# Domain Ontology Based Similarity and Analysis in Higher Education


**SOBIA HUSSAIN*, ASIF RAZA*, MUHAMMAD TANVEER MEERAN*, HAFIZ MUHAMMAD IJAZ**, AND SALMA JAMALI*****

* Department of Computer Science, Bahauddin Zakariya University, Multan, Pakistan.

** Department of Computer Science, The Islamia University of Bahawalpur, Bahawalpur, Pakistan.

*** Department of Information Technology, Quaid-e-Awam University of Engineering, Science &Technology, Nawabshah, Pakistan.

Authors E-Mail: (sobiahussain82@gmail.com, asifraza.raza14@gmail.com, tanveer_miran@yahoo.com, ijazhafiz7@gmail.com, salma.jamali@quest.edu.pk)



## ABSTRACT

**Text Re-use is a process of creating documents using existing ones. Text reuse is a common fact and rises, For example, copying text from different sites or re-using text in free blogs. Submitting someone else's work as your own, Cutting and pasting from sources without documenting, Media "borrowing" without documentation, Web publishing without permission of creators, Providing false documentation or data etc. Detecting text re-use has been studies in range of tasks and many applications. It is traditionally detected by computing similarity between contents as source text and possibly re-used text. Text similarity methods have been proposed for calculating similarity based on surface level or semantic features. But it is difficult to detect the similarity of text on concept bases. Conceptual similarity is silent problematic. Semantic based text re-use and its detection are receiving attention within the research community, where parts of certain source text has been re-used by using similar words or phrases. Ontology is explicit formal specifications of the terms in the domain and relations among them. Use of Ontologies for information extraction is common in information retrieval domain. These are limitedly used for concept mapping and extraction as well. To deal with cases of text re-use in which the text documents has been paraphrased. A DOBS (Domain Ontology Based Similarity) is proposed for detection of conceptual or topical similarity between two documents.**



**Key Words: Text Reuse, Semantics, Domain Ontology, N-Gram.**


## 1. INTRODUCTION

Plagiarism detection is a kind of Non-Topic based text classification task. Topic based and non-Topic based is the categories of the text classification. Traditionally, Textual content class is Topic based and typical example is news grouping. In Recent time, There has been and Increasing action in extent of Non-Topic classification in addition, for example in sub-tasks, genre classification, sentimentality classification, Language and encoding classification, spam identification, authorship attribution and plagiarism detection [1].

Text re-use is approach in the direction of changed making newly document from current document using text [2]. It is a standard practice in some situation (e.g. contents re-used in news media coverage. While in the other case of plagiarism it is unacceptable. In general access to web , extensive databases and media have change plagiarism into major issue for Scientists, Researches, Educational Organizations and Publishers. Most of people see copyright plagiarism as away to copy with the work of others, or acquiring the unique work of another person. Moreover terms like a "coping "&" borrowing" an disguise serious offensive. Today plagiarism is not limited to just cut and paste text. That is re-use, con also modified form expression, Sentences, Paragraphs to whole text document, change synonym and Translation Technologies and tools are giving to plagiarism a different dimension.

**Text Re-Used into Two Forms:** In syntactically, the Text document is examined by mapping linguistics structure (i.e. sentence structure). The Text is Checked by mapping linguistic structure as that is based on character, word, phrase, line, paragraphed, structure and document (Hash Function). In Semantics, The Text document checked or mapped through Ideas or concepts based that is based on concepts based word profiles, vector space demonstrate, fuzzy Semantics (WorldNet), semantics based on Ontologies [3]. Semantics copyright plagiarism. Following methods are used for plagiarism detection:

Detection of Intrinsic Plagiarism does not depend on the detection of plagiarism by any external document. In a Suspicious document, it tries to detect changes in the Writing style. For Example, if someone paid someone else to write a paper conclusion. Then conclusion was written by the same person as the rest of the paper should be detected intrinsically. The intrinsic plagiarism Detection is aimed as looking for passages within the document that tend to be semantically different from the rest of the document. To do these break the process into three stages:

Atomization-Deconstruction of the text document into passages.

