Ontology Matching for Patent Classification

Christoph Quix[1,2], Sandra Geisler[2], Rihan Hai[1], Sanchit Alekh[1]

[1] Databases and Information Systems, RWTH Aachen University, Germany
[2] Fraunhofer-Institute for Applied Information Technology FIT, Germany
[1]lastname@dbis.rwth-aachen.de
[2]firstname.lastname@fit.fraunhofer.de

Abstract. Interdisciplinary research and development projects in medical engineering benefit from well selected collaboration partners. The process of finding such partners from often unfamiliar fields is difficult, but can be supported by an expert profile that is based on patent analysis and classifying the patents to competence fields in medical engineering. Patent analysis and categorization are difficult and require the analysis of the semantic content. Hence, we propose a twofold approach using a large controlled vocabulary, a smaller competence field ontology, and an alignment between them to assign patents to a certain competence field. The approach has two parts: a *Topic Map* approach and a *Publication approach*. We evaluate these approaches and its components in several ways. Furthermore, we compare four different ways to assign a patent to a competence field and show that the semantic wealth of a large biomedical ontology is beneficial to the classification task.

1 Introduction

Ontology matching has been an active research area for more than 10 years [17,18]. Ontologies are used to describe a domain of interest by concepts and relationships between them, and to provide a formal description of these relationships. Thus, although the aim of ontology matching seems to be the matching of classes and properties, usually its actual intention is *to match elements of the domains described by the ontologies*. An example for such a 'domain matching' task is patent classification in which patents should be assigned to a class in a classification [4].

While a classification scheme or taxonomy can be easily represented as an ontology, representing the content of a patent as an ontology or describing the patent with elements of an ontology is more challenging. Patents have their own specific language and use a terminology that is different from a typical research publication. Patents are classified using the International Patent Classification (IPC) system; however, this is too general for a detailed patent analysis [12]. On the other hand, patent data is available in a structured form (usually XML) from patent offices, which simplifies the pre-processing and extraction of basic information such as title, abstract, and authors. Furthermore, they are often also available in multiple languages; at least, the bibliographic information and abstract is available in English, which solves the problem of multi-lingual documents.

We are aiming at building a recommender system for research projects in medical engineering (ME) [7] in the context of the mi-Mappa project[3]. In ME researchers from several disciplines (e.g., biology, medicine, mechanical engineering, computer science) work jointly on a research project. Furthermore, ME is a highly innovative domain with short product cycles requiring a fast translation of research results into applicable products [2]. While on the one hand, a publication list of a researcher provides a good basis for creating an author profile [14], on the other hand a list of patents allows to characterize the ability of a researcher to develop inventions and market-ready products. Hence, we concentrate mainly on the analysis of patents.

To address the problem of patent terminology, we exploit explicit references to scientific publications and their semantic annotations. In ME, most of the publications appear in journals or conferences that are indexed by PubMed[4]. PubMed uses MeSH[5], a rich controlled vocabulary with a hierarchical structure, to annotate the publications. Thus, to retrieve a MeSH annotation for a patent, we lookup the references to research articles in PubMed and retrieve the corresponding MeSH terms.

Using references to scientific publications is only one aspect in our approach for patent classification. The overall approach, depicted in Figure 1 consists of two complementary sub-approaches: the *Topic Map Approach* (TMA) and the *Publications Approach* (PBA). Both approaches utilize two ontologies - a competence field (CF) ontology and an ontology with comprehensive medical knowledge (MeSH) - and an alignment between them.

For the Publication Approach, excerpts of publication databases, as well as their associated MeSH terms are imported into our Data Lake (DL) system Constance [8]. The data lake can then be queried on-the-fly for publications cited by the currently processed patent, as well as the MeSH terms that are pertinent to each of these publications. For the categorization of the input patent with the TMA, the topic with the highest probability in the topic map (or multiple topics if they have the same probability) is retrieved. Each term characterizing the topic is compared with all concepts in the MeSH ontology resulting in a set of matching concepts.

