

# LSMatch Results for OAEI 2022

Abhisek Sharma<sup>1,\*†</sup>, Archana Patel<sup>2,†</sup> and Sarika Jain<sup>1,†</sup>

<sup>1</sup>National Institute of Technology Kurukshetra, India

<sup>2</sup>Eastern International University, Vietnam

## Abstract

The Large Scale Ontology Matching System (LSMatch) and its findings using OAEI 2022 datasets are presented in this paper. A string similarity and synonyms matcher is used in the element-level and label-based ontology matching system called LSMatch. The system's present efforts are concentrated on identifying common classes between the two ontologies. The OAEI effort on seven tracks—Anatomy, Conference, Multifarm, Bio-ML, Disease and Phenotype, Common Knowledge Graphs, and Knowledge Graph—is where LSMatch and LSMatch-Multilingual are collectively participating. LSMatch has shown encouraging outcomes across all seven tracks.

## Keywords

Ontology Matching, Knowledge Schema, Alignment, String similarity, Synonym matcher.

## 1. Presentation of the system

### 1.1. State, purpose, general statement

LSMatch (Large Scale Ontology Matching System) is an ontology matching system that finds correspondences between ontologies using lexical properties. It employs the Levenshtein string similarity measure and the synonyms matcher, which employs background knowledge containing synonyms to filter out concepts with similar meanings but different lexical representations [1]. This is LSMatch's second OAEI appearance, and it was tested on six tracks: Anatomy, Conference, Multifarm, Bio-ML, Common Knowledge Graphs, and Knowledge Graph. The LSMatch system was wrapped in the MELT framework [2], and it is performing at par with other systems in tracks.

### 1.2. Specific techniques used

The current version of LSMatch addresses monolingual ontology alignments, i.e., the concepts of the ontologies are in the same language, English [3]. We have called ontology as knowledge schema (KS) because the LSMatch system matches the classes only. The working of the LSMatch system is shown in figure 1. We introduce the multiple parts of the system by taking two Knowledge schemas (KS1 and KS2) as input to show the final set of alignments. LSMatch system takes input in any format and loaders loads the input KS (KS1 and KS2) as RDF graphs.

---

\*Corresponding author.

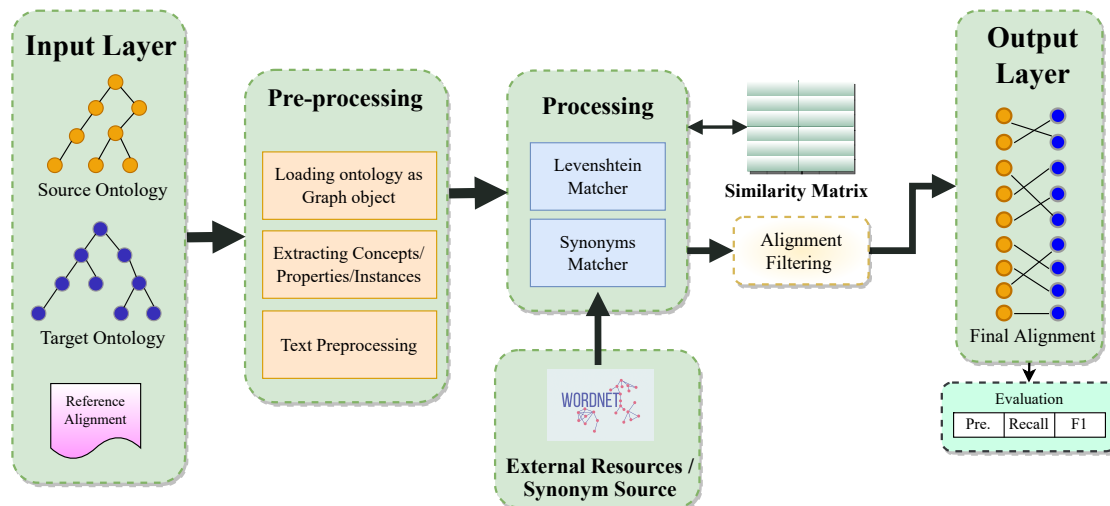
†These authors contributed equally.

