the highest almost every time (in the range of $0.88 - 1$). This is due to the fact that given a and b are semantically similar, they satisfy that similar relations (a with b (SM_1), hypernyms of a with hypernyms of b (SM_5), and hyponyms of a with hyponyms of b (SM_9)) are more likely to be similar than different relations (a with hypernyms of b (SM_2), a with hyponyms of b (SM_3), hypernyms of a with b (SM_4), hypernyms of a with hyponyms of b (SM_6), hyponyms of a with b (SM_7), and hyponyms of a with hypernyms of b (SM_8)). Thus, the similarity value between a and b is calculated as follows:

$$\begin{cases} Sim_{words}(a,b) = 1, \text{ if } a \text{ and } b \text{ are synonyms or one of them is a direct hyponym of the other} \\ Sim_{words}(a,b) = 0, \text{ if } [0.8 \times (SM_1 + SM_5 + SM_9) + 0.2 \times \sum_{\substack{i=2 \\ i \neq 5}}^{8} SM_i] \times \exp \frac{\sum_{\substack{i=1 \\ SM_i \neq 0}}^{9} 1}{9} \leq 1 \\ Sim_{words}(a,b) = \dfrac{[0.8 \times (SM_1 + SM_5 + SM_9) + 0.2 \times \sum_{\substack{i=2 \\ i \neq 5}}^{8} SM_i] \times \exp \frac{\sum_{\substack{i=1 \\ SM_i \neq 0}}^{9} 1}{9} - 1}{[0.8 \times (SM_1 + SM_5 + SM_9) + 0.2 \times \sum_{\substack{i=2 \\ i \neq 5}}^{8} SM_i] \times \exp \frac{\sum_{\substack{i=1 \\ SM_i \neq 0}}^{9} 1}{9} + 1} \text{, otherwise} \end{cases} \tag{16}$$

Step 2: Calculating the semantic similarity between sets of words. The similarity calculator uses the similarity measure between words (16) to compute the similarity between sets of words. Given two *entities* $e_1 \in S_1$ and $e_2 \in S_2$. Let $set_{e_1} = \{W_{1,1}, W_{1,2}, \ldots, W_{1,card(set_{e_1})}\}$ and $set_{e_2} = \{W_{2,1}, W_{2,2}, \ldots, W_{2,card(set_{e_2})}\}$ be their respective sets of words. The similarity calculator uses equation (17) to calculate the similarity between set_{e_1} and set_{e_2}.

$$Sim_{sets}(set_{e_1}, set_{e_2}) = \frac{1}{min(card(set_{e_1}), card(set_{e_2}))} \times \left(\sum_{i=1}^{card(set_{e_1})} max(m_{i,j})_{1 \leq j \leq card(set_{e_2})} \right) \tag{17}$$

Where $M = (m_{i,j})_{\substack{1 \leq i \leq card(set_{e_1}) \\ 1 \leq j \leq card(set_{e_2})}}$ is the similarity matrix. Its individual items are defined as follows $m_{i,j} = Sim_{words}(W_{1_i}, W_{2_j})$.

Next, we define the matches based on the similarity values.

C. The Post-matching Module

We applied our similarity measure (17) on the semantically related sets of words from the TEL schemas. The results formed a set of similarity values, each represents the similarity between two sets. The process of selecting the threshold value was based on reference matches we defined manually in order to identify the range of similarity values generated for semantically similar sets. We noticed that most matching sets have a similarity value greater than or equal to 0.8. Hence, we defined the threshold value 0.8 beyond which the pair of *entities* must be matched.

The post-matching module consists mainly of one major component, namely the *matches generator*, which uses the threshold value to eliminate *entity* pairs with very low similarity values, and match only pairs with high similarity values (≥ 0.8). Algorithm 5 summarizes this.

