Dynamic Ontology Matching Challenge

Maaike H.T. de Boer[1], Linda Oosterheert[1] and Roos M. Bakker[1,2]

[1]TNO - Netherlands Organisation for Applied Scientific Research, Anna van Buerenplein 1, 2595DA, The Hague, The Netherlands

[2]Universiteit Leiden, Reuvensplaats 3, Leiden, 2311BE, The Netherlands

Abstract

The labour market is currently heavily struggling with friction in demand and supply. Skills-based approaches deliver promising results. These approaches ask for a common and up-to-date skills language to achieve their full potential. Skills ontologies such as ESCO and O*NET exist, but tend to get outdated soon, as the labour market changes quickly and intensive manual expert-based labour is needed for updating the ontologies. The existence of multiple skills ontologies allows for applicability in different contexts (such as different countries), but requires mappings between them to be able to relate, reuse and transfer knowledge.

In this challenge, we invite participants to implement novel ideas on how to keep mappings between ontologies up-to-date in a dynamic context. We will provide different versions of the ESCO ontology and the mapping to O*NET as the ontology and data. It is necessary to use a hybrid method in this challenge, and it is allowed to use some human annotation in the challenge.

1. Introduction

The tension on the labour market has been significantly increased the last years. For example, the Netherlands faced a record high number of vacancies at the beginning of 2022[1]. In a large number of sectors there is an acute scarcity of staff (e.g., construction and installation, health care), whereas in other sectors staff must be out-flowed (e.g., financial and administrative) [1]. Employers cannot find the employees they need, and employees experience they cannot use large parts of their knowledge and skills in their current job or are not enabled to work more hours [2].

To understand and meet emerging skills demands and to empower individuals to learn, unlearn and relearn skills, a transition from a diploma-based to a skills-based learning and working system is ongoing. These skills-based approaches ask for a common and up-to-date skills language to achieve their full potential. Skills ontologies such as ESCO and O*NET exist, but skills ontologies tend to get outdated soon, as the labour market changes quickly and intensive manual expert-based labour is needed for updating the ontologies.

In A. Martin, K. Hinkelmann, H.-G. Fill, A. Gerber, D. Lenat, R. Stolle, F. van Harmelen (Eds.), Proceedings of the AAAI 2023 Spring Symposium on Challenges Requiring the Combination of Machine Learning and Knowledge Engineering (AAAI-MAKE 2023), Hyatt Regency, San Francisco Airport, California, USA, March 27-29, 2023.

maaike.deboer@tno.nl (M. H.T. d. Boer); linda.oosterheert@tno.nl (L. Oosterheert); roos.bakker@tno.nl (R. M. Bakker)

0000-0002-2775-8351 (M. H.T. d. Boer); ? (L. Oosterheert); 0000-0002-1760-2740 (R. M. Bakker)

[1]https://www.cbs.nl/nl-nl/nieuws/2022/20/arbeidsmarkt-nog-krapper-in-eerste-kwartaal

In this challenge, we invite participants to implement novel ideas on how to keep mappings between ontologies up-to-date in a dynamic context. Ontologies and knowledge models in general represent a part of the world. Ideally, they are connected to other ontologies to form a broader network of knowledge. However, the world is ever changing and so are the ontologies that describe it. This poses a problem in the mapping of ontologies. One change in an ontology can threaten the correctness of all its mappings. We believe a Hybrid solution is necessary, because current ontology mapping techniques do not perform sufficiently to allow a fully automatic solution.

The next section provides a short overview of related work on ontology mapping techniques. Section 3 gives insight in a few of the skills ontologies and section 4 describes the challenge.

2. Related Work

Finding correspondences between ontologies is often called ontology mapping or ontology matching. The result of this process is named an ontology alignment. Related work in this field can be divided into structure-level and element-level mapping [3, 4]. Harrow et al. [5] add knowledge as an additional level, which refers to data or facts stored in databases. Structure-level mapping is a mapping that includes the entire ontologies or groups of concepts with groups of concepts. In structure-level mapping, Euzenat et al. [4] make a difference between graph-based, taxonomy-based, model-based and instance-based techniques. All techniques view the ontology as a different type of domain model and use the known methods from their field on the ontology. The graph-based technique sees the ontology as a labelled graph and use graph algorithms. The taxonomy-based technique views the ontology as a taxonomy and takes only the specialisation relation into account. The model-based technique handles the input as a semantic interpretation and uses mainly logic reasoning techniques. The instance-based techniques compares sets of instances to match classes using set-theoretic reasoning.