✉ abhisek\_61900048@nitkkr.ac.in (A. Sharma); i.tiddi@vu.nl (A. Patel); jasarika@nitkkr.ac.in (S. Jain)



© 2022 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

CEUR Workshop Proceedings (CEUR-WS.org)



**Figure 1:** Overview of LSMatch system

- Levenshtein matcher: The LSMatch uses a string similarity matcher that calculates Levenshtein distance between the concepts [4]. The concepts are represented as `rdfs:label` or directly as the class name in the ontologies. The official definition of Levenshtein distance is stated as “The smallest number of insertions, deletions, and substitutions required to change one string or tree into another”<sup>1</sup>.
- Background knowledge [5]: To identify different lexical representations, LSMatch uses a synonym matcher that fetches synonyms Wordnet [6]. Python’s nltk library is used for wordnet inclusion.

For the purpose of storage and retrieval of alignments LSMatch uses dictionary. In the dictionary, we store information as `<key, value>` pairs where key is hashed [7, 8]. LSMatch stores the alignments received from both the matchers along with the similarity score. We target storing and updating the scores of pairs multiple times during the alignment process and having hashed keys allow us to do that efficiently. By default, LSMatch keeps all the alignments with a combined score (Levenshtein + Synonym) of 0.5 or above to check the alignments over variable thresholds. For the final selection of alignments the current version of LSMatch has used 0.95 as the threshold.

## 2. Results

This section describes the results of the LSMatch and LSMatch-multilingual system collectively on six tracks namely: Anatomy, Conference, Multifarm, Bio-ML, Common Knowledge Graphs, and Knowledge Graph. The results are presented collectively in Table 1. Differences from OAEI2021 are discussed in the subsections below.

<sup>1</sup><https://xlinux.nist.gov/dads/HTML/Levenshtein.html>

**Table 1**

Result summary of LSMatch at OAEI 2022 and OAEI 2021

Task	Year	Precision	F1	Recall									
--Anatomy--													
Mouse-Human	2022	0.952	0.761	0.634									
Mouse-Human	2021	0.997	0.763	0.618									
--Conference--													
OntoFarm (rar2-M3)	2022	0.83	0.55	0.41									
OntoFarm (rar2-M3)	2021	0.83	0.55	0.41									
OntoFarm (Sharp)	2022	0.88	0.57	0.42									
OntoFarm (Sharp)	2021	0.88	0.57	0.42									
OntoFarm (Discrete)	2022	0.87	0.66	0.53									
OntoFarm (Discrete)	2021	0.88	0.66	0.53									
OntoFarm (Continuous)	2022	0.88	0.67	0.54									
OntoFarm (Continuous)	2021	0.88	0.67	0.54									
DBpedia-OntoFarm	2022	0.5	0.55	0.6									
DBpedia-OntoFarm	2021	0.5	0.55	0.6									
--Bio-ML (Unsupervised (90% Test Mapping))--													
Equivalent Matching Results for OMIM-ORDO (Disease)	2022	0.65	0.329	0.221									
Equivalent Matching Results for NCIT-DOID (Disease)	2022	0.719	0.633	0.565									
Equivalent Matching Results for SNOMED-FMA (Body)	2022	0.809	0.132	0.072									
Equivalent Matching Results for SNOMED-NCIT (Pharm)	2022	0.982	0.706	0.551									
Equivalent Matching Results for SNOMED-NCIT (Neoplas)	2022	0.902	0.377	0.238									
--Bio-ML (Semi-supervised (70% Test Mapping))--													
Equivalent Matching Results for OMIM-ORDO (Disease)	2022	0.594	0.325	0.223									
Equivalent Matching Results for NCIT-DOID (Disease)	2022	0.665	0.611	0.565									
Equivalent Matching Results for SNOMED-FMA (Body)	2022	0.762	0.128	0.07									
Equivalent Matching Results for SNOMED-NCIT (Pharm)	2022	0.976	0.702	0.548									
Equivalent Matching Results for SNOMED-NCIT (Neoplas)	2022	0.877	0.374	0.238									
--Large BioMed and Disease & Phenotype track (2021)--													
FMA-NCI small	2021	0.979	0.876	0.792									
FMA-SNOMED small	2021	0.988	0.33	0.198									
HP-MP task	2021	1	0.421	0.267									
DOID-ORDO task	2021	1	0.463	0.301									
--Common KG Track--													
Nell-DBPedia	2022	0.96	0.84	0.75									
Nell-DBPedia	2021	0.99	0.87	0.78									
Yago-Wikidata	2022	0.96	0.76	0.63									
--Knowledge Graph Track--													
Task	Year	Class			Property			Instance			Overall		
		P	F1	R	P	F1	R	P	F1	R	P	F1	R
-	2022	0.97	0.78	0.64	0.73	0.71	0.69	0.66	0.63	0.6	0.66	0.63	0.61
-	2021	1	0.78	0.64	0	0	0	0	0	0	1	0.01	0

## 2.1. Anatomy

In anatomy overall result is almost same as last year with 2% improvement in recall, though overall F-measure got affected and it decreased by 0.2%.

