TERMS INTERRELATIONSHIP QUERY EXPANSION TO IMPROVE ACCURACY OF QURAN SEARCH

NUHU YUSUF

A thesis submitted in
Fulfilment of the requirement for the award of the
Doctor of Philosophy

Faculty of Computer Science and Information Technology
Universiti Tun Hussein Onn Malaysia

JUNE 2021

For my beloved mother and father

ACKNOWLEDGEMENT

ABSTRACT

Quran retrieval system is becoming an instrument for users to search for needed information. The search engine is one of the most popular search engines that successfully implemented for searching relevant verses queries. However, a major challenge to the Quran search engine is word ambiguities, specifically lexical ambiguities. With the advent of query expansion techniques for Quran retrieval systems, the performance of the Quran retrieval system has problem and issue in terms of retrieving users needed information. The results of the current semantic techniques still lack precision values without considering several semantic dictionaries. Therefore, this study proposes a stemmed terms interrelationship query expansion approach to improve Quran search results. More specifically, related terms were collected from different semantic dictionaries and then utilize to get roots of words using a stemming algorithm. To assess the performance of the stemmed terms interrelationship query expansion, experiments were conducted using eight Quran datasets from the Tanzil website. Overall, the results indicate that the stemmed terms interrelationship query expansion is superior to unstemmed terms interrelationship query expansion in Mean Average Precision with Yusuf Ali 68%, Sarawar 67%, Arberry 72%, Malay 65%, Hausa 62%, Urdu 62%, Modern Arabic 60% and Classical Arabic 59%.

ABSTRAK

Sistem capaian Al-Quran menjadi alat untuk pengguna mencari maklumat yang diperlukan. Mesin pencari adalah salah satu enjin carian yang paling popular yang berjaya dilaksanakan untuk mencari pertanyaan ayat yang relevan. Namun, cabaran utama mesin pencari Al-Quran adalah kesamaran kata, khususnya kesamaran leksikal. Dengan munculnya teknik pengembangan pertanyaan untuk sistem capaian Al-Quran, prestasi sistem pengambilan Al-Quran mempunyai masalah dan masalah dari segi mendapatkan maklumat yang diperlukan pengguna. Hasil teknik semantik semasa masih kekurangan nilai ketepatan tanpa mempertimbangkan beberapa kamus semantik. Oleh itu, kajian ini mencadangkan pendekatan pengembangan pertanyaan melalui hubungan antara istilah yang berdasar untuk meningkatkan hasil carian Al-Quran. Lebih khusus lagi, hubungan istilah dikumpulkan dari kamus semantik yang berlainan dan kemudian menggunakannya untuk mendapatkan akar kata tanpa imbuhan. Untuk menilai prestasi cadangan permintaan hubungan antara istilah yang dicetuskan, percubaan dilakukan menggunakan lapan set al-Quran dari laman web Tanzil. Secara keseluruhan, hasilnya menunjukkan bahawa pengembangan pertanyaan hubungan antara istilah tanpa imbuhan adalah lebih unggul daripada pengembangan pertanyaan istilah antara hubungan yang berimbuhan dalam Purata bagi Ketepatan Purata dengan Yusuf Ali 68%, Sarawar 67%, Arberry 72%, Bahasa Melayu 65%, Hausa 62%, Urdu 62% , Bahasa Arab Moden 60% dan Bahasa Arab Klasik 59%.

CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ALGORITHMS

LIST OF SYMBOLS AND ABBREVIATIONS

AP	-	Average Precision
AR	-	Average Recall
TS	-	Term Selection
WASS	-	Word Ambiguities Semantic Search
MAP	-	Mean Average Precision
BEP	-	Break Even Point
MRR	-	Mean Reciprocal Rank
BM25	-	Best Matching
TFID	-	Term Frequency Inverse Document Frequency
VSM	-	Vector Space Model
LSI	-	Latent Semantic Indexing
PLSI	-	Probabilistic Latent Semantic Indexing
Word2Vec	-	Word to Vector
ALQS	-	Apache Lucene Quran Search
P@K	-	Precision at K
F-Measure	-	F1 Measure
NDCG	-	Normalized Discounted Cumulative Gain
CBOW	-	Continuous Bag of Word
FAQ	-	Frequency Ask Question
SAS	-	Structure Activity Relationship
PC	-	Pearson Correlation
RMSE	-	Root Mean Squared Error
R@K	-	Recall at Rank K
DCG	-	Discounted Cumulative Gain
RDF	-	Resource Description Framework

OWL	-	Ontology Web Language
LSI-VSM	-	Latent Semantic Indexing-Vector Space Model
PLSI-VSM	-	Probabilistic Latent Semantic Indexing-Vector Space Model
Word2Vec	-	Word to Vector
SVD	-	Singular Value Decomposition
LDA-VSM	-	Latent Dirichlet Allocation
RI-VSM	-	Random Indexing-Vector Space Model
EM	-	Expectation-maximization Algorithm
WSD	-	Word Sense Disambiguation
JDK	-	Java Development Kit
IDE	-	Integrated Development Environment
TSV	-	Term Selection Value
TI-TSV		Terms Interrelationship - Term Selection Value
slTI-TSV	-	Statistical Linguistic Terms Interrelationship Term Selection Value
WslTI-TSV	-	Weighting Statistical Linguistic Terms Interrelationship Term Selection Value
TIQE	-	Terms Interrelationship Query Expansion

LIST OF APPENDICES

LIST OF PUBLICATIONS

JOURNAL:

I. **Nuhu Yusuf,** Mohd Amin Mohd Yunus, Norfaradilla Wahid. (2019) "A Comparative Analysis of Web Search Query: Informational Vs Navigational Queries". International Journal on Advanced Science, Engineering and Information Technology, 9(1),(**Published, Scopus Index**)

II. **Nuhu Yusuf,** Mohd Amin Mohd Yunus, Norfaradilla Wahid. (2019) "Query Expansion based on Explicit-Relevant Feedback and Synonyms for English Quran Translation Information Retrieval". International Journal of Advanced Computer Science and Applications, 10(5),(**Published, ISI and Scopus Index**)

III. **Nuhu Yusuf,** Mohd Amin Mohd Yunus, Norfaradilla Wahid. (2019) "Query Expansion Method for Quran Search Using Semantic Search and Lucene Ranking". Journal of Engineering Science and Technology (JESTEC), 15(1),675-692 (**Published, ISI and Scopus Index**)

IV. **Nuhu Yusuf,** Mohd Amin Mohd Yunus, Norfaradilla Wahid, Aida Mustapha, Nazri Mohd Nawi and Noor Azah Samsudin. (2020) "Arabic Text Semantic-Based Query Expansion". International Journal of Data Mining, Modelling and Management (IJDMMM), (In Press). (**Accepted, ISI and Scopus Index**)

CONFERENCE PROCEEDINGS:

I. **Nuhu Yusuf,** Mohd Amin Mohd Yunus and Norfaradilla Wahid. (2019) "Arabic Text Stemming Using Query Expansion". The 4th International Conference of Reliable Information and Communication Technology

2019 (IRICT 2019). Advances in Intelligent Systems and Computing (**Publication, ISI and Scopus Index**)

II.   **Nuhu Yusuf,** Mohd Amin Mohd Yunus and Norfaradilla Wahid. (2019) "Enhancing Query Expansion Method Using Word Embedding". The 2019 IEEE International Conference on System Engineering and Technology (ICSET 2019). IEEE (**Published, IEEEXplore and Scopus Index**)

III.   **Nuhu Yusuf,** Mohd Amin Mohd Yunus, Norfaradilla Wahid, Mohd Najib Mohd Saleh and Aida Mustapha (2020) "A Terms Interrelationship to Query Expansion Approach Based on Term Selection". The 5th International Conference of Reliable Information and Communication Technology 2019 (IRICT 2020). Advances in Intelligent Systems and Computing (**Accepted, ISI and Scopus Index**)

<div align="center">

CHAPTER 1

INTRODUCTION

</div>

1.1 Background Study

The increase in demand for search engine has become a worldwide phenomenon.
Today, many countries of the world are witnessing the growth of search engine users.
These users use the search engines not only for locating web pages but also for
research purposes such as finding research articles and papers. Along with search
engine users, there is also a drastic growth in the number of search engines [1]. This
led to improvements of such search engines to produce relevant search results to
users and minimize irrelevant ones. The effect of this is that users' information needs
largely depend on the search engines precision. The search engine is an information
retrieval application area that plays significant roles in ensuring relevant information
is obtained from available collections. In particular, the query is used to search
available or needed information within a document. Many search engines are now
available ranging from simplest to more complex ones. For instance, desktop search
[2], personal search [3], emails search [4], enterprise search [5] and mobile search
[6] [7] are an important search engine that is currently used by many users (personal,
organizations and government). The users use these search engines to search for
relevant information. Web Search has been the most successful search engine that
has been used in different applications to search for information. Such information
can come from different sources, especially web pages. This is however not
satisfactory because of the inability of a search engine to understand the query and
bring out needed information [1]. For a web search engine to display the relevant
documents based on users need, additional terms must be given to a query to

properly understand the meaning behind that query. For instance, documents that match a query with additional terms can be more likely to be relevant to the user.

Query expansion is one of the techniques used to improve search engine performance by adding additional information to a user query. Recently, query expansions have attracted the attention of many researchers and practitioners. This led to an increasing number of query expansion techniques and algorithms. Much of these query expansion techniques have been examined in [8] and [9] to add additional terms to the query. For example, the study [8] compared and contrast different query expansion techniques in terms of their datasets and query processing algorithms. It presents a survey of about 57 years' periods. In addition to that, the study [9] provides useful information on how to query expansion techniques based on word sense disambiguation, fuzzy rules and term weighting. Zhou et al [10] concentrated on how additional terms can be added to users profiles for better query expansion. The query and its weight can be used to measure the relevance of a document. A similar paper [11] takes account of user profiles to get relevant information. According to Liu et al [12], word embedding can be used to generate a list of words so that word similarities can be chosen. Other studies considered either Wikipedia [13] or query log [14] to improve search effectiveness. However, the query expansion still suffered a vocabulary mismatch [15], specifically in terms of word ambiguities of term selection [16] and term weighting [17] which lead to a lack of better search quality results. Consequently, linguistics and ontology methods have emerged among others as approaches of query expansion that concerned to improve performance.

The holy Quran is the holy book of guidance where laws, commandments, codes for their social and moral behaviors are laid down to Muslims. It contains a comprehensive religious philosophy. The language of the Holy Quran is Arabic but was translated into various languages all over the world. The holy Quran is divided into 114 surahs or chapters and each chapter consists of individual ayaat or verses. There are in total 6326 verses in the Holy Quran. The text of the Holy Quran has remained unchanged over the past 1440 years. The emerging Quran search on a Quran search system can be used to enhance a variety of task, such as searching for a specific verse. It can also enable the discovery of related verses of different keywords by allowing expanding query with different terms. A query can be expanded with synonyms of terms based on terms relationship. Afzal and Muktar

[18] suggested a Quran English WordNet solution as a linguistic approach to short queries, especially when linking to semantic similarities. Their work achieved significant improvement. Although semantic similarity proves effective, Moawad, Alromima and Elgohary [19] stated that absences of semantic resources were identified in many languages and as such alternative semantic approaches need to be developed. Bentrcia, Zidat and Marir [20] examined the possibility of using semantic relatedness instead of only semantic similarities. This paper used only semantic relatedness to obtain a relationship between two concepts using "AND" conjunctive. However, Lashkari, Bagheri and Ghorbani [21] identified based on the literature that depending only on semantic resource would not improve better search performance. Integrations of other methods with semantic were suggested to yield better results. This thesis utilizes ontology method-based query expansion to expand the query. An ontology method-based query expansion is the main tool for selecting query expansion term from knowledgebase based on their relatedness to the query [22].

Currently, the term selection assumptions are mainly concentrated on the probability of word occurrences. For instance, Jagendra et al [16] approach mainly assumed the chance of word occurrence. The term selection depends closely on how words relate to each other. Zhou et al [10] proposed a personalized query expansion. However, query expansion approaches are significantly improving the search engine performance in terms of relevance feedback, synonyms, ontology or related technique like semantic mapping. The combination of expert judgment, word synonyms and stemming require researchers' attention. Although some researchers such as Wasim et al. [23] try to address word ambiguities in term selection challenges by utilizing relevant assessment. Still, the basis of the work combines relevant feedback and word synonyms on biomedical document retrieval. Singh and Sharan [24] presented different term selection methods using relevant feedback. They combined the relevance feedback with a rank combination to improve search performance. This thesis is concerned with the Terms Interrelationship (TI) based on semantic search. Terms Interrelationship can effectively express different term relationships and improvement of query expansion.

1.2   Research Motivation

Terms interrelationship (TI) in semantic search can be used in query expansion through adding word meanings. Using TI semantic search in query expansion has been a topic of research from both researchers and practitioners. However, most of the studies such as Singh and Sharan [24] are focused on the pseudo-relevance feedback technique, which demands selecting terms for expansion from top retrieved documents. In the absence of such term selection methods, TI semantic search is used with synonyms [25]. Holy Quran is the holy book revealed by Allah to serve as guidance to Muslims. In the Quran search system; there was a need for a query expansion for expanding Quran query terms, which will ensure retrieving relevant verses (ayats) and chapters (surahs) of the holy Quran. The goal is to expand relevant query term with the similar term meaning to retrieve relevant information need. It does not necessarily consider the query and document terms to be the same because terms may be different but contained the same meanings. In Quran search query expansion, adding the semantic meaning of a term is interesting because Quran search always involves verse meaning to have term selection. For instance, Quran 7:31 said, "O children of Adam, take your adornment at every Masjid, and eat and drink, but be not excessive. Indeed, He likes not those who commit excess". The word "Masjid" is ambiguous because it may be a time or place of prayers (Yusuf Ali), place of worship (Pickthal) or Kabah at the time of Tawaf. Thus, adding the meaning of a term in query expansion might improve the query expansion results. Therefore, word ambiguities need to be addressed to properly understanding this query. Because of word ambiguities present in a query, commonly used technique, such as pseudo-relevance feedbacks are not appropriate for the Quran search system. Therefore, the first contribution of this thesis deals with the problem of providing term selection for query expansion with word ambiguities. The thesis proposed TI in semantic search solution for this problem.