Element-level mapping is a mapping in which each element in an ontology is considered independently from the other elements in the ontology. In element-level mapping, Euzenat et al. [4] make a distinction between formal resource-based, informal resource-based, constraint-based, string-based and language-based approaches. Formal resource-based means that formal external ontologies, such as upper ontologies are used as additional knowledge. Informal resource-based means that other external resources are used. Constraint-based includes ontological constraints such as domain and range or type of attributes to calculate a similarity. String-based approaches only use the string itself (names and/or descriptions of entities) to calculate a similarity. Often string distance metrics such as Levenshtein, Jaccard or TF-IDF are used. Language-based approaches rely on Natural Language Processing and use techniques such as tokenisation and lemmatisation, and external resources such as WordNet [6]. With the introduction of BERT [7], contextual embeddings provide a good performance on many NLP tasks. Neutel and de Boer [8] compare several alignment methods (fasttext labels, fasttext descriptions, BERT 'CLS' descriptions, BERT mean token descriptions and S-BERT description) in the mapping of ESCO and O*NET. The results showed that Sentence BERT has highest performance (in terms of coverage vs. Mean Reciprocal Rank) but it does not provide a ready-to-use alignment yet. de Boer et al. [9] compare performance of several contextual embeddings in the creation and mapping of evolving ontologies. Also recent top performer BERTmap [10, 11] and the Truveta mapper [12] use contextual embeddings.

2.1. Challenges in Ontology Matching

Otero-Cerdeira et al. [13] and Shvaiko and Euzenat [14] provide a survey of literature over the 2010s. Many ontology matching systems are created, such as AgreementMaker(light) [15], Anchor-Flood [16], ASMOV [17], LogMap [18], SAMBO [19] and SEMA [20]. Shvaiko and Euzenat [14] mention that there are still challenges ahead: large-scale evaluation, efficiency of ontology matching, matching with background knowledge, matcher selection and self-configuration, user involvement, explanations of ontology matching, collaborative and social ontology matching and alignment infrastructure. Otero-Cerdeira et al. [13] add automatic discovery of complex relations, many-to-one mappings, correct alignment of large ontologies and focus on applying automatically created mappings to practical applications.

Last year, Portisch et al. [21] provided a survey of the most recent development, mainly focused on the usage of background knowledge (one of the challenges mentioned above). They make a distinction between unstructured and structured background knowledge, and on the other axis domain-specific and general purpose background knowledge. Within structured background knowledge, there is a division between lexical and taxonomical (mono- or multilingual), factual database, semantic web database (single or linked) and pre-trained Neural Models (mono- or multilingual). The survey shows that general-purpose knowledge sources are more often used compared to domain-specific knowledge sources, and that WordNet (structured; lexical and taxonomical; monolingual) is most often used. Also, there is a bias towards biomedical matching tasks and monolingual (mainly English) matching. As white spots, logic-based and neural-based strategies (such as BERT) seem promising, as is exploration of unstructured and structured multilingual datasets. Furthermore, they mention that most system use direct label linking, but given links, fuzzy linking or Word Sense Disambiguation can be used as well. Trojahn et al. [22] adds that current tools fail to correctly capture the semantics behind concepts (which is also mentioned by Zeng et al. [23] about entity alignment), other relations than equivalences are largely neglected and that evaluations involving foundational ontologies are not yet addressed.

2.2. Evaluation

Evaluation of an alignment is often done in the Ontology Alignment Evaluation Initiative (OAEI) [2]. This initiative started in 2004 and has several tracks, such as anatomy, conference, multifarm, complex, food nutritional composition, bioML, but also interactive matching, crosswalks data schema matching and common knowledge graphs. Each track has different criterion on the matching, and different evaluation metrics. For evaluation different evaluation platforms are created, currently SEALS, MELT [24] and HOBBIT [25]. Also the alignment API [26] is used in the OAEI. For entity alignment, the OpenEA library exists [27].

[2]http://oaei.ontologymatching.org/

3. Labour Market Classifications

There are several classifications available for the labour market. First, occupation standards are available, such as the Standard Occupational Classification (SOC)[3] and the International Standard Classification of Occupations (ISCO)[4]. Both are taxonomies that only define occupations.

There are also classifications that describe occupations and skills, of which ESCO and O*NET are the most known. ESCO - the European Skills, Competences, Qualifications and Occupations - is an ontology that includes occupations, skills and competences [28]. A full first version is released in 2017 and is created with expert knowledge. It includes more than 13,000 skills and almost 3,000 occupations. Skills are classified in a strict hierarchy. Each occupation is associated with essential and optional skills. ESCO occupations are linked to ISCO, in which ESCO is a specification of ISCO.