Thus, for both approaches, we have a list of related concepts from the MeSH ontology. To establish a link to the competence field ontology, which we have created to describe the innovation areas in medical engineering (see section 2), we use ontology matching.

There are several questions arising when we analyze the presented approach. Creating an alignment between ontologies and the use of a huge medical ontology in this context require a high amount of resources in terms of memory and CPU power. Hence, we need to know if the effort using it is worth it. Furthermore,

[3] `http://www.dbis.rwth-aachen.de/mi-Mappa`

[4] `https://www.ncbi.nlm.nih.gov/pubmed/`

[5] Medical Subject Headings, `https://www.nlm.nih.gov/mesh/`

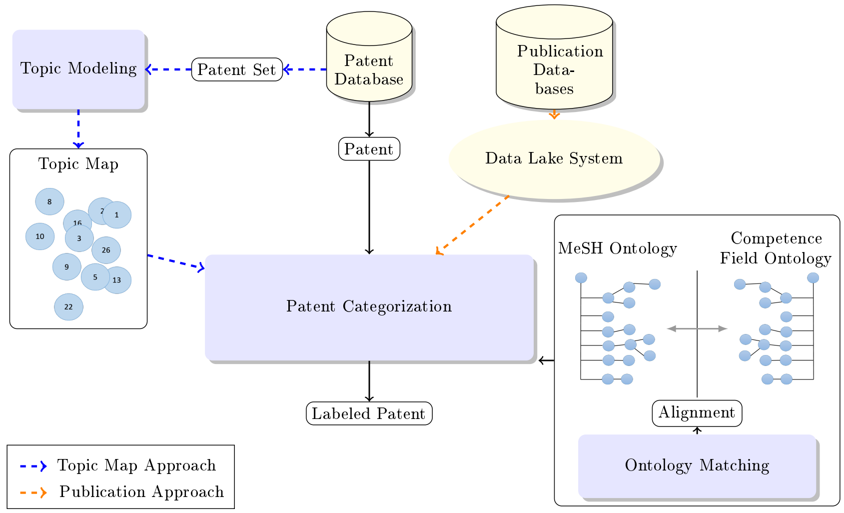

Fig. 1. The Overall Architecture

it is of interest if the quality and size of the alignment between the ontologies have an impact on the results. A special problem is to rate the quality of the alignment without a reference alignment. To answer these questions we present the following contributions in this paper:

- We analyze and select medical ontologies to use them as a basis for the creation of the CF ontology and as a single point of entry to identify the semantics of patents and publications.
- We describe the process of designing the competence field ontology and rate its quality based on approved methodologies.
- We create different alignments between the CF ontology and the medical ontology with different matcher configurations and compare their quality.
- We compare the results of four different approaches to categorize a patent: (1) Topic Map Approach with direct comparison of terms with concepts of the CF ontology (i.e., using no ontology matching techniques), (2) Publication Approach, (3) Topic Map Approach, (4) combination of Topic Map Approach and Publication Approach. Approach (2) and (3) use the alignment computed by ontology matching.

The rest of this paper is structured as follows. In Section 2 we explain the design of the CF ontology. Furthermore, the selection process of the utilized medical ontologies is explained (first results about these issues were reported in [7]). In Section 3 we describe the approaches to establish a link between patents and competence fields. In Section 4 the four approaches to categorize patents into competence fields will be evaluated. Finally, we discuss related work in Section 5 and conclude the paper in Section 6.

2 Modeling and Selection of Ontologies

Our assumption is that a huge medical ontology (or a set of them) and mappings to a smaller competence field ontology (CFO) will help to more easily classify patents into competence fields. The idea is somehow similar to a smart multi-level filter. First we retrieve terms describing the content of a patent (either from the topic map or the cited publications). These terms are compared to concept names in a huge medical taxonomy using string similarity measures. The most similar ones are selected, which results in a potentially long list of concepts. Afterwards we filter further and search for mappings from the concepts and their predecessors to concepts of the smaller competence field ontology using more intelligent matchers. This leads to scores which identify the membership confidence to the competence fields.

