Minimizing the Estimated Solution Cost with A* Search to Support Minimal Mapping Repair

Inne Gartina Husein*, Benhard Sitohang, Saiful Akbar
Institut Teknologi Bandung
Jalan Ganesha No. 10 Bandung 40132, Indonesia
*Universitas Telkom
Jalan Telekomunikasi No. 1 Terusan Buah Batu Bandung 40257, Indonesia
* correspondence email : Innegh235@students.itb.ac.id

Abstract— **Incoherent alignment has been the main focus in the matching process since 2010. Incoherent means that there is semantic or logic conflict in the alignment. This condition encouraged researches in ontology matching field to improve the alignment by repairing the incoherent alignment. Repair mapping will restore the incoherent to coherent mapping, by deleting unwanted mappings from the alignment. In order to minimize the impacts in the input alignment, repair process should be done as as minimal as possible. Definition of minimal could be (1) reducing the number of deleted mappings, or (2) reducing the total amount of deleted mappings' confidence values. Repair process with new global technique conducted the repair with both minimal definitions. This technique could reduce the number of deleted mappings and total amount of confidence values at the same time. We proposed A * Search method to implement new global technique. This search method was capable to search the shortest path which representing the fewest number of deleted mappings, and also search the cheapest cost which representing the smallest total amount of deleted mappings' confidence value. A* Search was both complete and optimal to minimize mapping repair size.**

Keywords—incoherent, repair, minimal, shortest path, cheapest cost

I. INTRODUCTION

Ontology matching is the process of identifying correspondence between elements of two ontologies. The elements are classes or entities in these ontologies [1]. A set of correspondence generated by ontology matching is called alignment. Ontology matching process is automatically done by a system called matching system. Correspondence also known as mapping. Alignment can be used as a reference to build interlinking on the Linked Data (LD). Scharffe (2013) says that the ontology matching is useful in the making of accurate links. It is known that there are over 31 billion RDF statements published as linked data, but there are only 500 million links or 1.6% of the total RDF statements [2]. The weak connectivity between LD can cause the weak interoperability on various open data sources on the internet. Ontology matching becomes important and fundamental thing in building knowledge interoperability on the LD [3]. Besides being used as a reference in building links on the LD, alignment can also be used as a reference to validate and fix the links in LD [4]. An illustration of the role of ontology matching in building interlinking can be seen in the figure 2 [5].

Along with the growth in the size and complexity of ontology, an effective and efficient matching method is needed in order to deal with heterogeneous and large-sized ontologies. Various techniques have been developed to generate the quality alignment from heterogeneous and large-sized ontologies. In 2010 there has been a phenomenon of incoherent alignment which resulted in poor quality alignment [6], [7]. This phenomenon has been encouraging researches in ontology matching field to improve the alignment by repairing the incoherent alignment [8]. Incoherent means that there is semantic or logic conflict in the alignment. Repairing mapping process will change (or restore) the incoherent to coherent mapping, by *deleting* unwanted mappings from the alignment. As it is known that alignment is an important resource for building links on LD, then deleting mapping should be done as little or as minimum as possible [9]. Minimal repair will generate coherent mapping with the minimum impacts or changes in the input of alignment M (see Figure 1).

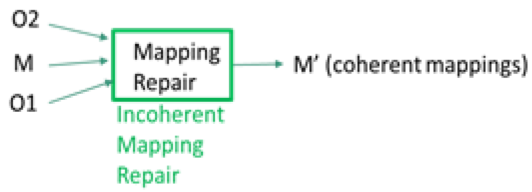

Figure 1. Mapping Repair Process

This paper does not discuss about how to find or determine the cause of incoherent mappings on an alignment, which is the first step before doing repair. This paper discuss more specific about the process of deleting unwanted mapping on an alignment. We propose the A* search method to support a repair with minimal number of deleted mappings and minimal confidence-value weighted.

The remainder of the paper is organizes as follows. Section 2 presents related work to minimal repair study. Section 3 presents solution with A* search method. Finally, section 4 provides the major conclusions.