Besides the presence of word ambiguities, a challenging aspect of term selection in the Quran search system is that the system-retrieved results may contain many irrelevant documents. In other words, the system retrieved results not only about term selection but also relevant to the query. For instance, a system may contain different term selection methods in query expansion but still contain many irrelevant results. Such situations can arise not only in the Quran search system but

also in also other systems. For a similar domain, Moaward et al [19] has proposed a bi-gram method of query expansion, where query terms are expanded with n-gram for effective results. They take into account improving relevant search results but not semantic search. Moreover, such improvement requires a better term selection method for word ambiguities in query expansion. This thesis identifies improving query expansion for TI in semantic search.

1.3   Problem Statement

Term selection is a challenge to query expansion, which refers to selecting terms that are similar to the query terms. The idea is that more documents that are relevant can be retrieved through query expansion if proper terms are selected similar to the query terms. Addressing the term selection required addressing word ambiguities [4]. Lexical ambiguity is more commonly addressed by researchers [5] [4] that concern on word ambiguities in a query which still led to many irrelevant documents retrieved. In addition, there are many different word ambiguities in a query. Hence, there is a need to address lexical ambiguities that are commonly found in Quran search to minimize irrelevant documents retrieval. Terms relationship in natural language understanding is addressing word ambiguities for semantic search that employs one semantic dictionary or WordNet. However, fewer concentrations concern about the terms relationship from more than one WordNet. Having stated the challenges of term selection in terms of lexical ambiguities, the researchers are now able to formulate the main problem in query expansion as:

> *How can the lexical ambiguities of term selection are exploited to produce the best possible query expansion for Quran system*

As a semantic search method, term relationship was used to capture different Quran terms of lexical ambiguities that could guide term selection in query expansion. The thesis presents meaningful definitions of relationships that apply to lexical ambiguities that may arise when selecting a term for query expansion. Terms interrelationship algorithms of lexical ambiguity will be designed to provide accurate query expansion and evaluated in Mean Average Precision (MAP), Mean Reciprocal

REFERENCES

[1]   G. Chandra and S. K. Dwivedi, "Query Expansion for Effective Retrieval Results of Hindi–English Cross-Lingual IR," *Appl. Artif. Intell.*, vol. 0, no. 0, pp. 1–27, 2019, doi: 10.1080/08839514.2019.1577018.

[2]   M. S. Raje, "Analysis of Desktop Search and Ranking of Their Results Based on Semantics from User Feedback," *2016 12th Int. Conf. Signal-Image Technol. Internet-Based Syst.*, pp. 241–245, 2016, doi: 10.1109/SITIS.2016.46.

[3]   X. Wang, N. Golbandi, M. Bendersky, D. Metzler, and M. Najork, "Position Bias Estimation for Unbiased Learning to Rank in Personal Search," 2018, doi: 10.1145/3159652.3159732.

[4]   Q. Ai, S. T. Dumais, N. Craswell, and D. J. Liebling, "Characterizing Email Search using Large-scale Behavioral Logs and Surveys.," in *Proceedings of the 26th International Conference on World Wide Web*, 2017, pp. 1511–1520.

[5]   U. Kruschwitz and C. Hull, "Searching the Enterprise," *Found. Trends® Inf. Retr.*, vol. 11, no. 1, pp. 1–142, 2017, doi: 10.1561/1500000053.

[6]   C. Luo, Y. Liu, T. Sakai, F. Zhang, M. Zhang, and S. Ma, "Evaluating Mobile Search with Height-Biased Gain," *Proc. 40th Int. ACM SIGIR Conf. Res. Dev. Inf. Retr. - SIGIR '17*, pp. 435–444, 2017, doi: 10.1145/3077136.3080795.

[7]   J. Mao, Y. Liu, N. Kando, Z. He, and M. Zhang, "A Two-Stage Model for User ' s Examination Behavior in Mobile Search," *Assoc. Comput. Mach.*, pp. 273–276, 2018.

[8]   H. K. Azad and A. Deepak, "Query expansion techniques for information retrieval: A survey," *Inf. Process. Manag.*, vol. 56, no. 5, pp. 1698–1735, 2019, doi: 10.1016/j.ipm.2019.05.009.

[9]   N. Kathuria, "A Comprehensive Survey on Query Expansion Techniques , their Issues and Challenges," *Int. J. Comput. Appl.*, vol. 168, no. 12, pp. 17–20, 2017.

[10]   D. Zhou, X. Wu, W. Zhao, S. Lawless, and J. Liu, "Query Expansion with Enriched User Profiles for Personalized Search Utilizing Folksonomy Data," *IEEE Trans. Knowl. Data Eng.*, vol. 29, no. 7, pp. 1536–1548, 2017.

[11]   S. Patel, N. Bhatt, and C. Shah, "Query expansion for effective retrieval from microblog," in *2017 International Conference on Computing Methodologies and Communication (ICCMC)*, Jul. 2017, no. ICCMC, pp. 394–397, doi: 10.1109/ICCMC.2017.8282716.

[12]   Q. Liu, H. Huang, J. Lut, Y. Gao, and G. Zhang, "Enhanced word embedding similarity measures using fuzzy rules for query expansion," in *2017 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, Jul. 2017, vol. L, pp. 1–6.

[13]   D. Puspitaningrum, G. Yulianti, and I. S. W. B. Prasetya, "Wiki-MetaSemantik: A Wikipedia-derived query expansion approach based on network properties," in *2017 5th International Conference on Cyber and IT Service Management, CITSM 2017*, 2017, pp. 1–6, doi: 10.1109/CITSM.2017.8089228.

[14]   A. Satter and K. Sakib, "A search log mining based query expansion technique to improve effectiveness in code search," in *19th International Conference on Computer and Information Technology, ICCIT 2016*, 2017, pp. 586–591, doi: 10.1109/ICCITECHN.2016.7860264.

[15]   L. Zhao and J. Callan, "Automatic term mismatch diagnosis for selective query expansion," *SIGIR'12 - Proc. Int. ACM SIGIR Conf. Res. Dev. Inf. Retr.*, pp. 515–524, 2012, doi: 10.1145/2348283.2348354.