To implement this approach two foremost things have to be done: (1) we have to model the competence field ontology and (2) we need to evaluate and select comprehensive medical ontologies. For the design of ontologies there exist several acknowledged methodologies, such as METHONTOLOGY [6], TOVE, or the work by Noy and McGuinness [13]. The NeOn methodology [19] is a more recent approach which combines ideas of the former methods. The methodology describes nine scenarios for building ontologies and ontology networks [19].

To create the CFO, we started from the descriptions in [15,3] and also used an extended description of ME domain experts. As the six competence fields are the categories we want to assign to the patents, we use these (and only these) as first level concepts in the ontology. All further concepts will be subconcepts of these. This approach corresponds to the *reusing and reengineering non-ontological resources* scenario of the NeOn methodology. To find subconcepts, we had analyzed the detailed description of the CFs by the domain experts. Firstly, we extracted a preliminary selection of 174 terms which we used to make a first draft of a preliminary ontology on which domain experts commented using a custom web front end for the review of ontologies.

In parallel we searched for one or multiple large biomedical taxonomies. We need these taxonomies for two things. First, we want to extend the basic CFO we created before with more terms to describe the competence fields in more detail. Second, we need the large ontology as entry point to find terms describing the patents and with the alignment to the CFO we can determine the corresponding competence fields. This corresponds to the sixth scenario of the NeOn methodology, namely *reusing, merging and reengineering ontological resources*. The first step in this scenario is the *ontological resource reuse process*, starting with the *Ontology Search* [19]. Hence, we searched for ontologies with domain specific search engines as described in [7]. We used the Bioportal[6] search engine, the Ontology Lookup Service[7], and the Ontobee[8] search engine using the preliminary list of terms to have a broad overview. Afterwards we carried out

[6] http://bioportal.bioontology.org
[7] http://www.ebi.ac.uk/ontology-lookup
[8] http://www.ontobee.org

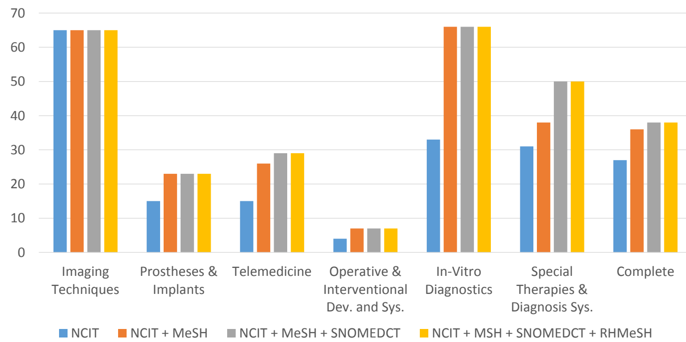

Fig. 2. Coverage based on Combination of Ontologies

the *Ontology Assessment and Comparison* steps [19]. The most promising four ontologies found are the National Cancer Institute (NCIT) Thesaurus, the Systematized Nomenclature of Medicine - Clinical Terms (SNOMEDCT), MeSH, and the Robert Hoehndorf Version of MeSH (RHMeSH). To identify if they satisfy our needs, we did a coverage analysis, where the coverage is the percentage of the competence field terms present in each of the ontologies. No single ontology covered all competence fields to a satisfying degree; some reached more than 60% for one competence field but only about 20% for the other fields (e.g., NCIT covers 'Imaging Techniques' well, but not the other fields).

Hence, we decided to analyze the coverage by adding one ontology after another to see the gain of adding further ontologies. We used the most promising ontologies identified before and started with the NCI Thesaurus. Figure 2 shows the results.