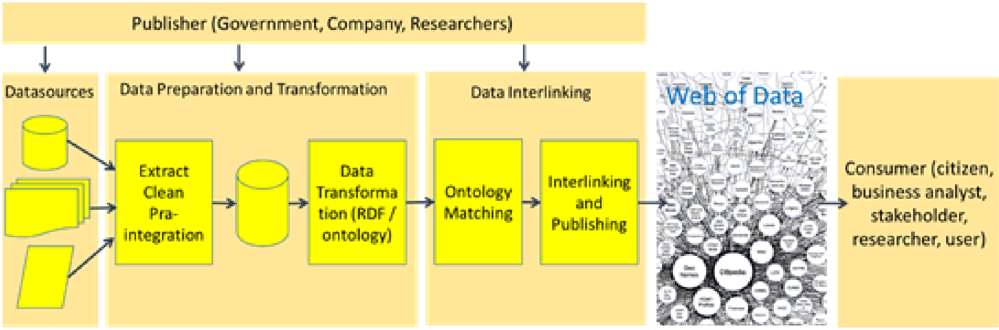

Figure 2. The Role of Ontology Matching in Building Interlinking

II. Related Work

This section explain the comparison between current repair systems and the way of search intersection mappings to be deleted to restore incoherent to coherent condition in alignment.

A. Repair Comparison

There are two different focuses in interpreting minimal mapping repair. Focus one is reducing the total amount of confidence-value of deleted mappings [8]. LogMap Repair is the repair system aimed at minimal focus one. LogMap implement local technique with Greedy search method to minimize confidence-value weighted.

Focus two is reducing the number of deleted mappings [10]. AML Repair with global technique and Heuristic search method will minimize the number of deleted mappings (or minimal focus two). Table 1 explains the differences of both systems.

Table 1 Repair System Comparison

No	Minimal Focus	System Name	Repair Technique	Search Technique
1	Reducing the total amount of confidence-value of deleted mappings	LogMap Repair	Local	Greedy Search
2	Reducing the number of deleted mappings	AML Repair	Global	Heuristic Search

An experiment conducted on 10 pairs of large-sized ontologies. The list of ontologies were being used in this experiment was shown table 2. Biomedical ontologies are the large-sized ontologies that have tens or even hundreds of thousands classes. They are semantically rich in vocabulary and known to have high levels of complexity. The biomedical ontology list can be seen in table 2 which is obtained from Bioportal[1] and OBO Foundry[2] websites. The repair size result showed that AML -with the global technique- was superior in reducing the number of the deleted unwanted mappings compared to LogMap with local technique.

Table 2 List of Biomedical Ontologies

No	Ontology	Acronym	Number of Classes	Sources
1	Bone Displasia Ontology	BDO	13,817	Bioportal
2	Cell Culture Ontology	CCONT	14,663	Bioportal
3	Experimental Factor Ontology	EFO	14,499	Bioportal
4	Cardiac Electophusiology	EP	81,957	Bioportal
5	Foundational Model of Anatomy	FMA	83,280	Bioportal
6	Mouse Adult Gross Anatomy Ontology	MA	3,205	OBO Foundry
7	NCI Thesaurus	NCI	105,347	Bioportal
8	Sleep Domain Ontology	SDO	1,382	BioPortal
9	Uber Anatomy Ontology	UBERON	15,773	OBO Foundry
10	Zebrafish Anatomy and Development Ontology	ZFA	2,955	OBO Foundry

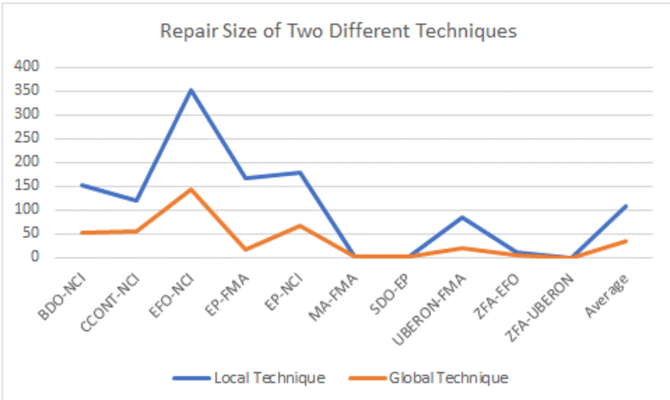

Figure 3. Repair Size Experiment Results Comparison [11]

In figure 3 it can be seen that global technique (red line) produced smaller repair size than local technique (blue line). It

[1] https://bioportal.bioontology.org

[2] www.obofoundry.org

meant global technique was better in minimizing the number of the deleted mappings, which was minimal focus two.