[16]   J. Singh *et al.*, "Fuzzy Logic Hybrid Model with Semantic Filtering Approach for Pseudo Relevance Feedback-based Query Expansion," in *Computational Intelligence (SSCI), 2017 IEEE Symposium Series*, 2017, pp. 1–7.

[17]   M. Santhanakumar, C. C. Columbus, and K. Jayapriya, "Multi term based co-term frequency method for term weighting in information retrieval," *Int. J. Bus. Inf. Syst.*, vol. 28, no. 1, p. 79, 2018, doi: 10.1504/ijbis.2018.10012193.

[18]   H. Afzal and T. Mukhtar, "Semantically Enhanced Concept Search of the Holy Quran: Qur'anic English WordNet," *Arab. J. Sci. Eng.*, vol. 44, no. 4, pp. 3953–3966, 2019, doi: 10.1007/s13369-018-03709-2.

[19]   I. Moawad, W. Alromima, and R. Elgohary, "Bi-Gram Term Collocations-based Query Expansion Approach for Improving Arabic Information

Retrieval," *Arab. J. Sci. Eng.*, vol. 43, no. 12, pp. 7705–7718, 2018, doi: 10.1007/s13369-018-3145-y.

[20] R. Bentrcia, S. Zidat, and F. Marir, "Extracting Semantic Relations from the holy Quran Based on Arabic Conjunctive Patterns," *J. King Saud Univ. - Comput. Inf. Sci.*, vol. 30, no. 3, pp. 382–390, 2018, doi: 10.1016/10.1016/j.jksuci.2017.09.004.

[21] F. Lashkari, E. Bagheri, and A. A. Ghorbani, "Neural embedding-based indices for semantic search," *Inf. Process. Manag.*, vol. 56, no. 3, pp. 733–755, 2019, doi: 10.1016/j.ipm.2018.10.015.

[22] J. A. Nasir, I. Varlamis, and S. Ishfaq, "A knowledge-based semantic framework for query expansion," *Inf. Process. Manag.*, vol. 56, no. 5, pp. 1605–1617, 2019, doi: 10.1016/j.ipm.2019.04.007.

[23] M. Wasim, M. N. Asim, M. U. Ghani, Z. U. Rehman, S. Rho, and I. Mehmood, "Lexical paraphrasing and pseudo relevance feedback for biomedical document retrieval," *Multimed. Tools Appl.*, vol. 78, pp. 29681–29712, 2019, doi: 0.1007/s11042-018-6060-z.

[24] J. Singh and A. Sharan, "Rank fusion and semantic genetic notion based automatic query expansion model," *Swarm Evol. Comput.*, vol. 38, no. September 2017, pp. 295–308, 2018, doi: 10.1016/j.swevo.2017.09.007.

[25] M. Franco-Salvador and L. A. Leiva, "Multilingual phrase sampling for text entry evaluations," *Int. J. Hum. Comput. Stud.*, vol. 113, no. January, pp. 15–31, 2018, doi: 10.1016/j.ijhcs.2018.01.006.

[26] Z.-Z. Hamid, "Quran Translations," *Tanzil Documents*, 2007. tanzil.net/trans/ (accessed Jul. 29, 2019).

[27] D. Chefrour and A. Amirat, "Using Fast String Search for Quran Text Auto-completion on Android," in *International Conference on Computer Science and its Applications*, 2019, vol. 50, pp. 92–101.

[28] M. Afifi Mohamad Safee *et al.*, "Hybrid Search Approach for Retrieving Medical and Health Science Knowledge from Quran," *Int. J. Eng. Technol.*, vol. 7, no. 4.15, p. 69, 2018.

[29] H. K. Azad and A. Deepak, "Query expansion techniques for information retrieval: A survey," *Inf. Process. Manag.*, vol. 56, no. 5, pp. 1698–1735, Sep. 2019.

[30] Y. M. Alginahi, "Quran search engines: challenges and design requirements,"

Int. J. Comput. Appl. Technol., vol. 57, no. 3, p. 237, 2018.

[31] J. Goikoetxea, A. Soroa, and E. Agirre, "Bilingual Embeddings with Random Walks over Multilingual Wordnets," *Knowledge-Based Syst.*, vol. 150, no. 15, pp. 218–230, 2018.

[32] B. S. Pedersen, M. Agirrezabal, S. Nimb, S. Olsen, and I. Rørmann, "Towards a principled approach to sense clustering – a case study of wordnet and dictionary senses in Danish," in *Proceedings of Global WordNet Conference*, 2018, pp. 1–8.

[33] A. R. Pal and D. Saha, "Word sense disambiguation in Bengali: An unsupervised approach," in *Proceedings of the 2017 2nd IEEE International Conference on Electrical, Computer and Communication Technologies, ICECCT 2017*, 2017, pp. 1–11.

[34] S. R. Kolhe and S. D. Sawarkar, "A concept driven document clustering using WordNet," in *2017 International Conference on Nascent Technologies in Engineering (ICNTE)*, 2017, pp. 1–5, doi: 10.1109/ICNTE.2017.7947888.

[35] A. Saif, N. Omar, U. Z. Zainodin, and M. J. Ab Aziz, "Building Sense Tagged Corpus Using Wikipedia for Supervised Word Sense Disambiguation," in *Procedia Computer Science*, 2018, vol. 123, pp. 403–412.

[36] M. Liu, W. Pan, M. Liu, Y. Chen, X. Peng, and Z. Ming, "Mixed similarity learning for recommendation with implicit feedback," *Knowledge-Based Syst.*, vol. 119, pp. 178–185, 2017.

[37] K. Albishre, Y. Li, and Y. Xu, "Effective pseudo-relevance for Microblog retrieval," in *Proceedings of the Australasian Computer Science Week Multiconference on - ACSW '17*, 2017, pp. 1–6.

[38] S. Jiang, T. F. Hagelien, M. Natvig, and J. Li, "Ontology-Based Semantic Search for Open Government Data," *2019 IEEE 13th Int. Conf. Semant. Comput.*, pp. 7–15, 2019, doi: 10.1109/icosc.2019.8665522.

[39] M. Esposito, E. Damiano, A. Minutolo, G. De Pietro, and H. Fujita, "Hybrid query expansion using lexical resources and word embeddings for sentence retrieval in question answering," *Inf. Sci. (Ny).*, vol. 514, pp. 88–105, 2020, doi: 10.1016/j.ins.2019.12.002.

[40] A. M. Hasan, A. M., Noor, N. M., Rassem, T. H., Noah, S. A. M., & Hasan, "A Proposed Method Using the Semantic Similarity of WordNet 3.1 to Handle the Ambiguity to Apply in Social Media Text," in *Information Science and*

Applications, K. J. Kim and H.-Y. Kim, Eds. Springer, Singapore., 2020, pp. 471–483.

[41] M. A. Zingla, C. Latiri, P. Mulhem, C. Berrut, and Y. Slimani, "Hybrid query expansion model for text and microblog information retrieval," *Inf. Retr. J.*, vol. 21, no. 4, pp. 337–367, 2018, doi: 10.1007/s10791-017-9326-6.