O*NET - the (North) American standard Occupational Information Network - is a thesaurus in the form of a database that contains occupations, workforce characteristics, occupational requirements, worker characteristics, worker requirements and experience requirements [29, 30]. A first version was already published in the last century, and a new version is available every few years, based on input from experts, job holders and job postings. O*NET contains around 1,000 occupations. Each occupation is associated with skills, abilities and work styles, with a degree of importance. O*NET occupations are linked to SOC, in which O*NET is a specification of SOC.

Mappings between those standards are previously created, and often done manually. Adjusting the classifications to changes in the demand for work is, therefore, labour-intensive. Further, changes in the demand for skills are not continuously processed in the skills classifications, despite the fast changes in the world of work. Recently, a report has been published in which AI in the form of a BERT model is used to provide a mapping from O*NET occupations (input) to ESCO occupations (output) based on (contextual) semantic textual similarity [31].

4. Challenge Description

In this challenge, we invite participants to implement novel ideas on how to keep mappings between ontologies up-to-date in the dynamic context of the labour market.

4.1. Dataset

We will use the ESCO ontology[5] as our ontology and data. In 2017, the first full version was published. In the beginning of 2022, a new version was published with 68 new occupations, 354 new skills and 158 new knowledge concepts, but also 2 obsolete occupations and 106 obsolete skills and knowledge concepts. Additions included skills for the green transition and occupations for emerging technologies. ESCO is mapped to other ontologies describing the labour market, of which we will use O*NET - Occupational Information Network[6].

[3] https://www.bls.gov/oes/current/oes_stru.htm
[4] https://www.ilo.org/public/english/bureau/stat/isco/
[5] https://ec.europa.eu/esco
[6] https://www.onetcenter.org

In this challenge, we provide the link to all ontologies as well as currently existing mappings. These are available at https://gitlab.com/tno-os/dynamic-ontology-matching-challenge.

4.2. Method

The goal is to create the (occupation) alignment of the ESCO 2017 version to O*NET (2019) (1) and the new ESCO (2022) version (2). It is necessary to use a hybrid method in this challenge, and it is allowed to use some human annotation in the challenge. Directions of solutions can be optimizing the combination of data-driven ontology mapping techniques and the (currently manually captured) expert-based knowledge (models), but also solutions that address the current challenges in ontology mapping.

We are looking for novel ideas, so it is not necessary to adhere to the evaluation platforms mentioned in the related work.

4.3. Evaluation

The winner of the challenge will be selected using the following criteria:

1. novelty of the method used
2. performance: the number of correct matches
3. scalability: can it work real-time or near real-time

Participants are invited to broaden the challenge to for example mapping to other ontologies, mapping of skills as well as occupations, use different type of relations (such as broader than / narrower than) and/or map to other languages.

More information on the submission of your solution and timelines will be made available through our Gitlab page (see section 4.1).

Acknowledgments

We would like to thank the Dutch 'Groeifonds' project Vaardig met Vaardigheden for their financial support, as well as the partners of the project. Furthermore, we would like to thank Quirine Smit for helping with the data preparation.

References

[1] F. Green, G. Henseke, Europe's evolving graduate labour markets: supply, demand, underemployment and pay, Journal for Labour Market Research 55 (2021) 1–13.
[2] G. Brunello, P. Wruuck, Skill shortages and skill mismatch in Europe: A review of the literature (2019).
[3] E. Rahm, P. A. Bernstein, A survey of approaches to automatic schema matching, the VLDB Journal 10 (2001) 334–350.
[4] J. Euzenat, P. Shvaiko, et al., Ontology matching, volume 18, Springer, 2007.