A study about global technique explained that the global technique ignored the calculation of the smallest confidence value of the deleted mapping [12]. Thus, global technique was not the smallest confidence weighted approach, and this became one advantage of the local technique. It meant local technique was better in minimizing the total amount of confidence value of deleted mappings, which was minimal focus one.

TABLE 3 MINIMAL EXPERIMENT COMPARISON

Technique / Minimal Focus	Local	Global	New Global
Number of deleted mappings	5	3	3
Total amount of confidence value of deleted mappings	0.37	0.274	0.21

Another study about new global technique explained the possibility to achieve two focuses minimal in one repair technique [13]. This new technique combined two focuses minimal of local and global, which were (1) minimal in the number of deleted mappings; and (2) minimal in the total amount of confidence values of deleted mappings. The result showed that *new global* technique produced the smallest number of deleted mappings and the least total amount of confidence value of deleted mappings, at the same time (see table 3). Compared to local and global technique, new global technique was closest to the goal of minimal mapping repair. In other words, new global improves the global technique in deleting unwanted mapping to restore the coherent condition.

B. *Mapping Deletion Locally and Globally*

The following will explain how to delete unwanted mapping locally, globally and new globally, to illustrate the fundamental differences in each of these techniques. First of all, a system will gather conflict mappings into conflict groups. Each mapping in conflict group (CG) is declared as incoherent mappings, so deletion must be done to restore coherent condition to other mappings (which are not deleted).

In local techniques, there will be deletion on a mapping in each CG so that the other existing mapping in the group become coherent (figure 4). Once a deletion occurs in the CG, the existing mappings (which are not deleted) will be coherent and the CG will no longer be stored on the conflict list. This technique will delete more mappings than others technique, because the more CG in system, the more mapping deletion. This will be the lack of this technique.

Global techniques will search for mapping that appear repeatedly on some CGs, and will delete this *intersection mapping*, resulting in fewer mapping deletion than local techniques (figure 5). The intersection mapping search is sorted from CG I and II, and then CG II and III, and then CG III and IV and so on. Instead of searches intersection mapping with small confidence values (to be deleted), this technique searches intersection mapping based on the CG order of the system. Thus, global technique ignores the smallest confidence value weighted, which is minimal focus two. Assuming there are 5 CGs which has a set of mappings and each confidence values, then the result of removal on global technique is as depicted in figure 6.

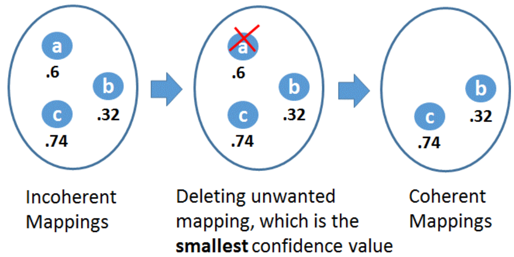

Figure 4. Mapping Deletion in Local Technique [13]

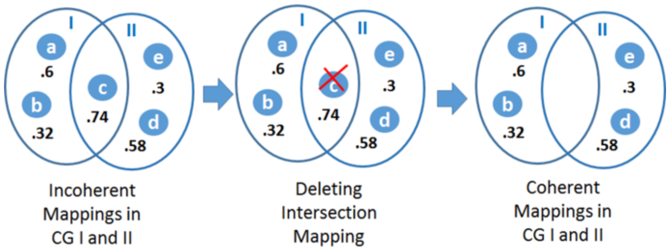

Figure 5. Mapping Deletion in Global Technique [13]

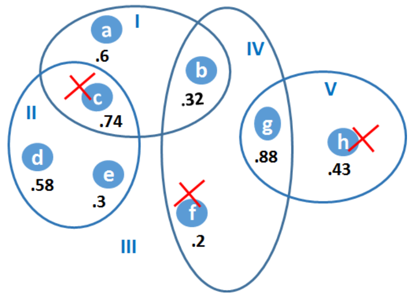

Figure 6. Intersection Searching in Global Technique

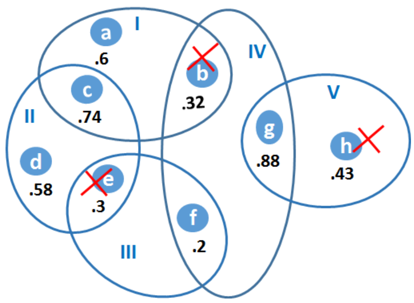

Figure 7. Intersection Searching in New Global Technique

New global technique improves the previous global technique by searching for smaller confidence value intersection mapping among the set of CGs, to be deleted. The result of removal on new global technique is as depicted in figure 7. This technique will reduce the number of deleted mappings and also reducing total amount of deleted mappings' confidence value. The result comparison of these three techniques is shown in table 3.