## 2.2. Conference

For conference track the result are exactly same as last year as due to some error we had to use the last year's LSMatch for this track.

## 2.3. Multifarm

This is the first entry of LSMatch in Multifarm track. For this track we specifically developed LSMatch-multilingual. Though both the versions of LSMatch were tested on Multifarm track, LSMatch-multilingual obtained best overall result among all the systems with 0.47 F1 score.

## 2.4. Bio-ML

The Bio-ML track is Machine Learning (ML) friendly Biomedical track. This track supersedes the previous largebio and phenotype tracks. There are 5 tasks in total (on which LSMatch was tested), all Equivalent matching have been performed with 5 ontology pairs, OMIN-ORDO(Disease), NCIT-DOID(Disease), SNOMED-FMA(Body), SNOMED-NCIT(Pharm), and SNOMED-NCIT(Neoplas). On OMIN-ORDO(Disease) and NCIT-DOID(Disease) LSMatch got average results. On SNOMED-FMA(Body), LSMatch has 6th best precision out of 9. On SNOMED-NCIT(Pharm) and SNOMED-NCIT(Neoplas), LSMatch has 2nd best precision just after LogMap-Lite. All the above stated results are on Unsupervised (90% Test Mapping). For Semi-supervised(70% Test Mappings), LSMatch has average performance in all tasks.

## 2.5. Common Knowledge Graphs

This year common Knowledge Graph track has one more task, namely Yago-Wikidata where LSMatch's performance was decent though need improvement. In Nell-DBPedia task, LSMatch has almost similar result to last year.

## 2.6. Knowledge Graph

In OAEI 2021 LSMatch only supported class matching, this year (OAEI 2022) LSMatch had added functionality to also match instance and properties. Class matching results this year are same as last year, with this year's property and instance matching overall result was 0.66, 0.63, and 0.61 precision, F1, and recall respectively. Which last year was 1, 0.01, and 0.

## 3. Conclusions

This year, the system was tested on six tracks, i.e., Anatomy, Conference, Multifarm, Bio-ML, Common Knowledge Graphs, and Knowledge Graph. The system achieved considerably good

precision in all the tracks but lacked behind in recall. In future versions, we will be adding a set of matchers and working to improve the utilization of background knowledge by which we can find better correlations between concepts that are not properly aligned using just the string similarity measures.

## References

- [1] S. Zhang, Y. Hu, G. Bian, Research on string similarity algorithm based on levenshtein distance, in: 2017 IEEE 2nd Advanced Information Technology, Electronic and Automation Control Conference (IAEAC), IEEE, 2017, pp. 2247–2251.
- [2] S. Hertling, J. Portisch, H. Paulheim, Melt-matching evaluation toolkit, in: International conference on semantic systems, Springer, Cham, 2019, pp. 231–245.
- [3] P. Shvaiko, J. Euzenat, Ontology matching: state of the art and future challenges, IEEE Transactions on knowledge and data engineering 25 (2011) 158–176.
- [4] T. T. A. Nguyen, S. Conrad, Ontology matching using multiple similarity measures, in: 2015 7th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management (IC3K), volume 1, IEEE, 2015, pp. 603–611.
- [5] Z. Aleksovski, W. Ten Kate, F. Van Harmelen, Exploiting the structure of background knowledge used in ontology matching., in: Ontology Matching, 2006, p. 13.
- [6] G. A. Miller, Wordnet: a lexical database for english, Communications of the ACM 38 (1995) 39–41.
- [7] P. Ochieng, S. Kyanda, Large-scale ontology matching: State-of-the-art analysis, ACM Computing Surveys (CSUR) 51 (2018) 1–35.
- [8] S. Anam, Y. S. Kim, B. H. Kang, Q. Liu, Review of ontology matching approaches and challenges, International Journal of Computer Science and Network Solutions 3 (2015) 1–27.