Myanmar Semantic Information Retrieval Using Self Organizing Map with Global Vector – MyanSeM

Thiri Haymar Kyaw, Thinn Thinn Wai, Thinn Mya Mya Swe
University of Information Technology
thirihaymarkyaw@gmail.com, thinnthinnwai@uit.edu.mm, thinnmyamyaswe@uti.edu.mm

Abstract

Nowadays, explosive growing the resources with Myanmar language on the Internet, the information retrieval (IR) for Myanmar web pages has increasingly important.The proposed system presents an effective semantic retrieval approach based on parallel Self Organizing Map (SOM), which uses document associationinstead of Euclidian distance in distance calculation, and GloVe (Global Vector for Word Representation) for word co-occurrence. *The Self Organizing Map (SOM) has been a promising method for document clustering and word sense disambiguation. This approach uses the parallel training the separate parts of SOM for document clustering, then combine, and re-cluster the documents.During the training of the parts of SOM, global vector is used for appearance of word co-occurrence and then combines the word categories based on semantic sense.Although many researchers have researched the various semantic information retrieval approaches, they have not yet adapted to retrieve the semantic information of Myanmar words and sentences. This approach can retrieve most semantically relevant web documents. This does not take too long time for SOM training because of parallel GPU approach.*

Keywords-Semantic Web, Self-Organizing Map, Information Retrieval.

1. Introduction

The World Wide Web serves the vastly distributed information services for every kind of information such as news, advertisements, blogs, customer relationship management, online learning, e-government, e-commerce, health services, context awareness services, etc. with custom languages. Myanmar language is same as the other languages like English. Some words are polysemy or homonyms and some words are synonyms. In addition, Myanmar words are un-segmented words and the delimitation of words is based on the typing of the user. Sometimes, the user does not type the word segments properly. The well-known search engines segment the Myanmar words according to the space delimitation and do not consider semantically related Myanmar words. They cannot retrieve the user-satisfied result. They retrieve a large number of irrelevant documents that are unable to meet the user's request. Analyzing the semantic meaning of words in user's query performs semantic information retrieval. The proposed system focuses on the Myanmar web pages for retrieving most relevant results.

There have been several researches applying Self Organizing Maps (SOMs) for word sense disambiguation [10], web log clustering [4, 5], document clustering and information retrieval [1, 3, 6, 7, and 10]. The Self-Organizing Feature Map (SOFM or SOM) is a clustering and data visualization technique based on a neural network proposed by Teuvo Kohonen [18, 19]. SOM is a model of unsupervised machine learning and an adaptive knowledge representation scheme. SOM consists of components called nodes or neurons. Associated with each node is a weight vector of the same dimension as the input data vectors and a position in the map space. The usual arrangement of nodes is a two-dimensional regular spacing in a hexagonal or rectangular grid. The self-organizing map describes a mapping from a higher dimensional input space to a lower dimensional map space. The procedure for placing a vector from data space onto the map is to find the node with the closest (smallest distance metric) weight vector to the data space vector [3].

The previous SOM based clustering approaches can organize the document map from word category [21] or WordNet [17] and enhance the information retrieval [10, 21]. Most of them are not consider for the semantic relation of words or words co-occurrence. They use one SOM for training the whole document collection and it takes huge amount of time and memory. Separating the parts of SOM and running in parallel [15] is the more promising approach for training the document collection. For those purpose of word representation, word vector approaches such as SVD based methods, continuous bag of words (CBOW), Skip-Gram, and Global Vectors (GloVe) [9]. The previous approaches are not convenient for the semantic meaning of the Myanmar words in the query and the documents very well. The proposed MyanSem method provides a deep SOM with GloVe for the retrieval of Myanmar documents.

2. Related Work

Pushpa et al. compared web page recommendation systems using K-means and Self Organizing Map [3].

Both the methods used historical browsers data for search key words and provided users with most relevant web pages. All users' click-through activity such as number of times he visited, duration he spent, and several other variables were stored in database. These systems used this database and process to cluster and rank them. The obtained results showed that the Self Organizing Map technique produced the most relevant results for a particular query word compared to K-means technique.

