

FCA-Map Results for OAEI 2016

Mengyi Zhao¹ and Songmao Zhang²

^{1,2}Institute of Mathematics, Academy of Mathematics and Systems Science,
Chinese Academy of Sciences, Beijing, P. R. China

¹myzhao@amss.ac.cn, ²smzhang@math.ac.cn

Abstract. FCA-Map is an automatic ontology matching system based on Formal Concept Analysis (FCA), which is a well developed mathematical model for analyzing individuals and structuring concepts. More precisely, we construct three types of formal contexts and extracts mappings from the lattices derived. Firstly, token-based formal context describes how class names, labels and synonyms share lexical tokens, leading to lexical mappings (anchors) across ontologies. Secondly, relation-based formal context describes how classes are in taxonomic or disjoint relationships with the anchors, leading to positive and negative structural evidence for validating the lexical matching. Lastly, after incoherence repair, positive relation-based context can be used to discover additional structural mappings. In this paper, we briefly introduce FCA-Map and its results of three tracks (i.e., Anatomy, Large Biomedical Ontologies, Disease and Phenotype) on OAEI 2016.

1 Presentation of the system

Among the first batch of OM algorithms and tools proposed in the early 2000s, FCA-Merge [4] distinguished in using Formal Concept Analysis (FCA) formalism to derive mappings from classes sharing textual documents as their individuals. Proposed by Wille [5], FCA is a well developed mathematical model for analyzing individuals and structuring concepts. FCA starts with a formal context consisting of a set of objects, a set of attributes, and their binary relations. Concept lattice, or Galois lattice, can be computed based on formal context, where each node represents a formal concept composed of a subset of objects (extent) with their common attributes (intent). The extent and the intent of a formal concept uniquely determine each other in the lattice. Further, a concept hierarchy can be derived where one formal concept becomes sub-concept of the other if its objects are contained in the latter. FCA can be naturally applied to ontology construction [3], and is also widely used in data analysis, information retrieval, and knowledge discovery.

Following the steps of FCA-Merge, several OM systems continued to use FCA as well as its alternative formalisms, exploiting different entities as the sets of objects and attributes for constructing formal contexts [1, 2, 6]. Different types of formal contexts decide the information used for ontology matching, and we observed that some intrinsic and essential knowledge of ontology has not been involved yet, including both textual information within classes (e.g., class names, labels, and synonyms) and relationships among classes (e.g., ISA, sibling, and disjointness relations). In order to empower

FCA with as much as ontological information as possible, we proposed FCA-Map, which generates three types of formal contexts and extracts mappings from the lattices derived. The next sub-sections provide more details about FCA-Map and then discuss our results of OAEI.

1.1 State, purpose, general statement

Given two ontologies, FCA-Map builds formal contexts and uses the derived concept lattices to cluster the commonalities among ontology classes, at lexical level and structural level, respectively. Concretely, FCA-Map performs step-by-step as follows.

1. **Acquiring anchors lexically.** The token-based formal context is constructed, and from its derived concept lattice, a group of lexical anchors \mathcal{A} across ontologies can be extracted.
2. **Validating anchors structurally.** Based on \mathcal{A} , the relation-based formal context is constructed, and from its derived concept lattice, positive and negative structural evidence of anchors can be extracted. Moreover, an enhanced alignment \mathcal{A}' without incoherences among anchors is obtained.
3. **Discovering additional matches.** Based on \mathcal{A}' , the positive relation-based formal context is constructed, and from its derived concept lattice, additional matches across ontologies can be identified.

1.2 Specific techniques used

The process of our system consists of the following successive steps.

Step 1: Constructing the token-based formal context to acquire lexical anchors.

The token-based formal context $\mathbb{K}_{lex} := (G_{lex}, M_{lex}, I_{lex})$ is described as follows. Names of ontology classes as well as their labels and synonyms, when available, are exploited after normalization that includes inflection, tokenization, stop word elimination, and punctuation elimination. In \mathbb{K}_{lex} , G_{lex} is the set of strings each corresponding to a name, label, or synonym of classes in two ontologies, M_{lex} is the set of tokens in these strings, and binary relation $(g, m) \in I_{lex}$ holds when string g contains token m , or a synonym or lexical variation of m . For the derived formal concepts, we restrict our attention to formal concepts whose *simplified extent* or *class-origin extent* contains exactly two strings or classes across ontologies, and extract two types of lexical anchors, namely **Type I anchor** for the exact match, and **Type II anchor** for the partial match, respectively.