It can be noted, that we gain about 10% coverage using all ontologies. The biggest gain is achieved by adding the MeSH ontology. Thus, we decided to use the NCIT and the MeSH ontologies to extend the CFO, as this was a good compromise between coverage and complexity. For the matching of the biomedical ontology to the CFO we first picked only one ontology to keep the computational overhead during runtime low. If it does not give us satisfying results, we will add more ontologies and also align them with the CFO. One possibility would be also to use the UMLS which is a superset of many medical ontologies, but it is really large, which could lead to performance problems. For now, we selected the Robert Hoehndorf MeSH[9] as it has a good coverage and is available in the OWL format.

The next steps to develop the CFO are the *ontology aligning and ontology merging* step and the *ontological resource engineering process*. We proceeded in these steps as follows. We took the extracted terms, the so far found concepts from the coverage analysis, and the detailed description of the innovation fields, and carried out an extended search in the MeSH Browser[10] and the NCIT Brow-

[9] `https://bioportal.bioontology.org/ontologies/RH-MESH`
[10] `https://meshb.nlm.nih.gov/search`

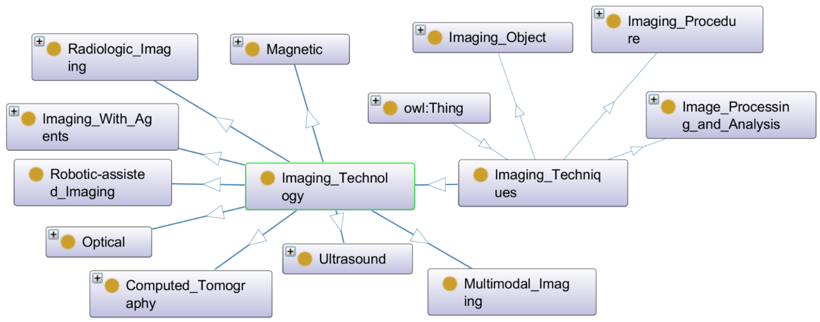

Fig. 3. The Imaging Technique Concept

ser[11] for these and related concepts. We analyzed the hierarchical structure of each of the found concepts and decided for each concept if it is adopted into the CFO. Where applicable we also adopted the inheritance relationship of concepts. We extended and restructured the CFO in cycles, i.e., according to [19] we did a *re-conceptualization* on different levels for the CFO and for the concepts from the biomedical ontologies. For the upper levels of the CFO we designed categories which fit better to our purposes for categorizing terms for medical engineering. We used a mind mapping technique and a bottom-up approach as for example described by Noy and McGuinness [13] to refine the design. As an example, the Imaging Techniques concepts and the concepts of the concept Imaging_Technology (2nd level) are visualized in Figure 3.

The ontology has been implemented in OWL using the NeOn toolkit[12]. We evaluated the CFO also in tests in the complete process of patent categorization. We noticed that the initial results were not satisfying because some competence fields were not represented well in the CFO. Hence, we did a frequency analysis of the MeSH terms from the Publications Approach. We made a ranked list of MeSH concepts based on how often they have been searched for, but did not lead to matches in the CFO. Based on this list we added more useful concepts to the CFO (no trivial, misleading terms, such as *Human*, but for example *Gene Expression Regulation*). The current CFO consists of 529 concepts and can be downloaded at `http://dbis.rwth-aachen.de/cms/projects/mi-mappa/CFO.owl`.

3 Matching of Ontologies and Topic Maps

As explained above, we are using three different basic approaches and one combined approach to classify patents. Figure 4 gives an overview of the different approaches.