Tarek F. Gharib, et al. proposed the semantic text document clustering approach that using WordNet lexical and Self Organizing Maps [17]. This approach used the WordNet to identify the importance concepts in the document. The SOM is used to enhance the effectiveness of document clustering algorithms. This approach took the advantages of the semantics available in knowledge base and the relationship between the words in the input documents. They have two reasons for using SOM that it is topologically preserving and clustering is performed non- linearly on the given input data sets. The topologically preserving property allows the SOM applied to document clustering, to group similar documents together in a cluster and organize similar clusters close together unlike most other clustering methods.

Xia Lin [21] proposed a Self-Organizing semantic map for retrieval of AI literature. The purpose of this system was to conceptualize art information retrieval approach, which used traditional search techniques as information filters and the semantic map as a browsing aid to support ordering, linking, and browsing information gathered by the filters.

Peter Gajdos and Pavel Moravec presented a simple modification of classic Kohonen SOM; they called Global-Merged SOM. Their approach allowed parallel processing of input data vectors or partitioning the problem for all vectors from the training in reducing the memory consuming. The set of input vectors is divided into a given number of parts. The classic SOM is applied on every part. In the final phase, pre-selected potential centroids of data clusters are used as weight vector. Their algorithm can utilized the power of batch processing in all inner parts (PSOM) [15].

3. The Proposed System

The focus of the MyanSeM method is to cluster Myanmar document collections using parallel SOM, in order to consider dimension reduction by finding word co-occurrence based on global vector [9]. The proposed method also addressed for the polysemy words of the query sentence choosing the correct sense.

3.1. Preprocessing

First, the information retrieval systems make the preprocessing for the crawled web pages such as removing HTML tagsand identifyingthe main content blocks. After removing HTML tags, the preprocessing tasks such as stopword removal, stemming, and handling of digits, hyphens, punctuations are also required [2]. Myanmar language is un-segmented language like Chinese, Japanese, and Thai. The tokenization or segmentation process is more difficult than space-delimited languages like English. The part of speech tagging of Myanmar words based on the corpus or lexicon is needed for constructing the term-document vector.The system also segments the query sentence to terms or words. A document in the vector space model [2] is represented as a weight vector, in which each component weight is computed based on some variation of TF (Term Frequency) and TF-IDF (Term Frequency-Inverse Document Frequency) scheme. In this paper, we do not explain detail about pre-processing and word segmentation.

3.2. Self-Organizing Map

Self-organization map (SOM) is an unsupervised learning method where the neural network organizes itself to form useful information [12]. Kohonen innovated this principle of topographic map formation [18, 19]. The SOM uses a set of neurons, regularly arranged in one dimension or two dimension rectangular or hexagonal grid, to form a discrete topological mapping of an input space. The architecture of two-dimensional hexagonal grid Kohonen self-organizing map is illustrated in Figure 1.

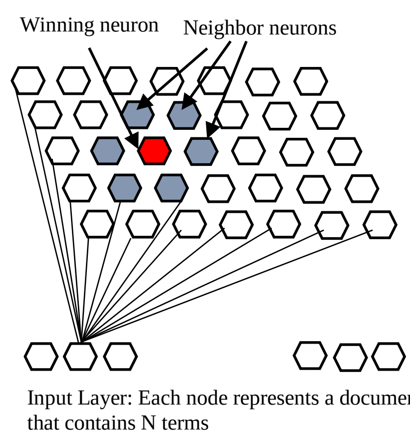


Figure. 1 Architecture of Self-organizing map

Continuous input values are presented sequentially in time through the input layer, without specifying the desired output. Each pattern is represented in the form of a vector, $x = (x_1, x_2, ..., x_n)$. Computational procedures of SOM [20] are summarized as follows:

Step 0. Initialize

- Small real random values to weights, w_{ij} for $i = 1$ to n and $j = 1$ to m.