III. A* SEARCH

It has been discussed earlier that new global technique searches intersection mappings with smaller confidence value, to be deleted. We propose to use A* Search strategy to find optimal path as a way to find intersection mappings with smaller confidence value. A* Search will find the shortest path and also the cheapest cost as an optimal path. The shortest path will need as little as possible nodes to be expanded. The cheapest cost will find the nodes with the lowest value to be expanded.

Search algorithms work by considering various possible sequences of action. The sequence of actions that may start in the initial state form a search *tree* with the initial state at the *root*. *Branches* are actions and *nodes* correspond to states in the state space of the problem. The process of expanding nodes on the frontier continues until either a solution is found or there are no more states to expand (see figure 8). They vary mainly according to how they choose which state will be developed next, this is called search strategy [14].

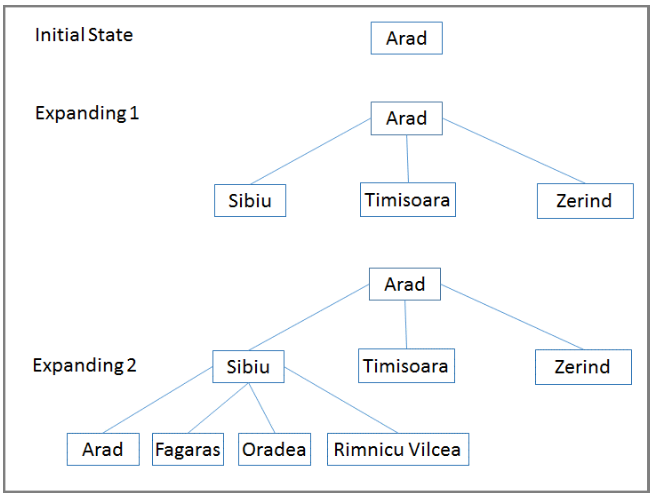

Figure 8. Search Strategy [14]

A. Heuristic Search Strategy

The search strategy with additional information about states beyond that provided in the problem definition is called informed search or heuristic search strategies. Heuristic is additional information (or knowledge) about node that have not yet been explored to decide which nodes to examine next. This knowledge guides the search algorithm and choose the next node to expand. The better (more informed) the heuristic, the fewer the nodes that need to be examined in the search tree to find a solution. In choosing the right heuristics, we usually assume that the heuristic that *reduces the number of nodes* that need to be examined in the search tree is a good heuristic [15]. In other words, the right heuristic can produce the path that closest to the goal, can be the fewest steps or the **lowest cost**.

B. Find Optimal Path with A* Search

A* search is a combination of $g(n)$ and $h(n)$ where $g(n)$ is the cost to reach the node, and $h(n)$ is the cost to get from the node to get the goal, as in :

$$f(n) = g(n) + h(n) \tag{1}$$

Since $g(n)$ gives the path cost from the start node to node n, and $h(n)$ is the estimated cost of the cheapest path from n to the goal, we have :

$f(n)$ = estimated cost of the cheapest solution through n (2)

A* search is both complete and optimal to find the cheapest solution with the lowest value of $g(n) + h(n)$ [14]. In our situation, $g(n)$ will be the confidence value of the current node

In our situation, $g(n)$ will be the confidence value of deleted mapping. The total confidence value from the current node to the final node.

and $h(n)$ will be the total sum of confidence values of deleted mappings. In other words h is the total confidence value of the current node to the final node

TABLE 4 MAPPING LIST

Mapping Code	Confidence Value
a	0.6
b	0.32
c	0.74
d	0.58
e	0.3
f	0.2
g	0.88
h	0.43

TABLE 5 CONFLICT GROUP LIST

Conflict Group	Mapping Code
I	a, b, c
II	c, d, e
III	e, f
IV	b, f, g
V	g, h

Using the case in section 2, we made table of mapping and conflict group (table 4 and table 5) as above. We want to represent these mappings and CGs into search tree element. CG I (first place in array) will be the *root*. A mapping will be the *node*. Deletion a mapping will be the *branch*. The process of expanding nodes on the frontier continues until there are *no more CG* to expand. An optimal path in this case will be the set of deleted mappings. A value of $g(n)$ will be the confidence value of deleted mapping. A value of $h(n)$ will be the total of confidence values from the current node to the final node. The logic algorithm of finding optimal path will be explained in textbox below. The calculation result of $g(n)+ h(n)$ will be shown in table 6.