#1: TMD (Topic Map with Direct Mapping): In this approach, we match the terms extracted from the topic maps directly with the competence

[11] `https://ncit.nci.nih.gov/ncitbrowser/`
[12] `http://neon-toolkit.org`

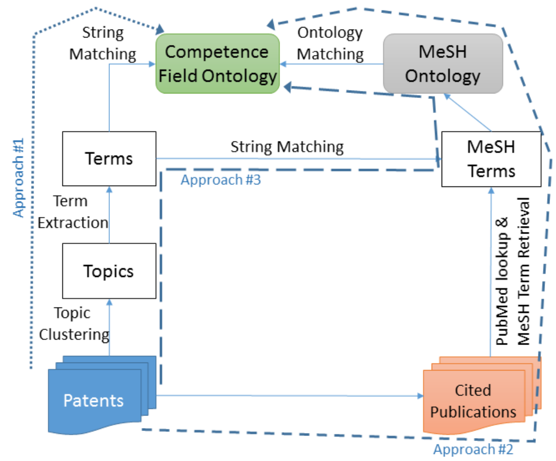

Fig. 4. The different approaches used for Evaluation

field ontology. This can be seen as a base line as it does not use a semantically rich ontology as intermediate component, but only uses string matching to match terms and ontology elements.

#2: PBA (Publication-Based Approach): This approach uses the MeSH terms attached to publications which are referenced by a patent. Then, we use an alignment between the CFO and MeSH to compute a score for the relationship between a patent and a competence field.

#3: TMA (Topic Map Approach): Here, we also use topic mapping (as in approach #1) to create initial clusters of patents and extract terms occurring frequently in these clusters. These terms are then matched with the concepts of the MeSH ontology. Using the same alignment as in the second approach, a relationship to the CFO is established.

#4: COM (Combined Approach of #2 & #3): This is a combination of PBA and TMA, with an emphasis on the results of PBA.

As the approaches TMD and TMA are based on topics, we first briefly explain this part, before we present how we did the alignment between of CFO and MeSH, and describe the publication-based approach.

3.1 Topic Mapping

A basic set of patents is used to build a topic map. Firstly, the corpus of documents is preprocessed (stemming, removing stop words, etc.) and a Document-Term-Matrix (DTM) is created. The matrix is input to a Latent Dirichlet Allocation (LDA) algorithm with the Gibbs sampling algorithm for estimation and variational expectation maximization [11]. The LDA determines a fixed number of topics which are each described by a fixed number of stemmed terms. To each patent in the basic patent set topics are assigned with a probability. The topic map and the assignments are stored in a database.

We evaluated different numbers of topics and different numbers of terms extracted for each topic (e.g., 10, 30, 50, etc.). As computation of the subsequent steps increases with a higher number of topics and terms, we used 50 topics and 50 terms for our evaluation in Section 4. As the TMD approach matches the terms directly with the CFO, no further processing on the extracted terms is done in this case. We just do a similarity calculation using a normalized *Longest Common Subsequence* [10] algorithm. In our tests, we found that a threshold value of 0.5 for the string similarity provides the best compromise.

For the categorization of the input patent with the TMA, the topic with the highest probability in the topic map (or multiple topics if they have the same probability) is retrieved. Each term characterizing the topic is compared with all concepts in the medical ontology resulting in a set of matching concepts. For each of the concepts in the set direct mappings and mappings of parent concepts are collected from the alignment and it is determined to which competence field the matching concept in the CF ontology belongs. From the similarities average scores are calculated for each term and each competence field. Based on this, an average score is calculated from all terms for the topic(s) of the patent. Hence, for each patent we have a score for each of the competence fields and normalize these, such that all scores add up to 1.

3.2 Ontology Matching

To rate how strong a patent or publication is related to a certain competence field, we need to match the describing terms either extracted from publications or from the topic map to terms describing the competence fields. In preparation to this step, we create an alignment between the selected MeSH ontology and the CFO. The alignment constitutes of a set of mappings between the concepts of the two ontologies. This means, for each mapping we have a pair of concepts and a similarity value. As we do not try to re-invent the wheel, we used AgreementMakerLight [5] as it produced constantly good results in the recent OAEI campaigns and also performs well for large biomedical ontologies. AgreementMakerLight is able to combine different matchers to create an alignment. We used the string matcher, the word matcher, the structural matcher, the lexical matcher, the cardinality filter, and the coherence filter. As a similarity threshold we used a value of 0.6. The matchers have been combined in a hierarchical way and the default settings for each matcher have been used.