- A neighborhood parameter, h and a learning rate, α, where $0 \leq \alpha \leq 1$

Step 1. Enter a new input vector, $x = (x_1, x_2, ..., x_n)$, to the input layer.

Step 2. Select winner neuron with the smallest distance to x.

Step 3. Modify Weights.

- Adjust the winner and its neighbors' weights according to the following formula:

$$w_j(t+1) = w_j(t) + \alpha(x(t) - w_j(t))$$

Step 4. Update learning rate α. Slowly reduce radius r at specified iterations.

Step 5. Continue from Step 1 to 4 until weights have stabilized.

The SOM algorithm always converges to a solution, i.e., that each of the winner weight vectors of the map converges to the mean of the data vectors for which it has been a winner, in a finite number of steps [18,19].

3.3. GloVe for Word Representation

In the *GloVe* model [9], the global corpus statistics are captured directly by the model. Let the matrix of word-word co-occurrence counts be denoted by X, whose entries X_{ij} tabulate the number of times word j occurs in the context of word i. Let $X_i = \sum_k X_{ik}$ be the number of times any word appears in the context of word i. Finally, let $P_{ij} = P(j|i) = X_{ij} / X_i$ be the probability that word j appear in the context of word i [9]. The simple example of Myanmar word co-occurrence probabilities is shown in Table 1.

Table. 1 Co-Occurrence Probabilities for Target Words တူ (hammer) and သတ် (kill) with Selected Context Words ထု (beat), ရိုက်(strike), စား(eat) and တော်စပ် (relate)

Probability and Ratio	k = ထု	k = ရိုက်	k = သေ	k = တော်စပ်
P(k| တူ)	3.4×10^{-3}	1.5×10^{-4}	0.3×10^{-4}	1.7×10^{-4}
P(k| သတ်)	2.9×10^{-3}	1.7×10^{-4}	4.1×10^{-4}	2.1×10^{-5}
P(k| တူ)/ P(k|သတ်)	1.17	0.88	7.3×10^{-2}	8.1

Consider two words i and j that exhibit a particular aspect of interest. The relationship of these words (တူ[hammer or chopstick],သတ်[kill],ထု[beat], ရိုက်[strike],စား[eat],တော်စပ် [relate]) can be examined by studying the ratio of their co-occurrence probabilities with various probe words, k. After doing vector difference, the highest occurrence probability ratio of words i, j to other work k is point out. For words k related to တူ but not သတ်, say k = တော်စပ်, we expect the ratio P_{ik}/ P_{jk} will be large. Similarly, for words k related to သတ် but not တူ, say k = သေ, the ratio should be small. For words k like ထု or ရိုက်,either these are related to both တူ and သတ်, or to neither, the ratio should be closeto one.

Theratio P_{ik} / P_{jk} depends on three words i, j, and k,the most general model takes the form,$F(w_i, w_j, \bar{w}_k) = P_{ik} / P_{jk}$ where $w \in \mathbb{R}^d$ are word vectors and $\bar{w} \in \mathbb{R}^d$ are separate context word vectors.

We trained our model on one lexicon based on 100 Myanmar news, commercial and blog web sites with 1 million tokens. We tokenizeand build a vocabulary of the 1000 most frequent words and then construct a matrix of co-occurrence counts. Currently, we omit the name entity recognition.

3.4. MyanSeM Algorithm

The proposed algorithm has two phases: Training phase and Query phase. In the training phase, it provides the clustering of Myanmar documents into related groups by caring the semantic categories and co-occurrence of words. In the query phase, it considers the semantic sense of the query sentence, maps to the most related clusters and retrieves the most relevant web pages.

Let T is the term vector and $T \in t_j$, where $j = 1$ to N. N is the number of terms in the document collection (D) and $D \in d_i$ where d_i represents each document and $i = 1$ to M. M is the number of documents. Each term (Myanmar word) comes from the Myanmar Lexicon that includes the Myanmar word, Myanmar meaning (definition), and English meaning and example Myanmar sentences of all homonyms. Term-document matrix is filled with weights of terms in each document.