Figure 9 describes the searching of optimal path using A* Search Method. The optimal path is the set of deleted mapping, which are mapping b, mapping e and mapping h. These deleted mappings have the cheapest value of f(n). The process of expanding CG will continue until no more CG in list of CG.

```
Logic Algorithm
Repeat
Take CG(array) from list of CG (ordered from I to N)
    Choose a mapping with lowest confidence value
    Compute
        save n into list of deleted mapping
        h(n) = h(n) +confidence value of deleted mapping
    Delete it (for example mapping A)
    Remove all CG containing A from list of CG
Until no more CG in list of CG
Repeat
    Compute f(n) = g(n) + h(n)
        f(n) = confidence value(n) + h(n)
Until no more mapping in list of deleted mapping
Print deleted mappings as an optimal path
```

TABLE 6 THE VALUE OF G (N) + H (N)

Expanding	node	g(node)	h(node)	f(node)
Initial State	root	0.0	(0.32 + 0.3 + 0.43) = 1.05	**1.05**
Expanding 1	a	0.6	(0.6 + 0.43) = 0.73	1.33
	b	0.32		**1.05**
	c	0.74		1.47
Expanding 2	c	(0.32 + 0.74) = 1.06	0.43	1.49
	d	(0.32 + 0.58) = 0.9		1.33
	e	(0.32 + 0.3) = 0.62		**1.05**
Expanding 3	g	(0. 62 + 0.88) = 1.5	0	1.5
	h	(0.62 + 0.43) = 1.05		**1.05**

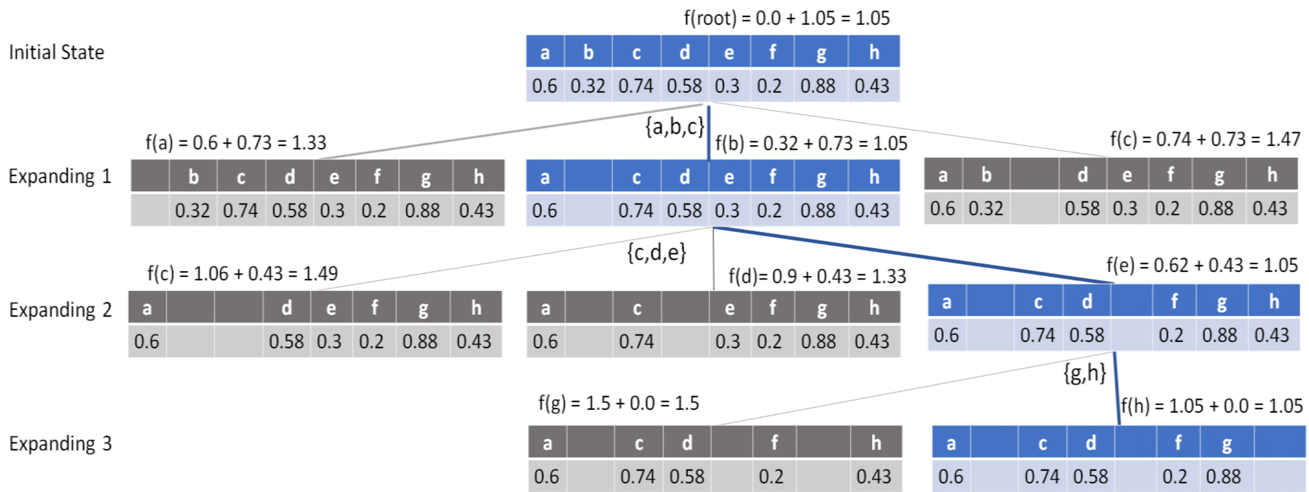

Figure 9. Finding Optimal Path using A* Search Method

A * Search method is the solution for minimizing the mapping repair. This method searches the shortest path which representing the fewest number of deleted mappings, and also searches the cheapest cost which representing the smallest total amount of deleted mappings' confidence value. Implementing this search strategy will fulfill the minimal with two focuses, and improve the quality of alignment.