Currently, we are also testing other settings and their impact on the quality of patent classification results. First experiments show, that slightly relaxed filter settings (e.g., not using a cardinality filter) increases the number of mappings and therefore, also improves the classification result.

3.3 Publication-based Approach

We queried the web service of EPMC[13] to retrieve the metadata of the papers referenced in our patent dataset. To extract the references from the patent

[13] European PubMed Central, `https://europepmc.org/`

data, we use a pattern-based approach similar to the FreeCite citation parser[14]. Luckily, the patent data is semi-structured such that the citations can be clearly identified. Nevertheless, for a large fraction of the patents, we are not able to retrieve MeSH terms from referenced publications (either because the referenced publication does not appear in PubMed or the citation is incorrect).

The retrieved metadata for each referenced publication is then stored in our Data Lake system Constance [8] from which it is accessed during patent processing.

Subsequently, we use a process which is similar to the TMA. In both cases, we have a list of MeSH terms as input. For each of the terms in the list, the mappings are determined as before and average scores per competence field are calculated and normalized for each patent.

3.4 Combined Approach

In the combined approach (COM), if both approaches TMA and PBA deliver results, the results are combined and overall scores for each competence field are determined. In all cases, we assign at most three competence fields to a patent. In most cases, only one competence field is assigned to a patent as the other competence fields do not exceed a certain threshold. Thus, we take the intersection of competence fields computed by TMA and PBA. If this is not empty, we take this result (because both approaches are sure about a result). If the intersection is empty, we take the competence fields with the highest scores from TMA and PBA.

4 Evaluation

In our experimental setup, we compare the aforementioned approaches. For the analysis of patents, we need a comprehensive data basis with high data quality. In the course of the mi-Mappa project, a subset of the PATSTAT database (2016 Spring edition, version 5.07) published by the European Patent Office (EPO) is used. For our purposes, we selected patents issued by a German (DE) or British (UK) authority after 2004, which are from the medical domain (CPC class A61), and which have an English abstract and title. This results in a set of 26,814 patents. For about 4,500 patents of this set, we are able to retrieve MeSH terms for the referenced publications. From this set, we randomly selected 59 patents to do a manual assignment to competence fields to evaluate our approaches. A more extensive expert evaluation is currently being setup. In addition, we plan also to evaluate our approach to the results of our project partners who apply a supervised learning approach using Support Vector Machines [9].

For TMA, we experimented with various configurations for the number of topics and their associated terms. We observe that with a relatively small number of topics and terms, e.g. 10 or 20, the terms are extremely broad-based and do

[14] http://freecite.library.brown.edu/

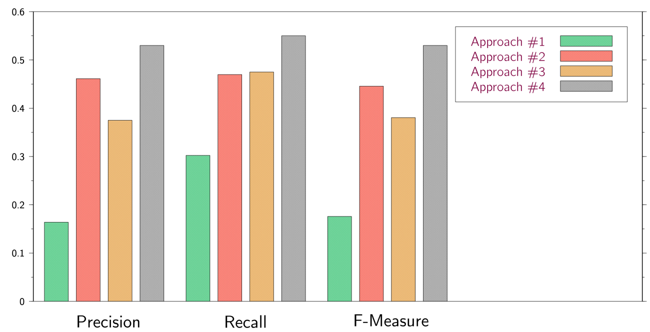

Fig. 5. Comparison of the precision, recall and f-measure for the different approaches

not provide meaningful matches with the MeSH ontology or the CFO. Therefore, based on the results, we chose the number of topics, as well as the number of terms to be 50 for our default test configuration.