SOM is composed of neurons that are also called nodes. Each node represents one document (d_i). Weight vector of each document $d_i = [w_1, w_2, ..., w_n]$. Each weight w_j of term j in document i represents TF(term frequency) * IDF (inverse document frequency) [2].

The proposed approach uses the two-dimensional hexagon shape SOM. In this approach, we use parallel SOM based on GPU-based SOM [15] that divides the document vector into parts of the vector. The collection of document D is divided into P number of parts. One part contains K number of documents.The partition of document collection is based on the number of documents in the collection.In the sample test in Section 4, we divide four partition (P) on 100 documents and 25 documents in each partition.

The stepbystep procedure for MyanSem algorithm is as follows:

[Training Phase]
Step 1: SOM with GloVe

– Train SOM for clustering documents within P partition
– Choose one document d among K documents randomly as an input vector.
– Find the most associate document $d_i \in K$, $i = 1$ to K.

Association of two documents: $\frac{|d \cap d_i|}{|d|}$

where
$| d \cap d_i |$ means number of words (terms) contain in both document d and d_i,
$| d |$ means the number of words (terms) contain in document d.

– Extract terms that contains in both input document and the most associate document are called shared terms S.
– Construct the Global Vector for those shared terms and the other terms that contains separately in two documents.
– Determine co-occurrence words (terms) that have highest ratios. Union those term vectors.
– Train SOM as traditional SOM until convergence for K documents. Terms have already reduced by combining co-occurrence terms.
– Compose L number of cluster that contains most associate documents. Select the cluster centroids (winner) for all L clusters.

Run Step 1 in parallel for P partitions.

Step 2: Combine the clusters

– Take one cluster randomly and the centroid node (document vector) within this cluster acts as the input vector.
– Train the SOM for all clusters until no changes occur.

[Query Phase]
Step 3: Sense the Query

– Construct the query word vector
– Check the homonyms of words using the Myanmar wordnet like lexicon.
– Construct local co-occurrence vector of query words and map to the examples of word in lexicon.

Table. 2 Query Word Co-Occurrence Matrix Applies to the Example Sentences in Lexicon

	တူ (hammer)	ဖြင့် (with)	ထု/ရိုက် (beat/strike)
တူ (hammer)	0	e1, e2	0
ဖြင့် (with)	e1, e2	0	e2
ထု/ရိုက် (beat/strike)	0	e2	0

Let consider the simple example query "တူဖြင့်ရိုက်". The word ″တူ″ has four different meanings (chopstick, hammer, relate, same) and the word ″ရိုက်″ has a synonym ″ထု″ that is already combine in the term vector. The examples are e1:"ခေါက်ဆွဲကိုတူဖြင့်စားသည်", e2:"သံကိုတူဖြင့်ထုသည်" and e3:"တူတော်စပ်သည်" e4: "ဆင်တူသည်".

In the query, word co-occurrence matrix shown in table 2, the example sentence of e2 is best match with the query. So, the correct sense of word (hammer) is selected.

Step 4: Map the Query to the document clusters

– The query vector enters as the input vector. The words that are not contains in the query sentence are filled with zero.
– Calculate the document similarity or association between the query vector and the cluster centroids.
– Choose the best match cluster and rank the documents within this cluster using query co-occurrence matrix applying these documents.

4. Discussion

We collect the 100 documents for training and the example sentences are shown in Figure 2. There are 50 total vocabulary (terms) used in the documents. The 50 × 100 document-term vector is separated into four 50 × 25 vectors. First we trains the SOM on first 50 × 25 vector with some epochs and construct the global vector for words (terms) within each cluster. Then we can reduce the 25 terms to 20 terms that have high co-occurrence probability. Then, we continue to train the SOM for 50 × 20 vector. As the above mentioned, we apply four SOM parallel and then combine after running the SOM until it reaches convergent. Finally, the semantically related documents are placed together within the same cluster. The sample clusters are shown in Figure 3.

