011. The success factors powering industry-academia collaboration. *IEEE software* 29, 2 (2011), 67–73.

[100] Michel Zitt and Elise Bassecoulard. 2006. Delineating complex scientific fields by an hybrid lexical-citation method: An application to nanosciences. *Information processing & management* 42, 6 (2006), 1513–1531.