[42] A. El Mahdaouy, S. O. El Alaoui, and E. Gaussier, "Word-embedding-based pseudo-relevance feedback for Arabic information retrieval," *J. Inf. Sci.*, vol. 45, no. 4, pp. 429–442, 2019, doi: 10.1177/0165551518792210.

[43] H. K. Azad and A. Deepak, "A new approach for query expansion using Wikipedia and WordNet," *Inf. Sci. (Ny).*, vol. 492, pp. 147–163, 2019, doi: 10.1016/j.ins.2019.04.019.

[44] D. Lucaci and A. Iftene, "Query expansion-automatic generation of semantic similar phrases using WordNet," 2018.

[45] C. Lucchese, F. M. Nardini, and R. Trani, "Efficient and Effective Query Expansion for Web Search," in *The 27th ACM International Conference on Information and Knowledge Management (CIKM)*, 2018, pp. 1551–1554, doi: 10.1145/3269206.3269305.

[46] G. Chandra and S. K. Dwivedi, "Query expansion based on term selection for Hindi - English cross lingual IR," *J. King Saud Univ. - Comput. Inf. Sci.*, vol. 32, no. 3, pp. 310–319, 2020.

[47] F. Beirade, H. Azzoune, and D. E. Zegour, "Semantic query for Quranic ontology," *J. King Saud Univ. - Comput. Inf. Sci.*, no. xxxx, pp. 1–9, 2019, doi: 10.1016/j.jksuci.2019.04.005.

[48] F. Zhang, H. Niu, I. Keivanloo, and Y. Zou, "Expanding Queries for Code Search Using Semantically Related API Class-names," *IEEE Trans. Softw. Eng.*, vol. 44, no. 11, pp. 1070–1082, 2017, doi: 10.1109/TSE.2017.2750682.

[49] Y. Gupta and A. Saini, "A novel Fuzzy-PSO term weighting automatic query expansion approach using combined semantic filtering," *Knowledge-Based Syst.*, vol. 136, pp. 97–120, 2017, doi: 10.1016/j.knosys.2017.09.004.

[50] A. M. Al-Zoghby, A. Elshiwi, and A. Atwan, "Semantic relations extraction and ontology learning from arabic texts—A survey," *Stud. Comput. Intell.*, vol. 740, pp. 199–225, 2018, doi: 10.1007/978-3-319-67056-0_11.

[51] B. Al-Khateeb, A. J. Al-Kubaisi, and S. T. Al-Janabi, "Query Reformulation Using WordNet and Genetic Algorithm," no. March, pp. 7–9, 2017, doi:

10.1109/NTICT.2017.7976138.

[52]   M. A. Yunus, R. Zainuddin, and N. Abdullah, "Semantic query with stemmer for quran documents results," *ICOS 2010 - 2010 IEEE Conf. Open Syst.*, no. Icos, pp. 40–44, 2010, doi: 10.1109/ICOS.2010.5720061.

[53]   J. Singh and A. Sharan, "Co-occurrence and Semantic Similarity Based Hybrid Approach for Improving Automatic," in *International Conference on Distributed Computing and Internet Technology, ICDCIT 2015: Distributed Computing and Internet Technology*, 2015, pp. 415–418.

[54]   F. Gonçalves, T. Oliveira, J. Neves, and P. Novais, "Semantic Capture Analysis in Word Embedding Vectors Using Convolutional Neural Network," *Adv. Intell. Syst. Comput.*, vol. 569, pp. 257–266, 2017, doi: 10.1007/978-3-319-56535-4.

[55]   A. Krishnan, P. Deepak, S. Ranu, and S. Mehta, "Leveraging semantic resources in diversified query expansion," *World Wide Web*, vol. 21, no. 4, pp. 1041–1067, 2018, doi: 10.1007/s11280-017-0468-7.

[56]   C. T. Zheng, C. Liu, and H. S. Wong, "Corpus-based topic diffusion for short text clustering," *Neurocomputing*, vol. 275, pp. 2444–2458, 2018, doi: 10.1016/j.neucom.2017.11.019.

[57]   C. Bhargavi and A. B. Reddy, "Transforming Unstructured Data into Conceptual Representation Using WORDNET," vol. 3, no. 1, pp. 389–395, 2018.

[58]   S. Zhang and G. H. Yang, "Deriving Differentially Private Session Logs for Query Suggestion," in *the 3rd ACM International Conference on the Theory of Information Retrieval, ICTIR*, 2017, vol. 17, pp. 51–58.

[59]   F. Shoeleh, M. S. Zahedi, and M. Farhoodi, "Search Engine Pictures : Empirical Analysis of a Web Search Engine Query Log," in *Web Research (ICWR), 2017 3th International Conference*, 2017, pp. 90–95.

[60]   Y. Jiang, W. Bai, X. Zhang, and J. Hu, "Wikipedia-based information content and semantic similarity computation," *Inf. Process. Manag.*, vol. 53, no. 1, pp. 248–265, 2017, doi: 10.1016/j.ipm.2016.09.001.

[61]   H. Holzmann, W. Nejdl, and A. Anand, "Exploring Web Archives Through Temporal Anchor Texts ∗," in *9th International ACM Web Science Conference*, 2017, pp. 1–10, doi: 10.1145/3091478.3091500.

[62]   A. Otsuka, K. Nishida, K. Bessho, H. Asano, and J. Tomita, "Query

Expansion with Neural Question-to-Answer Translation for FAQ-based Question Answering," in *Track: First International Workshop on Hybrid Question Answering with Structured and Unstructured Knowledge (HQA'18*, 2018, pp. 1063–1068, doi: 10.1145/3184558.3191537.

[63]   N. Yusuf, M. Amin, M. Yunus, and N. Wahid, "Query Expansion based on Explicit-Relevant Feedback and Synonyms for English Quran Translation Information Retrieval," *Int. J. Adv. Comput. Sci. Appl.*, vol. 10, no. 5, pp. 227–234, 2019.

[64]   J. Singh and A. Sharan, "A new fuzzy logic-based query expansion model for efficient information retrieval using relevance feedback approach," *Neural Comput. Appl.*, vol. 28, no. 9, pp. 2557–2580, 2017, doi: 10.1007/s00521-016-2207-x.

[65]   C. Layfield, J. Azzopardi, and C. Staff, "Experiments with Document Retrieval from Small Text Collections Using Latent Semantic Analysis or Term Similarity with Query Coordination and Automatic Relevance Feedback," in *Calì A., Gorgan D., Ugarte M. (eds) Semantic Keyword-Based Search on Structured Data Sources. IKC 2016. Lecture Notes in Computer Science*, 2017, vol. 10546, pp. 25–36.

[66]   D. Roy, D. Paul, M. Mitra, and U. Garain, *Using Word Embeddings for Automatic Query Expansion*. NY 14850, USA: Cornell University Library, 2016.