Feature Extraction Function-quantity that passages style by extracting stylometric features based on text's linguistic properties. Each passage is numerically expressed as a vector of feature value.
Classification compares the feature of the vectors of the passages: Return a result as that passage was plagiarized.

In Extrinsic plagiarism Detection methods, there are six common systems that are as follows:

Plagiarism Detection based on Grammar [4]
Semantics [5]
Clustering [6]
Cross lingual plagiarism Detection [7]
Citation based [8].
Character Based Plagiarism Detection [9].

**Extrinsic Plagiarism**: Detection is second method and more traditional model of Detecting Plagiarism. Extrinsic detection of some suspicious documents and a body of source document identifies pieces of suspicious document that make Plagiarism detection from the source with other document. We might test o document for plagiarism in real-world detection tool by comparing its text to a large number of internet documents. Instead, with corpus we used our suspicious documents compared to a large set of source documents (from which all plagiarized parts were taken). Because of time constraints, we can't just compare every word of suspicious document with every word of all source documents; it would take far too long. Use fingerprinting to create compact document representation to reduce the number of comparisons made, fast computations. But, more like in intrinsic identification, first need to deal with pieces of text of reasonable size before fingerprinting and comparison can be achieved. The following concept was inspired by this:

**Atomization:** Deconstruction of a text into passages.
**Fingerprinting:** Fingerprint passage to obtain a compact representation of that passage.
**Comparison:** Use similarity measures to find similarity between fingerprints obtained from suspicious documents and multiple sources.

For intrinsic plagiarism detection strategies there are 3 general methods that are as follows:

Grammar Semantics Hybrid Plagiarism Detection [4].
Structure Based Plagiarism Detection [10].
Syntax Similarity Based Detection [11].

Some of the popular web based plagiarism detection tools are:

Smallseo
Quetext
Turnitin

After compare above tool's results we can say that "Turnitin" is very strong and efficient tool than the others [12].

## 1.1 Semantics Based Text Similarity Structures

Graph similarity detection by graph is a graph-based model that combines text representation and computation of similarity within a combined context. When using Wikipedia or any other domain concepts as background knowledge, a description of a document (i.e. bi-part) graph is constructed and a graph is calculated for check similarity. This model can overcome the problem of semantic sensitivity by using background knowledge and at the same time iteratively calculating on (i.e. bi-part) graph. Therefore, as long as their combined concepts are connected, two documents do not need to share common terms or concepts to obtain a similarity score.

Ontology is a formal explicit specification of a shared conceptualization [13]. For current web content; websites today need to be supported by formal semantic representations. The Internet will be semantic and suitable for human and machine use when it happens. An ontology goal is to respect comprehension of typical data creation ideas with several collections.

Compose analyzable and reusable domain information,

Choose domain assumptions in an express manner

In order to compose a part between prepared data and domain information.

To isolate domain information. Gruber early defined ontology as an "explicit conceptualization specification" [14].

Word Net Ontology offers us a hierarchical approach to looking for re-usability of content. We can find indistinguishable substances for all objectives and purposes by using rational arrangement techniques such as string comparison, similarity of structure, and similarity of word. WordNet passes on the same-word social affair, which is called "synsets. "Every synset has a resource/thought social affair to interconnect with each other. There are several linguistic affiliations that synsets are: Antonym (opposite), Synonyms (related), Meronym (part of), Holonym (has-a), Hypernym (super concept) and Hyponymy (sub concept) is a hierarchy. Structural classification compares semantic association in synsets. Ontology is used for recommendation that is based on semantic analysis in social networks. Intelligent recommendation based on the analysis of the interaction between user and communities. Through ontology capture the basic semantics of user's thoughts and needs [15]. In article screening process for semantic reviews, ontology helps to identify received articles. Effective concepts and relations from Ontologies can be used as semantic knowledge [16].