Fig. 5 summarizes the findings from our experiments for the aforementioned approaches. It is obvious that all three of our proposed approaches #2, #3, and #4 perform significantly better than the baseline approach #1. All three evaluation parameters, i.e., precision, recall, and the f-measure are worse for the baseline approach. In contrast, when the MeSH ontology is used for matching the ontology terms (*#3*), the precision and f-score are 0.375 and 0.38, respectively, which are more than the doubled values of corresponding values produced by *#1*. The PBA performs even better, resulting in precision, recall, and f-score values of 0.46, 0.47 and 0.44, respectively. However, the combined approach #4 significantly outperforms all the others, and results in precision, recall, and f-measure values of 0.53, 0.55 and 0.53, respectively. Indeed, we found that in the case of the TMD-approach #1, there were a lot of erroneous matches, which led to non-distinctive results for the CFO assignment. These results affirm the superiority of techniques which use a comprehensive biomedical ontology and ontology matching for patent classification tasks.

5 Related Work

There are only few works that apply ontology matching in the context of patent analysis. Semantic similarities (based on ontology matching) and case-based reasoning have been applied in the design of invention processes which use patent analysis to study related works. Patent analysis using ontologies has been applied especially for patent search [1]. A patent search request can be represented as an ontology or as a set of concepts of an existing ontology, which is then matched with the ontologies representing the knowledge of patents [16]. Another example is the PatExpert system which uses a network of ontologies and knowledge bases to enable patent search, classification, and clustering [21]. Trappey et al. propose a system that calculates the conditional probability that, given a specific text chunk is present in the document, the chunk is mapped to a specific concept of a given ontology [20]. Patent similarity is then based on the number of common matched concepts. This approach restricts the clustering to the terms of the ontology which might lead to missing important terms not present in the ontology.

6 Conclusion

Patent analysis is a complex topic as patents use their own language and terminology. Even for humans used to research publications, patents are difficult to understand. Thus, typical approaches for classifying patents might fail.

In this paper, we investigated an ontology-based approach to assign patents to competence fields in medical engineering. We developed two different approaches and a combined approach that are based on a large biomedical ontology, its alignment to the competence field ontology designed by us, and other ontology matching techniques. We have shown that these more elaborated approaches outperform an approach that directly matches terms of patents with the competence field ontology.

However, the overall f-measure of about 55% for the combined approach is not yet satisfying. One problem is the small set of patents for which we have assigned competence fields that we can use as a ground truth. This will be extended with a larger expert evaluation in which patents will be classified by several experts. Even humans might disagree on the assignment of a patent to a competence field; therefore, we will have multiple expert opinions for one patent. We will also work on fine tuning and optimizing our approach. So far, we focused on the quality of the result, and did not worry too much about the performance. Still, we think that the area of patent classification is an interesting field which could benefit more from the results in ontology matching.

Acknowledgements

This work has been supported by the Klaus Tschira Stiftung gGmbH in the context of the mi-Mappa project (`http://www.dbis.rwth-aachen.de/mi-Mappa/`, project no. 00.263.2015). We thank our project partners from the Institute of Applied Medical Engineering at the Helmholtz Institute of RWTH Aachen University & Hospital, especially Dr. Robert Farkas, for the fruitful discussions of the approach and for providing the patent data.

References

1. D. Bonino, A. Ciaramella, F. Corno. Review of the state-of-the-art in patent information and forthcoming evolutions in intelligent patent informatics. *World Patent Information*, **32**(1):30–38, 2010.
2. BVMed. Branchenbericht Medizintechnologien 2015. www.bvmed.de/branchenbericht, June 2015.
3. Deutsche Gesellschaft für Biomed. Technik im VDE. Empfehlungen zur Verbesserung der Innovationsrahmenbedingungen für Hochtechnologie-Medizin. *Tech. rep.*, VDE, 2012.