Algorithm 5 MatchesGenerator($SETS'_1$, $SETS'_2$, V)

Input:
$SETS'_1$
$SETS'_2$
V
Output:
$Matches$

```
for each v in V do
    if (v ≥ 0.8) then
        Matches ← Matches ∪ (set_1, set_2)
    end if
end for
return Matches
```

V. Experimental Results

We conducted extensive experiments to evaluate xMatcher based on a real implementation. We focused on evaluating two major issues. (1) We verified the accuracy of the results of our similarity measure, by evaluating the correlation coefficient and the Mean Square Error. (2) We examined the accuracy of the matches generated by xMatcher, by evaluating *Precision*, *Recall*, *Overall*, and *F-Measure*.

A. Experimental Setup

Datasets: First, we experimented our measure on M&C dataset [24], which contains thirty word pairs (see Table IV). We then experimented xMatcher over the *Conference Track* used in OAEI 2018 and available on the Web[3]. The *Conference Track* involves 16 ontologies describing the domain of organizing academic conferences. It has been used by the research community for over 13 years. It has 21 reference alignments composed from 7 out of 16 real domain ontologies.

Implementation: In addition to our measure, we implemented four measures and distances Resnik, J&C, Lin, and Nababteh over WordNet. Then, we implemented xMatcher. Finally, since xMatcher was initially developed to take as input XML schemas and since the *Conference Track* includes ontologies, we implemented the converting process presented in [30] to transform ontologies into XML schemas.

Measures: For semantic similarity values (produced by all five measures), we used the correlation coefficient and Mean Square Error (MSE) to compare the returned results with the reference results [24]. The correlation coefficient measures how strong the relationship is between the returned values and the reference results. MSE measures the average of the squares of the errors between the returned values and the reference results. The lower the MSE is, the better.

For matching results, we used the previously published results produced by twelve ontology matching systems (SANOM [31], AML [13], LogMap [32], XMap [33], KEPLER [34], ALIN [9], DOME [11], Holontology [10], FCAMapX [35], [36], LogMapLt [32], ALOD2Vec [12], and Lily [37]) that

[3] http://oaei.ontologymatching.org/2018/

TABLE IV. SEMANTIC SIMILARITY VALUES BY WORD PAIR

Word pair	M&C	Resnik	J&C	Lin	Nababteh	Our measure
Automobile / Car	0.98	0.9962	1	1	1	1
Journey / Voyage	0.96	0.9907	0.9165	0.8277	0.857335	1
Gem / Jewel	0.96	1	1	0.2434	0.31453	1
Boy / Lad	0.94	0.9971	0.8613	0.6433	1	1
Coast / Shore	0.925	0.9994	0.9567	0.96	1	1
Asylum / Madhouse	0.9025	1	0.9379	0.769	0.879	1
Magician / Wizard	0.875	0.9999	1	0.1958	0.28158	1
Midday / Noon	0.855	0.9998	1	1	1	1
Furnace / Stove	0.7775	0.6951	0.593	0.2294	0.26674	0.79
Food / Fruit	0.77	0.9689	0.7925	0.0956	0.103839	0.98
Bird / Cock	0.7625	0.9984	0.8767	0.7881	0.930014	1
Bird / Crane	0.7425	0.9984	0.815	0	0.850943	0.95
Implement / Tool	0.7375	0.9852	0.977	0.914	1	1
Brother / Monk	0.705	0.8722	0.6656	0	1	1
Crane / Implement	0.42	0.8722	0.6526	0	0.513459	0.73
Brother / Lad	0.415	0.8693	0.6775	0.24	0.29735	0.62
Car / Journey	0.29	0	0.5883	0	0	0
Monk / Oracle	0.275	0.8722	0.6203	0.1828	0.191595	0.75
Food / Rooster	0.2225	0.5036	0.5885	0.0762	0.095302	0.66
Coast / Hill	0.2175	0.9867	0.8487	0.127	0.19414	0.49
Forest / Graveyard	0.21	0	0.484	0.1119	0.1706	0.61
Monk / Slave	0.1375	0.8722	0.6962	0.2011	0.34281	0.25
Coast / Forest	0.105	0	0.5179	0	0	0.2
Lad / Wizard	0.105	0.8722	0.6905	0.2241	0.34155	0.2
Cord / Smile	0.0325	0.8044	0.5845	0	0	0.11
Glass / Magician	0.0275	0.5036	0.5699	0.0663	0.09335	0
Rooster / Voyage	0.02	0	0.4168	0	0	0
Noon / String	0.02	0	0.4329	0	0	0