Plagiarism Detection Tools WS4J (WordNet Similarity for Java) [17]; this software provides APIs (Application Programming Interface) for several semantic similarity algorithms for any WordNet instance. Wu and Palmer is a word similarity algorithm used in WS4J. Through considering the depths of two synsets in WordNet taxonomies together with the depth of LCS (Least Common Subsumer), Wu and Palmer (scientists) measure similarity.

$$\text{Score} = \frac{2 \times \text{Depth}(1\text{cs})}{\left(\text{Depth} * \left(\text{s}1\right) + \text{Depth}\left(\text{sw}\right)\right)} \tag{1}$$

Equation (1) shows that '0 < score <= 1'. This value can never be zero because LCS depth is never zero (the taxonomy root depth is one). That value is one if there are two same input concepts. Required library for java WS4J for java library and two words with their part of speech are input parameters. The return value is the score of the similarity. If there is no direction between two meanings of terms, the negative number will be returned. If an error occurs, the error rate will be set to non-zero and created error string.

Limitations of the existing systems are that:

String based or pattern based plagiarism detection is only capable to detect verbatim cases of plagiarism.

Plagiarism detection will be difficult if text is paraphrased.

Strategies for syntactic similarity measures are insufficient to detect the concepts of sentences.

Ontology is used for plagiarism detection specially semantics base. In few cases a general ontology has not the ability for plagiarism detection in particular domain. Fig. 1 shows a general flow of detecting similarity of documents using Ontology. Our proposed ontology in the domain of computer hardware and software components is very effective and comprehensive for semantic plagiarism detection.

In this paper next section consists of proposed work; where we have defined about the DOBS Framework. A DOBS was developed to map the documents to ontology. We have also developed a method to test Ontology-Based semantic similarity by using N-gram similarity calculation. Then in next section we'll analyze results and introduces a synopsis of research as a whole and proposes future work that to be considered in conclusion and future work.

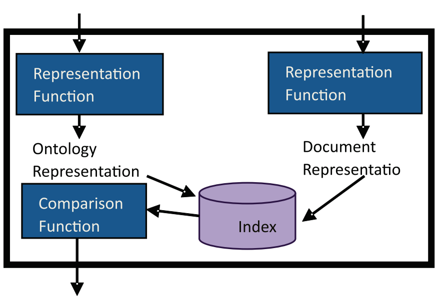


*FIG. 1. A GENERAL FLOW OF SIMILARITY DETECTION*

## 2. PROPOSED MODEL

A general flow of our model is as follows:

Fig. 2 shows framework of similarity of the documents with domain ontology. In this model similarity is detected by using token. Token are created by breaking the text documents into N-grams (uni-gram, bi-gram, tri-gram) of document. In the making of N-Grams/tokens first removing special characters and white spaces between words or before and after words in the text document.

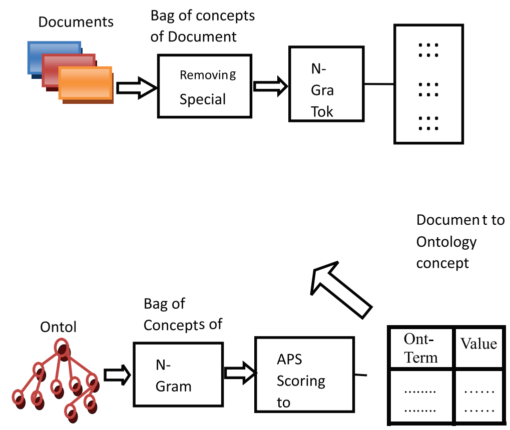


*FIG. 2. PROPOSED SYSTEM: DOBS FRAMEWORK*

If the user tries to include exceptional characters before and after those words in order to keep text detection re-useable. Our proposed algorithm will then catch similar words to measure conceptual similarity.