Sept 2: Constructing the relation-based formal context to validate lexical anchors. Structural relationships of ontologies are exploited to validate the matches obtained at the lexical level. [7] proposed using positive and negative structural evidence among anchors for the purpose of validation. In this step, we build the relation-based formal context to obtain both positive and negative structural evidence for lexical anchors. The relation-based formal context $\mathbb{K}_{rel} := (G_{rel}, M_{rel}, I_{rel})$ is described as

follows. Classes in two source ontologies are taken as object set G_{rel} , and lexical anchors prefixed with different relational labels are taken as attribute set M_{rel} . For example, relationships *ISA*, *SIBLING-WITH*, *PART-OF*, and *DISJOINT-WITH* are labeled by “(ISA)”, “(SIB)”, “(PAT)”, and “(I-D)” (or “(D-I)”), respectively. Binary relation $(g, m) \in I_{rel}$ holds if g has the corresponding relationship (as in the prefix of m) with the class from the same source ontology as g in the anchor of m . Formal concepts whose extents include both classes in some anchors indicate structural evidence. Such anchors are positive evidence to anchors with label “(ISA)”, “(SIB)” or “(PAT)” in the intent, and vice versa. On the other hand, they are negative evidence to anchors with label “(I-D)” or “(D-I)” in the intent, and vice versa. In this way, positive and negative structural evidence set of each anchor a can be obtained, denoted by $P(a)$ and $N(a)$, respectively. Then we utilize all the positive evidence sets \mathcal{P} and negative evidence sets \mathcal{N} to eliminate incorrect lexical anchors and retain the correct ones.

Setp 3: Constructing the positive relation-based formal context to discover additional matches. After incoherence repair and screening, anchors retained are those supported both lexically and structurally. Based on the enhanced alignment, FCA-Map goes further to build the positive relation-based formal context aiming to identify new, structural mappings. The way positive relation-based formal context \mathbb{K}'_{rel} constructed is similar to \mathbb{K}_{rel} , i.e., using classes in two source ontologies as object set and anchors prefixed with relationship labels as attribute set, where disjointedness relationship is no longer necessary. For the derived formal concepts, we restrict our attention to those with exactly two classes across ontologies in the *simplified extent*.

1.3 Link to the system and parameters file

SEALS wrapped version of FCA-Map for OAEI 2016 is available at <https://drive.google.com/open?id=0B810qAwN1CIoM0NMV3ZJMzVsTlk>.

1.4 Link to the set of provided alignments

The results obtained by FCA-Map during OAEI 2016 are available at <https://drive.google.com/open?id=0B810qAwN1CIodGdPUjVWY0M3U0U>.

2 Results

In this section, we present the results of FCA-Map achieved on OAEI 2016. Our system mainly focuses on Anatomy, Large Biomedical Ontologies, Disease and Phenotype.

2.1 Anatomy Track

The Anatomy track consists of finding an alignment between the Adult Mouse Anatomy and a part of the NCI Thesaurus describing the human anatomy. The results are shown in Table 1. The evaluation was run on a server with 3.46 GHz (6 cores) and 8GB RAM allocated. FCA-Map ranked fifth in Anatomy track.

Matcher	Precision	Recall	F-Measure	Runtime (s)
AML	0.95	0.936	0.943	47
CroMatcher	0.949	0.902	0.925	573
XMAP	0.929	0.865	0.896	45
LogMapBio	0.888	0.896	0.892	758
FCA-Map	0.932	0.837	0.882	117

Table 1: Results for Anatomy track

2.2 Large BioMed Track

The Large BioMed track consists of finding alignments between the Foundational Model of Anatomy (FMA), SNOMED CT, and the National Cancer Institute Thesaurus (NCI). The results obtained by FCA-Map for the small fragments of the FMA, NCI and SNOMED CT ontologies are summarize in Table 2. The evaluation of first two tasks was run on a Ubuntu Laptop with an Intel Core i7-4600U CPU @ 2.10GHz x 4 and 15Gb RAM allocated with 2 hours timeout. And the last task was run on a PC with Intel i7-4790 CPU @ 3.60GHz and 8GB RAM allocated. FCA-Map ranks second in the first two tasks.