participated in OAEI 2018 over the *Conference Track*. We used $Precision$ (18), $Recall$ (19), $Overall$ (20), and $F-Measure$ (21) [38] to evaluate the returned matches based on nine combinations of evaluation variants with crisp reference alignments: *ra1-M1*, *ra1-M2*, *ra1-M3*, *ra2-M1*, *ra2-M2*, *ra2-M3*, *rar2-M1*, *rar2-M2*, and *rar2-M3* (*ra1* is the original reference alignment; *ra2* is an extension of ra1; and *rar2* is an updated version of *ra2* that deals with violations of conservativity). *ra1-M1*, *ra2-M1*, and *rar2-M1* are used to evaluate only alignments between classes; *ra1-M2*, *ra2-M2*, and *rar2-M2* are used to evaluate only alignments between properties; and *ra1-M3*, *ra2-M3*, and *rar2-M3* are used to evaluate both alignments between classes and properties.

$$Precision = \frac{Correct\ Matches}{Correct\ Matches + Incorrect\ Matches} \quad (18)$$

(18) is the probability of correct matches among the matches returned by a matching system.

$$Recall = \frac{Correct\ Matches}{Missed\ Matches + Correct\ Matches} \quad (19)$$

(19) is the probability of correct matches returned by a matching system among the reference matches.

$$Overall = Recall \times (2 - \frac{1}{Precision}) \quad (20)$$

(20) quantifies the amount of manual post-effort necessary to remove false matches and add missed matches.

$$F - Measure = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (21)$$

(21) is the harmonic mean of *Precision* and *Recall*.

B. Results and Discussion

1) Semantic Similarity Measure: Experiment Results: We first applied our measure, Resnik, J&C, Lin, and Nababteh on M&C dataset (see the results in Table IV). We then used the results to calculate the correlation coefficient and MSE (see the overall results in Table V and the details about correlations in Fig. 3(a), Fig. 3(b), Fig. 3(c), Fig. 3(d), and Fig. 3(e)).

TABLE V. COMPARISON BETWEEN SOME STATE OF THE ART SIMILARITY MEASURES AND OUR MEASURE

Measure	Correlation coefficient	MSE
Resnik	0.6671	0.1373
J&C	0.8363	0.1018
Lin	0.6852	0.1188
Nababteh	0.7654	0.0699
Our measure	0.9102	0.0453

The findings indicate a strong positive correlation ($+0.9102$) between our measure and the reference results. They also indicate that our measure obtained the smallest MSE (0.0453) compared to the other measures. Thus, our measure outperforms the state of the art measures, showing that on the one hand information content-based measures cannot provide high accuracy results; on the other hand combining different WordNet information (hypernyms, direct hyponyms, senses, and *synsets*) is a good plus to obtain high accuracy results.