Ontology is the "representation of domain knowledge" in which vocabulary define a set of objectives and their relationships. In our DOBS model we use specific Domain "Computer" Ontology on computer parts, forms and their components. Fig. 3 shows computer domain ontology.

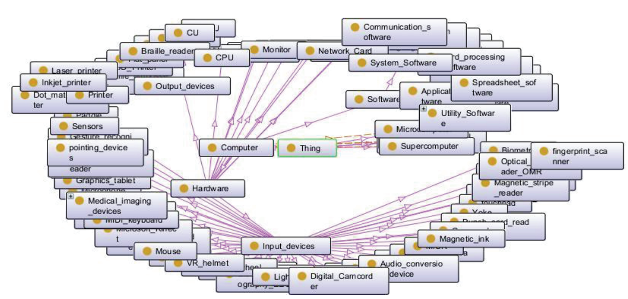


*FIG. 3. DOMAIN ONTOLOGY OF COMPUTER*

An ontology is defined as a DAG (Directed Acyclic Graph) where a node represents a primitive concept, and an edge models the relationship of binary specialization (is/a) between two concepts. Ontology thus define a hierarchy in which each term can have a set of sub-concepts known as descendants, however, not all instances of a concept must be a part of a sub-concept.

A common Ontology example is WordNet, where concepts represent groups of similar words (synonyms), and edges are is-subset-of (hypernyms) and part-of (hyponym) relations. Predicting a score that assigning to each item can be seen a problem. For example, the score could be a preference score or popularity rating. We assume that the score is a positive real-valued function that satisfies these assumptions: (1) the score depends on item features (2) each feature independently contributes to the score and (3) do not contribution to score of unknown and different features.

APS (A-Priori Score) (c) is based on concepts that is exist in ontology. The APS model expected an average user to score for each concept, but without using any using any information about the user. it is not used to predict the actual score, but only to estimate constants that decide how the actual score of the user propagates through ontology. Equation (2) calculating "APS" of concept (c) by using formula:

$$APS(c) = \frac{1}{n+1} \tag{2}$$

Where c is concept/term, n is number of node.

Fig. 4(a) an Ontology model where each node represents a concept. Fig. 4(b) the APS of those concepts by assigning the values to ontology terms in such a way that the top root level term has minimum value. Moving top/root level (general concept) to bottom level (specific concept), it has maximum value that present the scope of the concept in the domain. We take computer domain for our experiments.

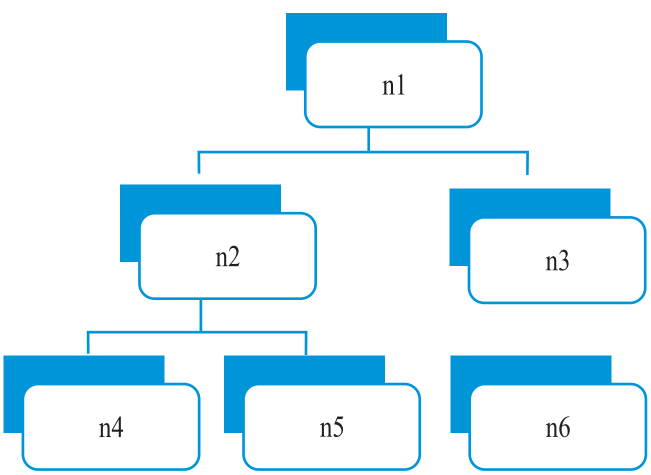


*FIG. 4(a). CONCEPTS REPRESENTATION IN ONTOLOGY MODEL*

| Concepts | $N_n$ | APS ($c_i$) |
|---|---|---|
| n1 | 5 | 0.14 |
| n2 | 2 | 0.25 |
| n3 | 1 | 0.33 |
| n4 | 0 | 0.50 |
| n5 | 0 | 0.50 |
| n6 | 0 | 0.50 |

*FIG. 4(b). A-PRIPRI-SCORE OF CONCEPTS IN ONTOLOGY MODEL*

Table 1 show the APS of Computer Domain Ontology's each level.