When the user enters the query "တူဖြင့်ရိုက်", first we determine the correct sense of every word in the query sentence as mentioned in Step 3. The query enters as an input to the combined SOM and chooses the winner and neighbors as the query results. We use the cosine of the angle (cosine similarity)[2] as the distance measure for choosing the winner of SOM as follows:

$$cosine(\mathbf{d}_j, \mathbf{q}) = \frac{\langle \mathbf{d}_j \bullet \mathbf{q} \rangle}{\| \mathbf{d}_j \| \times \| \mathbf{q} \|} = \frac{\sum_{i=1}^{|V|} w_{ij} \times w_{iq}}{\sqrt{\sum_{i=1}^{|V|} w_{ij}^2} \times \sqrt{\sum_{i=1}^{|V|} w_{iq}^2}}$$

Assume that q is the query vector q and $d_j$is the document vector, which is the centroid of each cluster. Most similar cluster centroid (winner node of SOM cluster) is selected. Finally, we construct the word vector of the query words (terms) and words (terms) in the documents within the winner cluster. The result of the above query is shown in Figure 4.

d1	ကလေးသည် ခေါက်ဆွဲကို တူနှင့် မစားတတ်ပဲ ဖြစ်နေသည်
2	ဂျပန်လူမျိုးများသည် ထမင်းကို တူဖြင့် စားသော အမူအကျင့် ရှိသည်
d3	တရုတ်လူမျိုးများသည် တူ ဖြင့် စားသော အလေ့အထ ရှိကြသည်
d4	တူကို ဝါးဖြင့် ပြုလုပ်သည်
d5	တူဖြင့် မသေမချင်း ထုသတ်ခဲ့သော လူသတ်သမားကို ထောင်ချခဲ့သည်
d6	ပန်းပဲသမားသည် သံကို တူဖြင့် ထု၍ လိုရာ ပုံဖော်သည်
d7	ဖားကို တူဖြင့် ရိုက်သတ်သည်
d8	မြမြသည် ခေါက်ဆွဲကို တူဖြင့် စားလေ့ရှိသည်
d9	သံကို တူဖြင့် ရိုက်နေသည်
d10	သူ ပစ်လိုက်သော တူချောင်းသည် နံရံတွင် စိုက်သွားသည်
d11	သူမတွင် တူနှစ်ယောက် ရှိသည်
d12	သူမတို့ ညီအစ်မသည် ရုပ်ဆင်း တူသည်
d13	သူမသည် နာမည်ကျော် မင်းသမီးနှင့် ရုပ်ချင်းဆင်သည်
d14	သူမသည် အမျိုးသားကို တူဖြင့် ထုသတ်ခဲ့သည်
d15	သူသည်ခိုးဝင်လာသူကို တူဖြင့် ထုသတ်လိုက်သည်
d16	သူသည် အဖေနှင့် ရုပ်ချင်း တူသည်
d17	အမြွှာနှစ်ယောက်သည် ရုပ်ဆင်း ချွတ်စွပ်တူသည်
d18	အိမ်မြှောင်ကို တူဖြင့် ထုသတ်မိသည်
d19	မောင်မောင်သည် မြမြ၏ တူတော်သည်
d20	ကြူကြူသည် အမေနှင့် ရုပ်ချင်းဆင်သည်

Figure. 2.Document Collection for Case Study

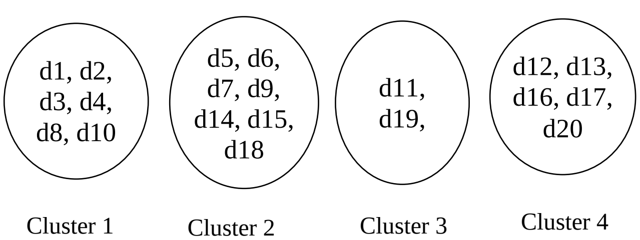


Figure. 2 Sample Clusters for 20 Documents

For the evaluation, the *F*-measure is computed as follows:

*F = 2 * ((Precision * Recall) / (Precision + Recall))*

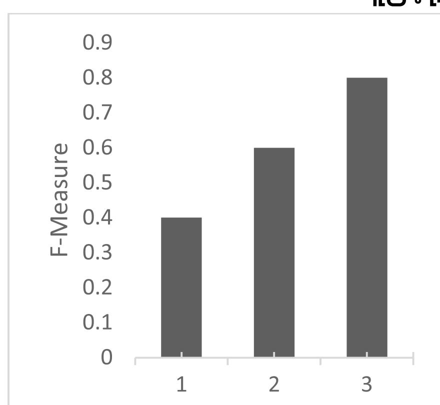
$$\text{precision} = \frac{|\{\text{relevant documents}\} \cap \{\text{retrieved documents}\}|}{|\{\text{retrieved documents}\}|}$$

$$\text{recall} = \frac{|\{\text{relevant documents}\} \cap \{\text{retrieved documents}\}|}{|\{\text{relevant documents}\}|}$$

The F-measure for three approaches is illustrated in Figure 5. The preliminary result shows that the proposed approach can retrieve more relevant Myanmar web documents. However, the evaluation in figure 5 is based on sample data and we plan to test with huge Myanmar document collection with large Myanmar lexicon like WordNet.

d9	သံကို တူဖြင့် ရိုက်နေသည်
d18	အိမ်မြှောင်ကို တူဖြင့် ထုသတ်မိသည်
d7	ဖားကို တူဖြင့် ရိုက်သတ်သည်
d14	သူမသည် အမျိုးသားကို တူဖြင့် ထုသတ်ခဲ့သည်
d15	သူသည် ခိုးဝင်လာသူကို တူဖြင့် ထုသတ်လိုက်သည်
d5	တူဖြင့် မသေမချင်း ထုသတ်ခဲ့သော လူသတ်သမားကို ထောင်ချခဲ့သည်
d6	ပန်းပဲသမားသည် သံကို တူဖြင့် ထု၍ လိုရာ ပုံဖော်သည်

Figure. 4. Result of the query "တူဖြင့်ရိုက်"



1: Myanmar Information Retrieval with no semantic information
2: Traditional SOM Clustering
3: SOM with GloVe

Figure. 5. F-Measure of Three Approaches

In addition, the traditional recall measure has problem that concerns the impossibility of collecting exhaustive relevance judgments in a realistically large document set. Any potentially relevant document has not been missed when making relevance judgments because judges would have to go through the entire document set of nearly a million documents, which is infeasible [16]. Therefore, wewill use the pooling method to evaluate the proposed system in future. The document pool to be manually judged is constructed by putting together the top N retrieval results from a set of n systems [16].As the future work, we will evaluate and prove the result of the proposed approach by comparing traditional SOM clustering approach.

5. Conclusion

Although the semantic information retrieval for many languages have been improved, semantic information retrieval for Myanmar language still needs more researches. In this paper, we propose the novel algorithm MyanSem that is based on the parallel SOM with GloVe. This approach also modifies the distance calculation of SOM to appropriate the document association. GloVe is the word representation model that determine the words co-occurrence. This system provides the semantic information retrieval for Myanmar web pages. Therefore, the proposed system can retrieve the most relevant pages that meet the user-satisfied results. It can remove the irrelevant, meaningless pages or advertisement pages from the query result.

6. References

[1] A. Ahmad and R. Yusof, A Modified Kohonen Self Organizing Map (KSOM) Clustering for four Categorical Data, Journal of Technology (Science and Engineering), 2016, pp 75-80.

[2] B. Liu, "Web Data Mining: Exploring Hyperlinks, Contents, and Usage Data", *Data-Centric Systems and Application Series, Springer-Verlag Berlin Heidelberg*, 2011.

[3] C. N. Pushpa, J. Thriveni, K. R. Venugopal and L. M. Patnaik, "Web Page Recommendation System using Self Organizing Map Technique", *International Journal of Current Engineering and Technology*, Inpressco, 2014, Vol. 4, No.5, pp. 3270-3277.