4. C. J. Fall, A. Törcsvári, K. Benzineb, G. Karetka. Automated categorization in the international patent classification. In *ACM SIGIR Forum*, pp. 10–25. 2003.
5. D. Faria, C. Pesquita, B. S. Balasubramani, C. Martins, J. Cardoso, H. Curado, F. M. Couto, I. F. Cruz. OAEI 2016 results of AML. In *Proc. 11th Intl. Workshop on Ontology Matching*, pp. 138–145. 2016.
6. M. Fernández-López, A. Gómez-Pérez, N. Juristo. Methontology: from ontological art towards ontological engineering. In *Proc. Symposium on Ontological Engineering of AAAI*. 1997.
7. S. Geisler, R. Hai, C. Quix. An Ontology-based Collaboration Recommender System using Patents. In *Proc. Intl. Conf. on Knowledge Engineering and Ontology Development (KEOD)*, pp. 389–394. Lisbon, Portugal, 2015.
8. R. Hai, S. Geisler, C. Quix. Constance: An Intelligent Data Lake System. In F. Özcan, G. Koutrika, S. Madden (eds.), *Proc. Intl. Conf. on Management of Data (SIGMOD)*, pp. 2097–2100. ACM, San Francisco, CA, USA, 2016.
9. N. Hamadeh, M. Bukowski, T. Schmitz-Rode, R. Farkas. Cooperative Patent Classification as a mean of validation for Support Vector Machine Learning: Case Study in Biomedical Emerging Fields of Technology. In *51. Jahrestagung der Biomedizinischen Technik (BMT)*. 2017.
10. D. S. Hirschberg. Algorithms for the longest common subsequence problem. *Journal of the ACM (JACM)*, **24**(4):664–675, 1977.
11. K. Hornik, B. Grün. topicmodels: An R package for fitting topic models. *Journal of Statistical Software*, **40**(13):1–30, 2011.
12. K.-K. Lai, S.-J. Wu. Using the patent co-citation approach to establish a new patent classification system. *Information Processing & Mgmt.*, **41**(2):313–330, 2005.
13. N. F. Noy, D. L. McGuinness. Ontology Development 101: A Guide to Creating Your First Ontology. *Tutorial*, Stanford University, 2001.
14. J. Portenoy, J. D. West. Visualizing Scholarly Publications and Citations to Enhance Author Profiles. In *Proc. WWW*, pp. 1279–1282. Perth, Australia, 2017.
15. C. Schlötelburg, C. Weiß, P. Hahn, T. Becks, A. C. Mühlbacher. Identifizierung von Innovationshürden in der Medizintechnik. *Tech. rep.*, Bundesministeriums für Bildung und Forschung, October 2008.
16. A. Segev, J. Kantola. Patent Search Decision Support Service. In *7th Intl. Conf. on Information Technology: New Generations (ITNG)*, pp. 568–573. 2010.
17. P. Shvaiko, J. Euzenat. A Survey of Schema-Based Matching Approaches. *Journal on Data Semantics*, **IV**:146–171, 2005. LNCS 3730.
18. P. Shvaiko, J. Euzenat. Ontology Matching: State of the Art and Future Challenges. *IEEE Transactions on Knowledge and Data Engineering*, **25**(1):158–176, 2013.
19. M. C. Suárez-Figueroa. *NeOn Methodology for building ontology networks: specification, scheduling and reuse*. Ph.D. thesis, Univ. Politecnica de Madrid, 2010.
20. A. J. Trappey, C. V. Trappey, F.-C. Hsu, D. W. Hsiao. A fuzzy ontological knowledge document clustering methodology. *IEEE Trans. on Systems, Man, and Cybernetics, Part B*, **39**(3):806–814, 2009.
21. L. Wanner, R. Baeza-Yates, S. Brügmann, J. Codina, B. Diallo, E. Escorsa, M. Giereth, Y. Kompatsiaris, S. Papadopoulos, E. Pianta, et al. Towards content-oriented patent document processing. *World Patent Information*, **30**(1):21–33, 2008.