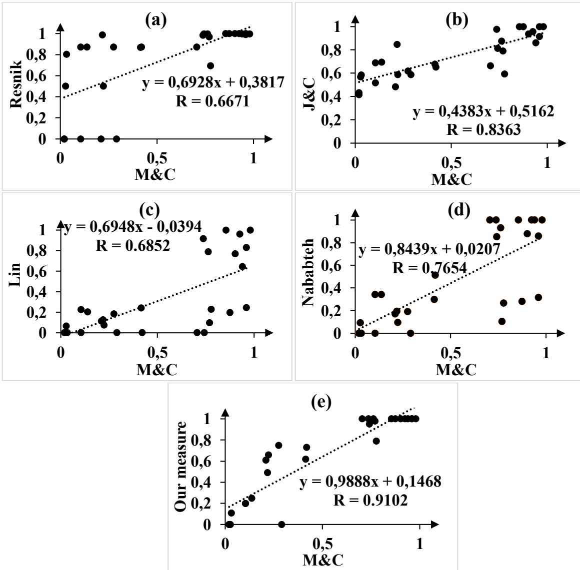

Fig. 3. Regression lines for (a) Resnik vs. M&C; (b) J&C vs. M&C; (c) Lin vs. M&C; (d) Nababteh vs. M&C; and (e) Our measure vs. M&C

2) xMatcher: Experiment Results: We first generated the matches using xMatcher. We then calculated, for all matches for which there is a reference alignment, *Precision*, *Recall*, *Overall*, and *F-Measure* nine times, each time with a different reference alignment. Fig. 4(a), Fig. 4(b), Fig. 4(c), Fig. 4(d), Fig. 4(e), Fig. 4(f), Fig. 4(g), Fig. 4(h), and Fig. 4(i) present the new and previously published results.

On the one hand, the previously published results indicate noticeable changes in *Precision*, *Recall*, *Overall*, and *F-Measure*: overall, they achieved good matching accuracy when evaluated based on *ra1-M1*, *ra1-M3*, *ra2-M1*, *ra2-M3*, *rar2-M1*, and *rar2-M3*; and low accuracy even null sometimes (Lily and ALIN) with *ra1-M2*, *ra2-M2*, and *rar2-M2*. On the other hand, xMatcher obtained high accuracy matches, outperforming all systems almost every time except from *ra1-M2* and *ra2-M2* where AML surpassed it slightly ($Precision = 1$).

While xMatcher matches both classes and properties, Lily and ALIN match only classes the reason why they failed to produce high accuracy matches with *ra1-M2*, *ra2-M2*, and *rar2-M2*; SANOM, AML, LogMap, and XMap match some but not all properties which explain their negative *Overall* with *ra1-M2*, *ra2-M2*, and *rar2-M2*; and KEPLER, DOME, Holontology, FCAMapX, LogMapLt, and ALOD2Vec match very few properties which justify their negative *Overall* and low *Precision*, *Recall*, and *F-Measure* with *ra1-M2*, *ra2-M2*, and *rar2-M2*. We can conclude that (1) SANOM, AML, LogMap, XMap, KEPLER, ALIN, DOME, Holontology, FCAMapX, LogMapLt, ALOD2Vec, and Lily work well with the reference alignments that consider classes or both classes and properties. However, they fail to match correctly with the reference alignments that consider only properties; and (2) xMatcher succeeds to achieve superior accuracy matches regardless of the reference alignment it is compared to.

Overall, xMatcher obtained the highest accuracy matches (see Fig. 4.j which displays the average matching accuracy): $Precision = 0.89$ suggests that most matches are correct;