[67]   R. Kumar, G. Bhanodai, and R. Pamula, "Book search using social information, user profiles and query expansion with Pseudo Relevance Feedback," *Appl. Intell.*, pp. 1–23, Jan. 2019, doi: 10.1007/s10489-018-1383-z.

[68]   A. Montazeralghaem, H. Zamani, and A. Shakery, "Term Proximity Constraints for Pseudo-Relevance Feedback," in *Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval - SIGIR '17*, 2017, pp. 1085–1088, doi: 10.1145/3077136.3080728.

[69]   A. Keikha, F. Ensan, and E. Bagheri, "Query expansion using pseudo relevance feedback on wikipedia," *J. Intell. Inf. Syst.*, vol. 50, no. 3, pp. 455–478, 2017, doi: 10.1007/s10844-017-0466-3.

[70]   S. H. Na and K. Kim, "Verbosity normalized pseudo-relevance feedback in

information retrieval," *Inf. Process. Manag.*, vol. 54, no. 2, pp. 219–239, 2018.

[71] Y. Wang, H. Huang, and C. Feng, "Query Expansion Based on a Feedback Concept Model for Microblog Retrieval," in *Proceedings of the 26th International Conference on World Wide Web - WWW '17*, 2017, pp. 559–568.

[72] F. S. Utomo, N. Suryana, and M. S. Azmi, "New instances classification framework on Quran ontology applied to question answering system," *TELKOMNIKA (Telecommunication Comput. Electron. Control.*, vol. 17, no. 1, pp. 139–146, 2019.

[73] Z. Chen, Z. He, X. Liu, and J. Bian, "Evaluating semantic relations in neural word embeddings with biomedical and general domain knowledge bases," *BMC Med. Inform. Decis. Mak.*, vol. 18, no. 2, p. 65, 2018, doi: 10.1186/s12911-018-0630-x.

[74] F. Beirade, H. Azzoune, and D. E. Zegour, "Semantic query for Quranic ontology," *J. King Saud Univ. - Comput. Inf. Sci.*, no. xxxx, pp. 1–9, Apr. 2019.

[75] R. L. Wang, S. Edwards, and C. Ives, "Ontology-based semantic mapping of chemical toxicities," *Toxicology*, vol. 412, pp. 89–100, 2019, doi: 10.1016/j.tox.2018.11.005.

[76] H. Gong, J. Du, and W. Wang, "Microblog query expansion based on ontology expansion and borda count rank," in *Chinese Intelligent Systems Conference*, 2018, vol. 459, pp. 271–280.

[77] D. M. Banisakher *et al.*, "Ontology-based supervised concept learning for the biogeochemical literature," in *Proceedings - 2018 IEEE 19th International Conference on Information Reuse and Integration for Data Science, IRI 2018*, 2018, pp. 402–410, doi: 10.1109/IRI.2018.00066.

[78] S. Sharma, A. Kumar, and V. Rana, "Ontology based informational retrieval system on the semantic web: Semantic web mining," in *Proceedings - 2017 International Conference on Next Generation Computing and Information Systems, ICNGCIS 2017*, 2018, pp. 43–47, doi: 10.1109/ICNGCIS.2017.21.

[79] H. L. Wang, S. H. Wu, and K. K. Wang, "Semantic Search on Internet Tabular Information Extraction for Answering Queries," in *Ninth International Conference on Information and Knowledge Management (CIKM)*, 2000, no.

January, pp. 243–249.

[80] P. Clark, J. Thompson, H. Holmback, and L. Duncan, "Exploiting a Thesaurus-Based Semantic Net for Knowledge-Based Search," in *IAAI 2000 Proceedings*, 2000, pp. 988–995.

[81] F. S. Utomo, N. Suryana, and M. S. Azmi, "New instances classification framework on Quran ontology applied to question answering system," *TELKOMNIKA (Telecommunication Comput. Electron. Control.*, vol. 17, no. 1, pp. 139–146, 2019, doi: 10.12928/telkomnika.v17i1.9794.

[82] P. Yang and J. Lin, "Reproducing and Generalizing Semantic Term Matching in Axiomatic Information Retrieval," in *European Conference on Information Retrieval*, 2019, vol. 68, no. 5, pp. 369–381, doi: 10.1007/978-3-030-15712-8_24.

[83] A. Jain, S. Vij, and O. Castillo, "Hindi Query Expansion based on Semantic Importance of Hindi WordNet Relations and Fuzzy Graph Connectivity Measures," *Comput. y Sist.*, vol. 23, no. 4, pp. 1337–1355, Dec. 2019, doi: 10.13053/cys-23-4-2984.

[84] A. Bouziri, C. Latiri, and E. Gaussier, "LTR-expand: query expansion model based on learning to rank association rules," *J. Intell. Inf. Syst.*, pp. 1–26, Mar. 2020, doi: 10.1007/s10844-020-00596-8.

[85] R. A. Kadir, R. A. Yauri, and A. Azman, "Semantic ambiguous query formulation using statistical Linguistics technique," *Malaysian J. Comput. Sci.*, vol. 31, no. 5, pp. 48–56, 2018.

[86] H. ALMarwi, M. Ghurab, and I. Al-Baltah, "A hybrid semantic query expansion approach for Arabic information retrieval," *J. Big Data*, vol. 7, no. 1, pp. 1–19, 2020, doi: 10.1186/s40537-020-00310-z.

[87] T. Khatoon and A. Govardhan, "Query Expansion with Enhanced-BM25 Approach for Improving the Search Query Perfor- mance on Clustered Biomedical Literature Retrieval," *J. Digit. Inf. Manag.*, vol. 16, no. 2, pp. 85–98, 2018.

[88] H. K. Azad and A. Deepak, "A New Approach for Query Expansion using Wikipedia and WordNet," *Inf. Sci.*, vol. 492, pp. 147–163, 2019, [Online]. Available: http://arxiv.org/abs/1901.10197.

[89] O. El Midaoui, B. El Ghali, A. El Qadi, and M. D. Rahmani, "Geographical Query reformulation using a Geographical Taxonomy and WordNet,"

Procedia Comput. Sci., vol. 127, pp. 489–498, 2018, doi: 10.1016/j.procs.2018.01.147.

[90] Y. Gupta and A. Saini, "A novel Fuzzy-PSO term weighting automatic query expansion approach using combined semantic filtering," *Knowledge-Based Syst.*, vol. 136, no. 15 November 2017, pp. 97–120, 2017.

[91] W. Li, S. Wang, and Z. Yu, "Deep Learning and Semantic Concept Spaceare Used in Query Expansion," *Autom. Control Comput. Sci.*, vol. 52, no. 3, pp. 175–183, 2018, doi: 10.3103/S0146411618030082.

[92] I. G. Veul, "Locally versus Globally Trained Word Embeddings for Automatic Thesaurus Construction in the Legal Domain," in *Benelearn 2017: Proceedings of the Twenty-Sixth Benelux Conference on Machine Learning Programme Committee Conference Track*, 2017, no. June, pp. 9–10, [Online]. Available:

https://pure.tue.nl/ws/files/72619856/benelearn_2017.pdf#page=113.