TABLE 1. COMPUTER DOMAIN ONTOLOGY'S LEVELS SCORE

| Root Level-1 | Level-2 | Level-3 | Level-4 | Level-5 | Value/Weight |
|---|---|---|---|---|---|
| Computer | - | - | - | - | 0.007 |
| | Hardware | - | - | - | 0.009 |
| | - | Bus | - | - | 0.06 |
| | - | - | System Bus | - | 0.20 |
| | - | - | - | Control Bus | 0.50 |
| | - | - | - | Address Bus | 0.50 |

## 3. EXPERIMENTAL ANALYSIS

Experiments are done on CLUE (Cross-Language Urdu-English) dataset [18] that are classified into three categories, (1) Fully relevant, (2) Partially relevant and (3) Non relevant to the domain ontology.

**Analyzing DOBS (Semantic Similarity) Against N-Gram (Syntactic Similarity) on Fully Relevant Documents:** We have selected 40 documents and map it to ontology. **Test document** is randomly selected from 40 documents and rests are the source documents. Here are the results:

**Fully Relevancy:** Table 2 after we get these results and compare them, we can see that Document-14 has maximum difference value showing that Document-14 is conceptual dissimilar to Document-08 whereas Document-10 and Document-10 have minimum difference values. We know that minimum difference value between two documents show the maximum similarity and vice versa. The negative difference values equal to zero.

TABLE 2. TEST DOCUMENT: 08 (DOCUMENT-08 APS = 0.1181)

| No. | Document (Number) | Source Document Ontology-to-Documents APS Score | Semantic Matching<br>Test Document -Source Documents | Syntactic Matching<br>N-Gram Document -to- Document |
|---|---|---|---|---|
| 1. | 3 | 0.1136 | 0.0045 | 0.0000 |
| 2. | 5 | 0.1084 | 0.0097 | 0.0000 |
| 3. | 10 | 0.1237 | Negative Values (0.0000) | 0.0000 |
| 4. | 14 | 0.1014 | 0.0167 | 0.0011 |
| 5. | 18 | 0.1357 | Negative Values (0.0000) | 0.0000 |
| 6. | 20 | 0.1142 | 0.0039 | 0.0000 |

Table 3 after we get these results and compare them, we can see that Document-14 has maximum difference value showing that Document-14 is conceptual dissimilar to Document-10 whereas Document-18 has minimum difference values. We know that minimum difference value between two documents show the maximum similarity and vice versa. The negative difference values equal to zero.

TABLE 3. TEST DOCUME NT: 10 (DOCUMENT-10 APS = 0.1237)

| No. | Document (Number) | Source Document Ontology-to-Documents APS Score | Semantic Matching | Syntactic Matching |
|---|---|---|---|---|
| | | | Test Document-Source Documents | N-Gram Document-to-Document |
| 1. | 3 | 0.1136 | 0.0101 | 0.0000 |
| 2. | 5 | 0.1084 | 0.0153 | 0.0000 |
| 3. | 8 | 0.1181 | 0.0056 | 0.0000 |
| 4. | 14 | 0.1014 | 0.0223 | 0.0011 |
| 5. | 18 | 0.1357 | Negative Value (0.0000) | 0.0000 |
| 6. | 20 | 0.1142 | 0.0095 | 0.0000 |

Table 4 after we get these results and compare them, we can see that Document-14 has maximum difference value showing that Document-14 is conceptual dissimilar to Document-18 whereas Document-10 has minimum difference values. We know that minimum difference value between two documents show the maximum similarity and vice versa. The negative difference values equal to zero.