Task	Precision	Recall	F-Measure	Runtime (s)
FMA-NCI (small)	0.954	0.917	0.935	236
FMA-SNOMED (small)	0.936	0.803	0.865	1,865
SNOMED-NCI (small)	0.914	0.666	0.771	13,542

Table 2: Results of FCA-Map for the Large BioMed Track

2.3 Disease and Phenotype Track

The Pistoia Alliance Ontologies Mapping project team organises this track based on a real use case where it is required to find alignments between disease and phenotype ontologies. Specifically, the selected ontologies are the Human Phenotype Ontology (HPO), the Mammalian Phenotype Ontology (MP), the Human Disease Ontology (DOID), and the Orphanet and Rare Diseases Ontology (ORDO). The evaluation was run on a Ubuntu Laptop with an Intel Core i7-4600U CPU @ 2.10GHz x 4 and 15Gb RAM allocated.

Matcher	Task	Precision	Recall	F-Measure	Sum F-Measure	Precision	Recall	F-Measure	Sum F-Measure
		Silver 2	Silver 2	Silver 2	Silver 2	Silver 3	Silver 3	Silver 3	Silver 3
LogMap	HP-MP	0.9354	0.9125	0.9238	1.8372	0.7732	0.9729	0.8617	1.7828
	DOID-ORDO	0.9520	0.8779	0.9134		0.9052	0.9375	0.9211	
FCA-Map	HP-MP	0.9836	0.7543	0.8539	1.8162	0.9421	0.9244	0.9332	1.8706
	DOID-ORDO	0.9662	0.9586	0.9624		0.8880	0.9926	0.9374	
AML	HP-MP	0.9305	0.7998	0.8602	1.7684	0.8536	0.9446	0.8968	1.7714
	DOID-ORDO	0.8532	0.9708	0.9082		0.7784	0.9981	0.8747	
PhenoMF	HP-MP	0.7568	0.9164	0.8290	1.7149	0.6292	0.9452	0.7555	1.6905
	DOID-ORDO	0.9498	0.8301	0.8859		0.9472	0.9233	0.9351	

Table 3: Results against silver standard with vote 2 and 3

Table 3 shows the results against the silver standard which is automatically built by voting the outputs of the participating systems. LogMap is the system closer to the mappings voted by at least 2 systems, and FCA-MAP produces results very close to the silver standard with vote 3.

3 General comments

This is the first time FCA-Map system participates in the OAEI campaign. It is competitive with other systems in some tracks such as Anatomy, Large Biomedical Ontologies, Disease and Phenotype. Three types of formal contexts are constructed one-by-one, and their derived concept lattices are used to cluster the commonalities among classes at lexical and structural level, respectively. The tokens shared by two classes in these mappings are unique to their names. The lexical matching method of FCA-Map is suitable for domain ontologies having class names, labels, or synonyms from domain-specific vocabulary.

4 Conclusions

In this paper, we have presented FCA-Map and its results of three tracks (i.e., Anatomy, Large Biomedical Ontologies, Disease and Phenotype) on OAEI 2016. The evaluation results show the good performance of FCA-Map. Future work would introduce more elements of ontology into FCA-Map including properties, individuals, and logical constructors and axioms. Optimization techniques for handling large-scale FCA contexts will also be worth exploring.

Acknowledgements. This work has been supported by the National Key Research and Development Program of China under grant 2016YFB1000902, the Natural Science Foundation of China under No. 61232015, the Knowledge Innovation Program of the Chinese Academy of Sciences (CAS), Key Lab of Management, Decision and Information Systems of CAS, and Institute of Computing Technology of CAS.

References

1. de Souza, K.X.S., Davis, J.: Aligning ontologies and evaluating concept similarities. In: OT-M Confederated International Conferences” On the Move to Meaningful Internet Systems”, Springer (2004) 1012–1029
2. Guan-yu, L., Shu-peng, L., et al.: Formal concept analysis based ontology merging method. In: Computer Science and Information Technology (ICCSIT), 2010 3rd IEEE International Conference on. Volume 8., IEEE (2010) 279–282
3. Obitko, M., Snsel, V., Smid, J.: Ontology design with formal concept analysis. *CLA* **128**(3) (2004) 1377–1390
4. Stumme, G., Maedche, A.: Fca-merge: Bottom-up merging of ontologies. In: IJCAI. Volume 1. (2001) 225–230
5. Wille, R.: Restructuring lattice theory: an approach based on hierarchies of concepts. In: Ordered sets. Springer (1982) 445–470

6. Xu, X., Wu, Y., Chen, J.: Fuzzy fca based ontology mapping. In: 2010 First International Conference on Networking and Distributed Computing, IEEE (2010) 181–185
7. Zhang, S., Bodenreider, O.: Experience in aligning anatomical ontologies. International journal on Semantic Web and information systems **3**(2) (2007) 1