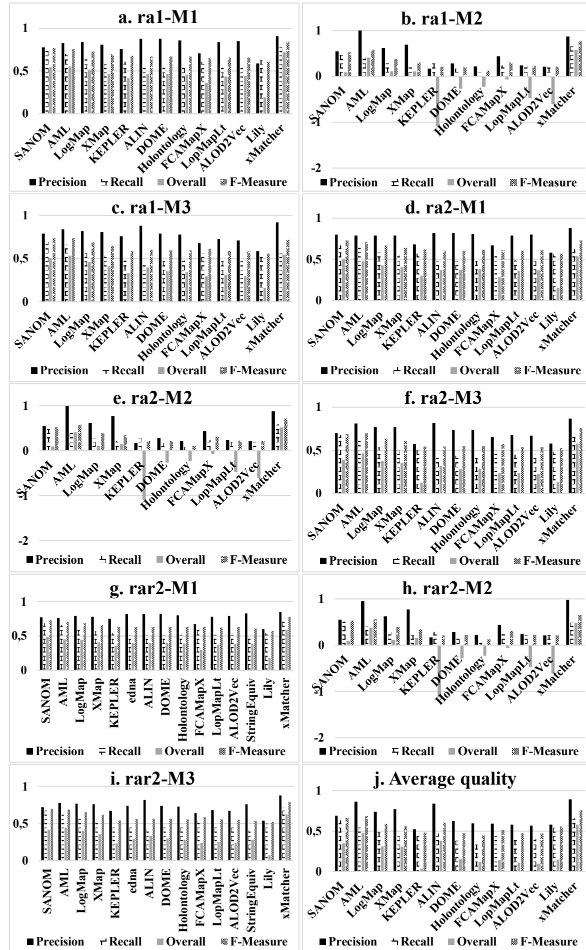

Fig. 4. Accuracy of the Matches

$Recall = 0.66$ suggests that xMatcher missed only few matches; and $Overall = 0.57$ implies that xMatcher needs only a small amount of manual post-effort to correct the results.

To prove scalability of xMatcher (note that due to space limitation, we do not display the results in figures in this paper), we applied xMatcher on more datasets, for instance the TEL (Travel, Entertainment and Living) datasets which contain five different datasets that are publicly available on the Web. The Travel group includes two various domains: *Car Rentals* and *Airfare*; the Entertainment group contains two different domains as well: *Movies* and *Books*; and, the Living group involves mainly one single domain: *Jobs*. The results show once again the capability of xMatcher to reach a high matching accuracy, which proves that xMatcher is scalable.

VI. Conclusion

We have demonstrated that the use of WordNet combined with our semantic similarity measure is an effective way to capture semantic correspondences in XML schemas. Current matching systems are error-prone and human-dependent. Thus, we have developed xMatcher, an approach to automatically match XML schemas and provide accurate matches.

Given two XML schemas S_1 and S_2, our main idea is to first generate sets of words from S_1 and S_2, then determine semantically related sets, and finally identify semantic correspondences between related sets. We evaluated xMatcher over the *Conference Track*. The results show that xMatcher achieves better accuracy than twelve state of the art matching systems. Future research includes the following:

- **Improving the accuracy of the matches.** An interesting direction is to achieve better correlation, MSE, *Precision*, *Recall*, *Overall*, and *F-Measure*.
- **Considering other matching quality factors.** In this paper, we focused on achieving high matching accuracy. A future direction is to propose techniques that consider other quality factors.
- **Matching other data representations.** xMatcher takes as input XML schemas. An interesting direction is to match different data representations.

References

[1] C. Zhang, L. Chen, H. Jagadish, M. Zhang, and Y. Tong, "Reducing uncertainty of schema matching via crowdsourcing with accuracy rates," *IEEE Transactions on Knowledge and Data Engineering*, 2018.

[2] Y. Lee, M. Sayyadian, A. Doan, and A. S. Rosenthal, "etuner: tuning schema matching software using synthetic scenarios," *The VLDB Journal—The International Journal on Very Large Data Bases*, vol. 16, no. 1, pp. 97–122, 2007.

[3] L. Otero-Cerdeira, F. J. Rodríguez-Martínez, and A. Gómez-Rodríguez, "Ontology matching: A literature review," *Expert Systems with Applications*, vol. 42, no. 2, pp. 949–971, 2015.

[4] F. Ardjani, D. Bouchiha, and M. Malki, "Ontology-alignment techniques: survey and analysis," *International Journal of Modern Education and Computer Science*, vol. 7, no. 11, p. 67, 2015.

[5] L. Mukkala, J. Arvo, T. Lehtonen, T. Knuutila, *et al.*, "Current state of ontology matching. a survey of ontology and schema matching," 2015.