TABLE 4. TEST DOCUMENT: 18 (DOCUMENT-18 APS = 0.1357)

| No. | Document (Number) | Source Document Ontology-to-Documents APS Score | Semantic Matching | Syntactic Matching |
|---|---|---|---|---|
| | | | Test Document-Source Documents | N-Gram Document -to-Document |
| 1. | 3 | 0.1136 | 0.0221 | 0.0000 |
| 2. | 5 | 0.1084 | 0.0273 | 0.0000 |
| 3. | 8 | 0.1181 | 0.0176 | 0.0000 |
| 4. | 10 | 0.1237 | 0.0120 | 0.0000 |
| 5. | 14 | 0.1014 | 0.0343 | 0.0011 |
| 6. | 20 | 0.1142 | 0.0215 | 0.0000 |

**Analyzing DOBS (Semantic Similarity) Against N-Gram (Syntactic Similarity) on Partially Relevant Documents:** We have selected 40 documents and map it to ontology. Test document is randomly selected from 40 documents and rests are the source documents. Here are the results:

Table 5 after we get these results and compare them, we can see that Document-15 has maximum difference value showing that Document-15 is conceptual dissimilar to Document-01 whereas Document-26 has minimum difference values. We know that minimum difference value between two documents show the maximum similarity and vice versa. The negative difference values equal to zero.

TABLE 5. TEST DOCUMENT: 01 (DOCUMENT-01 APS=0.0506)

| No. | Document (Number) | Source Document Ontology-to-Documents APS Score | Semantic Matching | Syntactic Matching |
|---|---|---|---|---|
| | | | Test Document-Source Documents | N-Gram Document -to-Document |
| 1. | 4 | 0.0343 | 0.0163 | 0.0000 |
| 2. | 9 | 0.0368 | 0.0138 | 0.0000 |
| 3. | 13 | 0.0312 | 0.0194 | 0.0007 |
| 4. | 15 | 0.0298 | 0.0208 | 0.0000 |
| 5. | 18 | 0.0361 | 0.0145 | 0.0000 |
| 6. | 26 | 0.0399 | 0.0107 | 0.0003 |

Table 6 after we get these results and compare them, we can see that Document-15 has maximum difference value showing that Document-15 is conceptual dissimilar to Document-9 whereas Document-1 and Document-26 have minimum difference values. We know that minimum difference value between two documents show the maximum similarity and vice versa. The negative difference values equal to zero.

TABLE 6. TEST DOCUMENT: 09 (DOCUMENT-9 APS=0.0368)

| No. | Document (Number) | Source Document Ontology-to-Documents APS score | Semantic Matching | Syntactic Matching |
|---|---|---|---|---|
| | | | Test Document-Source Documents | N-Gram Document -to-Document |
| 1. | 1 | 0.0506 | Negative Value (0.0000) | 0.0000 |
| 2. | 4 | 0.0343 | 0.0025 | 0.0000 |
| 3. | 13 | 0.0312 | 0.0056 | 0.0007 |
| 4. | 15 | 0.0298 | 0.0070 | 0.0000 |
| 5. | 18 | 0.0361 | 0.0007 | 0.0000 |
| 6. | 26 | 0.0399 | Negative Value (0.0000) | 0.0003 |

Table 7 after we get these results and compare them, we can see that Document-15 has maximum difference value showing that Document-15 is conceptual dissimilar to Document-26 whereas Document-1 and Document-9 have minimum difference values. We know that minimum difference value between two documents show the maximum similarity and vice versa. The negative difference values equal to zero.