[5] I. Harrow, R. Balakrishnan, E. Jimenez-Ruiz, S. Jupp, J. Lomax, J. Reed, M. Romacker, C. Senger, A. Splendiani, J. Wilson, et al., Ontology mapping for semantically enabled applications, Drug discovery today 24 (2019) 2068–2075.
[6] G. A. Miller, Wordnet: a lexical database for English, Communications of the ACM 38 (1995) 39–41.
[7] J. Devlin, M.-W. Chang, K. Lee, K. Toutanova, Bert: Pre-training of deep bidirectional transformers for language understanding, arXiv preprint arXiv:1810.04805 (2018).
[8] S. Neutel, M. H. T. de Boer, Towards Automatic Ontology Alignment using BERT, in: AAAI Spring Symposium: Combining Machine Learning with Knowledge Engineering, 2021.
[9] M. H. T. de Boer, R. M. Bakker, M. Burghoorn, Creating dynamically evolving ontologies: A use case from the labour market domain, in: AAAI Spring Symposium: Combining Machine Learning with Knowledge Engineering, 2023.
[10] Y. He, J. Chen, D. Antonyrajah, I. Horrocks, BERTMap: A BERT-based Ontology Alignment System, arXiv preprint arXiv:2112.02682 (2021).
[11] Y. He, J. Chen, H. Dong, E. Jiménez-Ruiz, A. Hadian, I. Horrocks, Machine learning-friendly biomedical datasets for equivalence and subsumption ontology matching, in: The Semantic Web–ISWC 2022: 21st International Semantic Web Conference, Virtual Event, October 23–27, 2022, Proceedings, Springer, 2022, pp. 575–591.
[12] M. Amir, M. Baruah, M. Eslamialishah, S. Ehsani, A. Bahramali, S. Naddaf-Sh, S. Zarandioon, Truveta Mapper: A Zero-shot Ontology Alignment Framework, arXiv preprint arXiv:2301.09767 (2023).
[13] L. Otero-Cerdeira, F. J. Rodríguez-Martínez, A. Gómez-Rodríguez, Ontology matching: A literature review, Expert Systems with Applications 42 (2015) 949–971.
[14] P. Shvaiko, J. Euzenat, Ontology matching: state of the art and future challenges, IEEE Transactions on knowledge and data engineering 25 (2011) 158–176.
[15] D. Faria, C. Pesquita, E. Santos, M. Palmonari, I. F. Cruz, F. M. Couto, The agreementmakerlight ontology matching system, in: OTM Confederated International Conferences" On the Move to Meaningful Internet Systems", Springer, 2013, pp. 527–541.
[16] M. H. Seddiqui, M. Aono, An efficient and scalable algorithm for segmented alignment of ontologies of arbitrary size, Journal of web semantics 7 (2009) 344–356.
[17] Y. R. Jean-Mary, E. P. Shironoshita, M. R. Kabuka, Ontology matching with semantic verification, Journal of Web Semantics 7 (2009) 235–251.
[18] E. Jiménez-Ruiz, B. Cuenca Grau, Logmap: Logic-based and scalable ontology matching, in: International Semantic Web Conference, Springer, 2011, pp. 273–288.
[19] P. Lambrix, H. Tan, SAMBO—a system for aligning and merging biomedical ontologies, Journal of Web Semantics 4 (2006) 196–206.
[20] V. Spiliopoulos, A. G. Valarakos, G. A. Vouros, V. Karkaletsis, SEMA: Results for the ontology alignment contest OAEI 2007, in: Proceedings of the Second International Workshop on Ontology Matching, Citeseer, 2007, pp. 244–254.
[21] J. Portisch, M. Hladik, H. Paulheim, Background knowledge in ontology matching: A survey, Semantic Web (2022) 1–55.
[22] C. Trojahn, R. Vieira, D. Schmidt, A. Pease, G. Guizzardi, Foundational ontologies meet ontology matching: A survey, Semantic Web 13 (2022) 685–704.

[23] K. Zeng, C. Li, L. Hou, J. Li, L. Feng, A comprehensive survey of entity alignment for knowledge graphs, AI Open 2 (2021) 1–13.

[24] S. Hertling, J. Portisch, H. Paulheim, MELT - matching evaluation toolkit, in: Semantic Systems. The Power of AI and Knowledge Graphs - 15th International Conference, SEMANTiCS 2019, Karlsruhe, Germany, September 9-12, 2019, Proceedings, 2019, pp. 231–245.

[25] E. Jiménez-Ruiz, T. Saveta, O. Zamazal, S. Hertling, M. Roder, I. Fundulaki, A. N. Ngomo, M. Sherif, A. Annane, Z. Bellahsene, et al., Introducing the HOBBIT platform into the ontology alignment evaluation campaign, in: 13th International Workshop on Ontology Matching (OM), volume 2288, 2018, pp. 49–60.

[26] J. David, J. Euzenat, F. Scharffe, C. Trojahn dos Santos, The alignment API 4.0, Semantic web 2 (2011) 3–10.

[27] Z. Sun, Q. Zhang, W. Hu, C. Wang, M. Chen, F. Akrami, C. Li, A benchmarking study of embedding-based entity alignment for knowledge graphs, Proceedings of the VLDB Endowment 13 (2020) 2326–2340.

[28] J. De Smedt, M. le Vrang, A. Papantoniou, ESCO: Towards a Semantic Web for the European Labor Market, in: Ldow@ www, 2015.

[29] N. G. Peterson, M. D. Mumford, W. C. Borman, P. Jeanneret, E. A. Fleishman, An occupational information system for the 21st century: The development of O*NET, American Psychological Association, 1999.

[30] M. Cifuentes, J. Boyer, D. A. Lombardi, L. Punnett, Use of O*NET as a job exposure matrix: a literature review, American journal of industrial medicine 53 (2010) 898–914.

[31] The crosswalk between ESCO and O*NET, Technical Report, European Commission, 2022.