[93] Q. Huang and H. Wu, "QE-integrating framework based on Github knowledge and SVM ranking," *Sci. China Inf. Sci.*, vol. 62, no. 5, pp. 1–16, 2019, doi: 10.1007/s11432-017-9465-9.

[94] R. Kumar, "Social Book Search : Reranking based on Document and query expansion with keyword Filtering," in *Advanced Computing and Communication Systems (ICACCS), 2017 4th International Conference*, 2017, pp. 1–5.

[95] B. Selvaretnam and M. Belkhatir, "Coupled intrinsic and extrinsic human language resource-based query expansion," *Knowl. Inf. Syst.*, vol. 60, no. 3, pp. 1397–1426, 2019, doi: 10.1007/s10115-018-1267-x.

[96] B. Audeh, P. Beaune, and M. Beigbeder, "SMERA: Semantic mixed approach for web query expansion and reformulation," *Stud. Comput. Intell.*, vol. 665, pp. 159–180, 2017, doi: 10.1007/978-3-319-45763-5_9.

[97] M. M. Alqahtani and E. Atwell, "Developing Bilingual Arabic-English Ontologies of Al-Quran," in *2nd IEEE International Workshop on Arabic and Derived Script Analysis and Recognition, ASAR 2018*, 2018, pp. 96–101, doi: 10.1109/ASAR.2018.8480237.

[98] H. Z. Rad, S. Tiun, and S. Saad, "Lexical scoring system of lexical chain for Quranic document retrieval," *GEMA Online J. Lang. Stud.*, vol. 18, no. 2, pp. 59–79, 2018, doi: 10.17576/gema-2018-1802-05.

[99] S. Balaneshin-kordan and A. Kotov, "Embedding-based Query Expansion for Weighted Sequential Dependence Retrieval Model," *Proc. 40th Int. ACM SIGIR Conf. Res. Dev. Inf. Retr. - SIGIR '17*, pp. 1213–1216, 2017, doi: 10.1145/3077136.3080764.

[100] D. Li and E. Kanoulas, "Bayesian Optimization for Optimizing Retrieval Systems," in *Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining*, 2018, pp. 360–368.

[101] D. Zhou, X. Wu, W. Zhao, S. Lawless, and J. Liu, "Query Expansion with Enriched User Profiles for Personalized Search Utilizing Folksonomy Data," *IEEE Trans. Knowl. Data Eng.*, vol. 29, no. 7, pp. 1536–1548, 2017, doi: 10.1109/TKDE.2017.2668419.

[102] A. Keikha, F. Ensan, and E. Bagheri, "Query expansion using pseudo relevance feedback on wikipedia," *J. Intell. Inf. Syst.*, vol. 50, no. 3, pp. 455–478, 2017, doi: 10.1007/s10844-017-0466-3.

[103] F. C. Fernández-Reyes, J. Hermosillo-Valadez, and M. Montes-y-Gómez, "A Prospect-Guided global query expansion strategy using word embeddings," *Inf. Process. Manag.*, vol. 54, no. 1, pp. 1–13, 2018, doi: 10.1016/j.ipm.2017.09.001.

[104] P. Kotoula and C. Makris, "Query Disambiguation Based on Clustering Techniques," in *IFIP International Conference on Artificial Intelligence Application and Innovations*, 2018, pp. 133–145, doi: 10.1007/978-3-319-92016-0.

[105] A. Kanavos, P. Kotoula, C. Makris, and L. Iliadis, "Employing query disambiguation using clustering techniques," *Evol. Syst.*, pp. 1–11, 2019, doi: 10.1007/s12530-019-09292-7.

[106] B. Selvaretnam and M. Belkhatir, "Natural language technology and query expansion: Issues, state-of-the-art and perspectives," *J. Intell. Inf. Syst.*, vol. 38, no. 3, pp. 709–740, 2012, doi: 10.1007/s10844-011-0174-3.

[107] J. Mennes, "SenseDisclosure : A new procedure for dealing with problematically ambiguous terms in cross-disciplinary communication," *Lang. Sci.*, vol. xxx, pp. 1–11, 2018, doi: 10.1016/j.langsci.2018.06.003.

[108] A. Ferrari and A. Esuli, "An NLP approach for cross-domain ambiguity detection in requirements engineering," *Autom. Softw. Eng.*, vol. 26, no. 3, pp. 559–598, 2019, doi: 10.1007/s10515-019-00261-7.

[109] A. Dumitrache, L. Aroyo, and C. Welty, "Capturing Ambiguity in Crowdsourcing Frame Disambiguation," in *The Sixth AAAI Conference on Human Computation and Crowdsourcing (HCOMP 2018)*, 2018, no. Hcomp, pp. 12–20.

[110] O. Inel and L. Aroyo, "Validation Methodology for Expert-Annotated Datasets : Event Annotation Case Study," in *2nd Conference on Language, Data and Knowledge (LDK 2019). Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik, 2019.*, 2019, no. 12, pp. 1–12.

[111] K. Bousquet, T. Y. Swaab, D. L. Long, K. Bousquet, T. Y. Swaab, and D. L. Long, "The use of context in resolving syntactic ambiguity : structural and semantic influences in fl uences," *Lang. Cogn. Neurosci.*, vol. 0, no. 0, pp. 1–15, 2019, doi: 10.1080/23273798.2019.1622750.

[112] X. Yin, Y. Huang, B. Zhou, A. Li, L. Lan, and Y. Jia, "Deep Entity Linking via Eliminating Semantic Ambiguity With BERT," *IEEE Access*, vol. 7, pp. 169434–169445, 2019.

[113] D. E. Mohamed, "Ambiguity and Vagueness in the Qur'ān: A Semantic Investigation Into Significant Examples," *QURANICA-International J. Quranic Res.*, vol. 11, no. 2, pp. 38–52, 2019.

[114] B. A. Abu-ghararah, "The Translation of Qur ' anic Lexical Ambiguity," *J. Arts Humanit.*, vol. 08, no. 12, pp. 73–79, 2020.

[115] N. Abdelaal, "Lexical and Semantic Problems in Translation," *Arab. Transl. between English*, pp. 95–120, 2020, doi: 10.1007/978-3-030-34332-3_4.

[116] M. I. E. Ghembaza, "Specialized Quranic Semantic Search Engine : Features and Architecture," *Int. J. Comput. Sci. Inf. Secur.*, vol. 17, no. 2, pp. 63–74, 2019.

[117] S. J. Musahar, H. Talib, R. Musahar, F. AZIM, and M. Z. Zakaria, "Ambiguity in Holy Quran Commentaries: The Use of Polysemic Words 'Imam & Ummah,'" in *Proceedings of the Regional Conference on Science, Technology and Social Sciences (RCSTSS 2016)*, 2016, pp. 443–454.