TABLE 7. TEST DOCUMENT: 26 (DOCUMENT-26 APS=0.0399)

| No. | Document (Number) | Source Document Ontology-to-Documents APS Score | Semantic Matching | Syntactic Matching |
|---|---|---|---|---|
| | | | Test Document-Source Documents | N-Gram Document-to-Document |
| 1. | 1 | 0.0506 | Negative value (0.0000) | 0.0000 |
| 2. | 4 | 0.0343 | 0.0056 | 0.0000 |
| 3. | 9 | 0.0368 | 0.0031 | 0.0000 |
| 4. | 13 | 0.0312 | 0.0087 | 0.0007 |
| 5. | 15 | 0.0298 | 0.0101 | 0.0000 |
| 6. | 18 | 0.0361 | 0.0361 | 0.0000 |

So we can say that DOBS is a good approach for detecting the conceptual similarity between the documents instead of Simple Textual (Syntactic) Similarity (N-Gram). Above the table analysis looks closely DOBS results effectiveness. So, we can say that documents that's not syntactically same but conceptually they are same.

**Analyzing DOBS (Semantic Similarity) Against N-Gram (Syntactic Similarity) on Non-Relevant Documents:** We have selected 20 documents (Non-relevant to ontology concepts) and done the analysis of DOBS against N-Gram.

Test Document is randomly selected and rests are the source documents. Here are the results.

Table 8 this is the extreme case where DOBS did not detect the conceptual similarity but N-gram detects the somehow syntactic similarity.

TABLE 8. TEST DOCUMENT: 08 (DOC-08 APS=0.0000)

| No. | Document (Number) | Source Document | Semantic Matching | Syntactic Matching |
|---|---|---|---|---|
| | | Ontology-to-Documents APS score | Test Document-Source Documents | N-Gram Document-to-Document |
| 1. | 1 | | | 0.0019 |
| 2. | 2 | | | 0.0002 |
| 3. | 3 | | | |
| 4. | 4 | | | 0.0000 |
| 5. | 5 | | | |
| 6. | 6 | | | |
| 7. | 7 | | | 0.0025 |
| 8. | 9 | | | |
| 9. | 10 | | | 0.0000 |
| 10. | 11 | 0.0000 | 0.0000 | |
| 11. | 12 | | | |
| 12. | 13 | | | 0.0007 |
| 13. | 14 | | | |
| 14. | 15 | | | 0.0000 |
| 15. | 16 | | | |
| 16. | 17 | | | 0.0001 |
| 17. | 18 | | | 0.0000 |
| 18. | 19 | | | |
| 19. | 20 | | | 0.0014 |

## 4. RESULT AND DISSCUSION

The DOBS is a good approach to detect Conceptual text similarity where documents are slightly paraphrased or by adding synonyms to the documents instead of N-Gram (syntactic) similarity. But there is a verse case where N-Gram is a well approach to detecting syntactic plagiarism when we need to test plagiarism on documents near to copy. So, DOBS is therefore good technique for detecting similarities in conceptual text and N-Gram is a good technique for detecting plagiarism near copies.

## 5. CONCLUSION

Concept based text detection is an active area of research. The main purpose of this research is to analysis conceptually similarity with the use of domain ontology. Where degree of domain ontology based similarity of two different documents means both texts are totally dissimilar and somehow similar on concept based.

The main aim of this research is to develop technique for detection of conceptually similarity by using domain ontology. The particular focus was on detecting extrinsic plagiarism when original text has been altered by using synonyms or rephrased. The detection of text's concept matching with extensive rephrasing is still in its most primitive stages and an open test.

## 6. FUTURE WORK

Semantic or concept based similarity in cross lingual to figure proportion of text re-usability between two cross lingual text documents using ontology. Dynamically Ontology Enhancement. Improvement of concept Extraction in NLP.

## ACKNOWLEDGEMENT

Thanks to ALLAH Almighty, Respected Parents and faculty, whose encourage and provide me research data (Ontology for analysis, Dataset), which belongs to the Institute of Computing, Bahauddin Zakariya University, Multan, Pakistan.

## REFERENCES

[1] Khmelev, D.V., and Teahan, W.J., "A Repetition Based Measure for Verification of Text Collections and for Text Categorization", Proceedings of ACM 26th Annual International Conference on Research and Development in Information Retrieval, pp. 104-110, July, 2003.