[6] S. Anam, Y. S. Kim, B. H. Kang, and Q. Liu, "Review of ontology matching approaches and challenges," *International journal of Computer Science and Network Solutions*, vol. 3, no. 3, pp. 1–27, 2015.

[7] G. A. Miller, "Wordnet: a lexical database for english," *Communications of the ACM*, vol. 38, no. 11, pp. 39–41, 1995.

[8] D. Faria, C. Pesquita, B. S. Balasubramani, T. Tervo, D. Carriço, R. Garrilha, F. M. Couto, and I. F. Cruz, "Results of aml participation in oaei 2018.," in *OM@ ISWC*, pp. 125–131, 2018.

[9] J. da Silva, K. Revoredo, and F. A. Baião, "ALIN results for OAEI 2018," in *Proceedings of the 13th International Workshop on Ontology Matching co-located with the 17th International Semantic Web Conference, OM@ISWC 2018, Monterey, CA, USA, October 8, 2018.*, pp. 117–124, 2018.

[10] P. Roussille, I. Megdiche Bousarsar, O. Teste, and C. Trojahn, "Holontology: results of the 2018 oaei evaluation campaign," CEUR-WS: Workshop proceedings, 2018.

[11] S. Hertling and H. Paulheim, "Dome results for oaei 2018.," in *OM@ ISWC*, pp. 144–151, 2018.

[12] J. Portisch and H. Paulheim, "Alod2vec matcher.," in *OM@ ISWC*, pp. 132–137, 2018.

[13] D. Faria, C. Pesquita, E. Santos, M. Palmonari, I. F. Cruz, and F. M. Couto, "The agreementmakerlight ontology matching system," in *OTM Confederated International Conferences" On the Move to Meaningful Internet Systems"*, pp. 527–541, Springer, 2013.

[14] I. F. Cruz, F. P. Antonelli, and C. Stroe, "Agreementmaker: efficient matching for large real-world schemas and ontologies," *Proceedings of the VLDB Endowment*, vol. 2, no. 2, pp. 1586–1589, 2009.

[15] L. Meng, R. Huang, and J. Gu, "A review of semantic similarity measures in wordnet," *International Journal of Hybrid Information Technology*, vol. 6, no. 1, pp. 1–12, 2013.

[16] B. Poorna and A. S. Ramkumar, "Semantic similarity measures: an overview and comparison," *International Journal of Advanced Research in Computer Science*, vol. 9, no. Special Issue 1, p. 100, 2018.

[17] M. H. El Yazidi, A. Zellou, and A. Idri, "Towards a fuzzy mapping for mediation systems," in *2012 IEEE International Conference on Complex Systems (ICCS)*, pp. 1–4, IEEE, 2012.

[18] A. Gupta, A. Kumar, J. Gautam, A. Gupta, M. A. Kumar, and J. Gautam, "A survey on semantic similarity measures," *IJIRST-International Journal for Innovative Research in Science & Technology*, vol. 3, p. 12, 2017.

[19] A. Yousfi, M. H. Elyazidi, and A. Zellou, "Assessing the performance of a new semantic similarity measure designed for schema matching for mediation systems," in *International Conference on Computational Collective Intelligence*, pp. 64–74, Springer, 2018.

[20] A. M. Abdelrahman and A. Kayed, "A survey on semantic similarity measures between concepts in health domain," *American Journal of Computational Mathematics*, vol. 5, no. 02, p. 204, 2015.

[21] M. H. E. Yazidi, A. Zellou, and A. Idri, "FMAMS: fuzzy mapping approach for mediation systems," *Int. J. Appl. Evol. Comput.*, vol. 4, no. 3, pp. 34–46, 2013.

[22] M. H. E. Yazidi, A. Zellou, and A. Idri, "Mapping in GAV context," in *10th International Conference on Intelligent Systems: Theories and Applications, SITA 2015, Rabat, Morocco, October 20-21, 2015*, pp. 1–5, 2015.