[4] C. Sadhana, L. Mary, I. Sheela, "Enhanced Self Organizing Map Algorithm for Web Usage Mining Through Neural Network", *International Journal of Trend in Research and Development*, Volume 3(6), 2016.

[5] D. Qi and C.C Li, Self Organizing Map based Web Pages Clustering using Web Logs, Proceedings of the 16th International Conference of Software Engineering and Data Engineering, SEDE, Las Vegas, Nevada, July, 2007, pp 265-270.

[6] E. Chifu and C. Cenan, "Discovering Web Document Clusters with Self –Organizing Maps", *Scientific Annals of the "Alexandru Ioan Cuza" University of Iaşi Computer Science Section*, Tome XIV, 2004, pp. 1-10.

[7] H. C. Yang, C. H. Lee, and K.-L. Ke, "TSOM: A Topic-Oriented Self-Organizing Map for Text Organization", *World Academy of Science, Engineering and Technology, International Journal of Computer and Information Engineering*, Vol:4, No:5, 2010.

[8] H. Yin, "The Self-Organizing Maps: Background, Theories, Extensions and Applications", *Studies in Computational Intelligence (SCI)* 115, 715–762, 2008.

[9] J. Pennington, R. Socher, C. D. Manning, "GloVe: Global Vectors for Word Representation", *Empirical Methods in Natural Language Processing*, 2014, pp. 1532-1543.

[10] L. Zhang, "An Intelligent Information Retrieval Algorithm based on Knowledge Discovery and Self-Organizing Feature Map Neural Network", *International conference on Inventive Computation Technologies (ICICT)*, 2016.

[11] Martin Marcel Couturier, Disambiguating Words with Self Organizing Maps, Master Thesis, Massachusetts Institute of Technology, June, 2011.

[12] M. Negnevitsky, Artificial Intelligence, A Guide to Intelligent Systems, 2nd Edition, Pearson Education Limited 2005.

[13] M. Sasaki, "Latent Semantic Word Sense Disambiguation Using Global Co-Occurrence Information Using Non-Negative Matrix Factorization", Journal of Computer Science Applications and Information Technology, Vol (2). No. (3), pp. 1-4, 2017.

[14] N. Ampazis and S. J. Perantonis, LSISOM – A Latent Semantic Indexing Approach to Self-Organizing Maps of Document Collections, *Neural Processing Letters*, Vol 19, pp. 157-173, 2004.

[15] P. Gajdos and P. Moravec, Two-step Modified SOM for Parallel Calculation, *J. Pokorny, V. Sna sel, K. Richta (Eds.): Dateso*, 2010, pp. 13-21.

[16] S. Teufel, "An Overview Of Evaluation Methods In Trec Ad Hoc Information Retrieval And Trec Question Answering", L. Dybkjær et al. (eds.), Evaluation of Text and Speech Systems, 2007, pp. 163–186.

[17] T. F. Gharib, "Self Organizing Map based Document Clustering Using WordNet Ontologies", *JCSI International Journal of Computer Science Issues*, Vol. 9, Issue 1, No 2, January 2012.

[18] T. Kohonen, Self-organization and Associative Memory, *Springer-Verlag, N.Y*, 3rd edition. 1989.

[19] T. Kohonen, S. Kaski, K. Lagus, J. Salojärvi, J. Honkela, V. Paatero, A. Saarela: Self-organization of a massive document collection, *IEEE Transactions on Neural Networks*, Vol. 11, no. 3, pp. 574-585, 2000.

[20] T. Munakata, Fundamentals of the New Artificial Intelligence, Neural, Evolutionary, Fuzzy and More, 2nd Edition, Springer-Verlag London Limited 2008.

[21] X. Lin, D. Soergel, G. Marchionini, "A Self-organizing Semantic Map for Information Retrieval", *Proceedings of the 14th annual international ACM SIGIR conference*, 1991, pp. 262-269.00000.00.

[22] http://learning.maxtech4u.com/self-organizing-map-som/.