[2] Clough, P., and Gaizauskas, R., "Corpora and Text Re-Use", Handbook of Corpus Linguistics, pp. 1249-1271, 2009.

[3] Nawab, R.M.A., "Mono-Lingual Paraphrased Text Reuse and Plagiarism Detection", Ph.D. Dissertation, University of Sheffield, 2012.

[4] Ali, A.M.E.T., Abdulla, H.M.D., and Snasel, V., "Overview and Comparison of Plagiarism Detection Tools", Database, Texts, Specification, Objects, pp. 161-172, 2011.

[5] Clough, P., "Old and New Challenges in Automatic Plagiarism Detection", National UK Plagiarism Advisory Service, 2003.

[6] UKEssays.com. (2019), "Search Results", [online] Available at: https://www.ukessays.com/search.php?q=a-survey-of-plagiarism-detection-methods+information-+technology-essay&cat=63&sa= [Accessed 10th July, 2019].

[7] Osman, A.H., Salim, N., and Abuobieda, A., "Survey of Text Plagiarism Detection", Computer Engineering and Applications Journal, Volume 1, No. 1, pp. 37-45, 2012.

[8] Garfield, E., "Citation Indexes for Science: A New Dimension in Documentation through Association of Ideas", International Journal of Epidemiology, Volume 35, No. 5, pp. 1123-1127, 2006.

[9] En.wikipedia.org. (2019), "Plagiarism Detection", [online] Available at: http://en.wikipedia.org/wiki/Plagiarism_detection [Accessed 3rd September, 2019].

[10] Chow, T.W., and Rahman, M.K.M., "Multilayer SOM with Tree-Structured Data for Efficient Document Retrieval and Plagiarism Detection", IEEE Transactions on Neural Networks, Volume 20, No. 9, pp. 1385-1402, 2009.

[11] Alzahrani, S.M., Salim, N., and Abraham, A., "Understanding Plagiarism Linguistic Patterns, Textual Features, and Detection Methods", IEEE Transactions on Systems, Man, and Cybernetics, Part-C, Applications and Reviews, Volume 42, No. 2, pp. 133-149, 2011.

[12] Hussain, S., "Semantic Based Text Reused Detection and Analysis in Higher Education", Master Thesis, Department of Computer Science, Bahauddin Zakariya University, Multan, Pakistan, 2018.

[13] Mihoubi, H., Simonet, A., and Simonet, M., "An Ontology Driven Approach to Ontology Translation", International Conference on Database and Expert Systems Applications, pp. 573-582, Springer, Berlin, Heidelberg, September, 2000.

[14] Gruber, T.R., "Toward Principles for the Design of Ontologies Used for Knowledge Sharing?",International Journal of Human Computer Studies, Volume 43, Nos. 5-6, pp. 907-928, 1995.

[15] Abdelaziz, K., Ouchani, S., and Chohra, C., "Recommendations-Based on Semantic Analysis of Social Networks in Learning Environments", Computers in Human Behavior, Volume 101, pp. 435-449, 2019.

[16] Xiaonan, J., Ritter, A., and Yen, P.-Y., "Using Ontology-Based Semantic Similarity to Facilitate the Article Screening Process for Systematic Reviews", Journal of Biomedical Informatics, Volume 69, pp. 33-42, 2017.

[17] Citeseerx.ist.psu.edu. (2019), "CiteSeerX-Unknown File Type", [online] Available at:http://citeseerx.ist.psu.edu/viewdoc/ download?doi=10.1.1.294.9744&rep=rep1&type=pdf\ [Accessed 20th November, 2019].

[18] Israr, H., Rao, M.A.N., Arbab, A., Jamshed, H., Riaz, S., and Munir, E.U., "Cross-Language Urdu–English (Clue) Text Alignment Corpus", Working Notes Papers of the CLEF, 2015.