IV. CONCLUSION

Ontology matching is the process of identifying correspondence between elements of two ontologies. Output this process is a set of correspondence, known as alignment. There had been a phenomenon of incoherent alignment which resulted in decreased quality of alignment. This phenomenon had been encouraging researches to repair the alignment. Incoherent meant that there was semantic or logic conflict in alignment. Repairing mapping process would restore the incoherent to coherent mapping, by deleting unwanted mappings from the alignment. But then we learnt that deleting mapping should be done as as minimum as possible, in order to minimize the impacts in the input alignment. Definition of minimal could be (1) reducing the number of deleted mappings, or (2) reducing the total amount of deleted mappings' confidence values. After comparing local and global techniques in mapping repair, we found that new global technique were much more better due to repair with two minimal focus. This technique could reduce the number of deleted mappings and total amount of deleted mappings' confidence values at the same time. We proposed A * Search method to implement new global technique. This search method was capable to search the shortest path which representing the fewest number of deleted mappings, and also search the cheapest cost which representing the smallest total amount of deleted mappings' confidence value. A* Search was both complete and optimal to minimize mapping repair size.

Acknowledgment

This research was supported by Directorate General of Resource for Science, Technology and Higher Education, Indonesia. We thank Dr. Giabbanelli and Dr. Freedman who provided expertise and materials that greatly assisted the research.

References

[1] L. Otero-cerdeira, F. J. Rodríguez-martínez, and A. Gómez-rodríguez, "Ontology Matching : A literature review," *Int. J. Expert Syst. With Appl.*, vol. 42, no. 2, pp. 949–971, 2015.

[2] V. Bryl, C. Bizer, R. Isele, M. Verlic, S. G. Hong, S. Jang, M. Y. Yi, and K. Choi, "Interlinking dan Knowledge Fusion," in *LNSC 8661*, 2014, vol. 8661, pp. 70–89.

[3] M. Ehrig, *Ontology Alignment - Bridging the Semantic Gap*. Springer, 2007.

[4] J. Euzenat and P. Shvaiko, *Ontology Matching*. Springer, 2013.

[5] I. G. Husein, B. Sitohang, and S. Akbar, "Peningkatan Interlinking pada Linked Data Melalui Ontology Alignment," in *Seminar Nasional Teknologi Informasi dan Multimedia 2016*, 2016.

[6] J. Euzenat, A. Ferrara, C. Meilicke, A. Nikolov, J. Pane, F. Scharffe, P. Shvaiko, H. Stuckenschmidt, O. Zamazal, V. Svatek, and C. Trojahn, "Results of the Ontology Alignment Evaluation Initiative 2010," in *Proc. 8th ISWC workshop on ontology matching (OM)*, 2013, pp. 61–100.

[7] P. Lambrix, G. Qi, and C. Meilicke, "ISWC tutorial 2011," 2011. [Online]. Available on : http://www.ida.liu.se/~patla00/courses/ISWC11/. [Accessed: 26-Mar-2016].

[8] E. Jiménez-Ruiz and B. Cuenca Grau, "LogMap: Logic-based dan scalable ontology matching," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2011, vol. 7031 LNCS, no. PART 1, pp. 273–288.

[9] E. Santos, D. Faria, C. Pesquita, and F. Couto, "Ontology alignment repair through modularization and confidence-based heuristics," 2013.

[10] I. F. Cruz, F. P. Antonelli, and C. Stroe, "AgreementMaker : Efficient Matching for Large Real-World Schemas and Ontologies," *Pvldb*, vol. 2, no. 2, pp. 1586–1589, 2009.

[11] D. Faria, C. Pesquita, E. Santos, M. Palmonari, I. F. Cruz, and F. M. Couto, "The AgreementMakerLight ontology matching system," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2013, vol. 8185 LNCS, pp. 527–541.

[12] E. Jiménez-Ruiz, C. Meilicke, B. C. Grau, and I. Horrocks, "Evaluating mapping repair systems with large biomedical ontologies," in *CEUR Workshop Proceedings*, 2013, vol. 1014, pp. 1000–1010.

[13] I. G. Husein, B. Sitohang, S. Akbar, and F. N. Azizah, "Comparisons of Diagnosis in Mapping Repair Systems," in *2016 International Conference on Data and Software Engineering (ICoDSE)*, 2016.

[14] S. Russell and P. Norvig, *Artifical Intelligence: A Modern Approach (Third Edition)*. Prentice Hall, 2010.

[15] B. Coppin, *Artificial Intelligence Illuminated*, 1st ed. Canada: Jones and Bartlett Publishers, 2004.