[118] F. T. da Silva and J. E. B. Maiai, "Query Expansion in Text Information Retrieval with Local Context and Distributional Model," *J. Digit. Inf. Manag.*, vol. 17, no. 6, p. 313, Dec. 2019, doi: 10.6025/jdim/2019/17/6/313-320.

[119] N. Soudain, I. Bounhas, and S. Ben Babbis, "Ambiguity aware Arabic document indexing and query expansion: A morphological knowledge

learning-based approach," in *Proceedings of the 31st International Florida Artificial Intelligence Research Society Conference, FLAIRS 2018*, 2018, pp. 230–235.

[120] H. Wu, J. Li, Y. Kang, and T. Zhong, "Exploring noise control strategies for UMLS-based query expansion in health and biomedical information retrieval," *J. Ambient Intell. Humaniz. Comput.*, pp. 1–12, May 2018, doi: 10.1007/s12652-018-0836-x.

[121] L. Hou, B. Le Bruyn, and R. Kager, "Disambiguate or not? – The role of prosody in unambiguous and potentially ambiguous anaphora production in strictly Mandarin parallel structures," in *INTERSPEECH 2017*, 2017, pp. 1393–1397, doi: 10.21437/Interspeech.2017-1214.

[122] Z. Zhang, Y. Kudo, T. Murai, and Y. Ren, "Addressing Complete New Item Cold-Start Recommendation : A Niche Item-Based Collaborative Filtering via Interrelationship Mining," *Appl. Sci.*, vol. 9, no. 1894, 2019.

[123] Z. Zhang, Y. Kudo, T. Murai, and Y. Ren, "Improved covering-based collaborative filtering for new users ' personalized recommendations," *Knowl. Inf. Syst.*, vol. 62, no. 3, pp. 3133–3154, 2020.

[124] V. Klyuev and V. Oleshchuk, "Semantic retrieval : an approach to representing , searching and summarising text documents Vitaly Klyuev * Vladimir Oleshchuk," *Int. J. Inf. Technol. Commun. Converg.*, vol. 1, no. 2, pp. 221–234, 2011.

[125] G. Sidorov, "Latent semantic analysis (LSA): Reduction of dimensions," *SpringerBriefs Comput. Sci.*, pp. 17–19, 2019, doi: 10.1007/978-3-030-14771-6_4.

[126] F. gao Niu, "Basic Co-Occurrence Latent Semantic Vector Space Model," *J. Classif.*, vol. 1, no. 71503151, pp. 1–18, 2018, doi: 10.1007/s00357- 018-9283-9.

[127] F. S. Al-Anzi and D. AbuZeina, "Enhanced search for Arabic language using latent semantic indexing (LSI)," in *2018 International Conference on Intelligent and Innovative Computing Applications, ICONIC 2018*, 2019, pp. 1–4, doi: 10.1109/ICONIC.2018.8601096.

[128] N. Yusuf, M. A. M. Yunus, and N. Wahid, "A Comparative Analysis of Web Search Query: Informational Vs Navigational Queries," *Int. J. Adv. Sci. Eng. Inf. Technol.*, vol. 9, no. 1, pp. 136–141, 2019, doi:

http://dx.doi.org/10.18517/ijaseit.9.1.7578.

[129] F. D. Ahmad, "A Malay language document retrieval system: An experimental approach and analysis," Universiti Kebangsaan Malaysia, 1995.

[130] N. H. M. Noor, S. Sapuan, and F. Bond, "Creating the open wordnet Bahasa," in *PACLIC 25 - Proceedings of the 25th Pacific Asia Conference on Language, Information and Computation*, 2011, pp. 255–264.

[131] C. Dictionary, "Cambridge dictionary," *Cambridge University Press*, 2018. https://dictionary.cambridge.org/dictionary/english-malaysian/ (accessed Dec. 24, 2018).

[132] F. Christiane, *Wordnet: An Electronic Lexical Database*. Cambridge, MA, 1998.

[133] E. WordNet, "Euro WordNet English," *Euro WordNet*, 2020. https://en-word.et (accessed Mar. 20, 2020).

[134] L. Abouenour, K. Bouzoubaa, and P. Rosso, "On the evaluation and improvement of Arabic WordNet coverage and usability," *Lang. Resour. Eval.*, vol. 47, no. 3, pp. 891–917, 2013, doi: 10.1007/s10579-013-9237-0.

[135] Bab.la, "Online Dictionary for 28 Languages," *Oxford University Press*, 2018. https://en.bab.la/dictionary/english-arabic/ (accessed Dec. 24, 2018).

[136] W. Black *et al.*, "Introducing the Arabic WordNet project," *GWC 2006 3rd Int. Glob. WordNet Conf. Proc.*, pp. 295–299, 2005.

[137] CLE, "Urdu Wordnet 1.0," *CLE*, 2020. www.cle.org.pk/software/ling-resources/UrduWordNetWordlist.html (accessed Apr. 03, 2020).

[138] Ijunoon, "English to Urdu Text Translation," *Ijunoon*, 2018. https://www.ijunoon.com/urdudic/ (accessed Dec. 24, 2018).

[139] A. Saeed, R. M. A. Nawab, M. Stevenson, and P. Rayson, "A sense annotated corpus for all-words Urdu word sense disambiguation," *ACM Trans. Asian Low-Resource Lang. Inf. Process.*, vol. 18, no. 4, 2019, doi: 10.1145/3314940.

[140] Shamsuddeen, "Kamus," *Kamus*, 2020. http://kamus.com.ng/index.php (accessed Apr. 30, 2020).

[141] R. NL, "Stop words list in over 40 languages," 2020. http:www.ranks.nl.stopwords (accessed Mar. 16, 2020).

[142] M. F. Porter, "An algorithm for suffix stripping," *Program*, vol. 14, no. 3, pp. 130–137, 1980.

[143] A. Al-Omari and B. Abuata, "Arabic light stemmer (ARS)," *J. Eng. Sci.*

Technol., vol. 9, no. 6, pp. 702–717, 2014.

[144] A. Bimba, N. Idris, N. Khamis, and N. F. M. Noor, "Stemming Hausa text: using affix-stripping rules and reference look-up," *Lang. Resour. Eval.*, vol. 50, no. 3, pp. 687–703, Sep. 2016.

[145] H. A. Robbani, "Sastrawi." MIT, USA, 2018.

[146] A. Kumari and D. K. Lobiyal, "Word2vec's Distributed Word Representation for Hindi Word Sense Disambiguation," in *16th International Conference, ICDCIT 2020*, 2020, vol. 5966 LNCS, pp. 325–335, doi: 10.1007/978-3-030-36987-3.

[147] H. Karamti, M. Tmar, M. Visani, T. Urruty, and F. Gargouri, "Vector space model adaptation and pseudo relevance feedback for content-based image retrieval," *Multimed. Tools Appl.*, vol. 77, no. 5, pp. 5475–5501, 2018, doi: 10.1007/s11042-017-4463-x.