[23] F. Couto and A. Lamurias, "Semantic similarity definition," *Encyclopedia of bioinformatics and computational biology*, vol. 1, 2019.

[24] P. Resnik, "Using information content to evaluate semantic similarity in a taxonomy," *arXiv preprint cmp-lg/9511007*, 1995.

[25] J. J. Jiang and D. W. Conrath, "Semantic similarity based on corpus statistics and lexical taxonomy," *arXiv preprint cmp-lg/9709008*, 1997.

[26] D. Lin, "Principle-based parsing without overgeneration," in *31st annual meeting of the association for computational linguistics*, pp. 112–120, 1993.

[27] N. Mohammed and D. Mohammed, "New modified semantic similarity measure based on information content approach," *International Journal of Computer Science and Network Security (IJCSNS)*, vol. 17, no. 3, p. 73, 2017.

[28] A. Tversky, "Features of similarity.," *Psychological review*, vol. 84, no. 4, p. 327, 1977.

[29] Z. Zhou, Y. Wang, and J. Gu, "New model of semantic similarity measuring in wordnet," in *2008 3rd International Conference on Intelligent System and Knowledge Engineering*, vol. 1, pp. 256–261, IEEE, 2008.

[30] L. Mukkala, J. Arvo, T. Lehtonen, and T. Knuutila, "Trc-matcher and enhanced trc-matcher. new tools for automatic xml schema matching," 2017.

[31] M. Mohammadi, W. Hofman, and Y. Tan, "SANOM results for OAEI 2018," in *Proceedings of the 13th International Workshop on Ontology Matching co-located with the 17th International Semantic Web Conference, OM@ISWC 2018, Monterey, CA, USA, October 8, 2018.*, pp. 205–209, 2018.

[32] E. Jiménez-Ruiz, B. C. Grau, and V. Cross, "Logmap family participation in the OAEI 2018," in *Proceedings of the 13th International Workshop on Ontology Matching co-located with the 17th International Semantic Web Conference, OM@ISWC 2018, Monterey, CA, USA, October 8, 2018.*, pp. 187–191, 2018.

[33] W. E. Djeddi, S. B. Yahia, and M. T. Khadir, "Xmap: results for OAEI 2018," in *Proceedings of the 13th International Workshop on Ontology Matching co-located with the 17th International Semantic Web Conference, OM@ISWC 2018, Monterey, CA, USA, October 8, 2018.*, pp. 210–215, 2018.

[34] M. Kachroudi, G. Diallo, and S. B. Yahia, "KEPLER at OAEI 2018," in *Proceedings of the 13th International Workshop on Ontology Matching co-located with the 17th International Semantic Web Conference, OM@ISWC 2018, Monterey, CA, USA, October 8, 2018.*, pp. 173–178, 2018.

[35] M. Zhao and S. Zhang, "Fca-map results for oaei 2016.," in *OM@ ISWC*, pp. 172–177, 2016.

[36] G. Chen and S. Zhang, "Fcamapx results for OAEI 2018," in *Proceedings of the 13th International Workshop on Ontology Matching co-located with the 17th International Semantic Web Conference, OM@ISWC 2018, Monterey, CA, USA, October 8, 2018.*, pp. 160–166, 2018.

[37] Y. Tang, P. Wang, Z. Pan, and H. Liu, "Lily results for OAEI 2018," in *Proceedings of the 13th International Workshop on Ontology Matching co-located with the 17th International Semantic Web Conference, OM@ISWC 2018, Monterey, CA, USA, October 8, 2018.*, pp. 179–186, 2018.

[38] I. Kastner and F. Adriaans, "Linguistic constraints on statistical word segmentation: The role of consonants in arabic and english," *Cognitive science*, vol. 42, pp. 494–518, 2018.