Modification of Keyword Selection Process to Get Least List with Weighted Keywords by using Essence of both 'Baxendale' and 'Swanson' Experiment

Sandip Ghosh Assistant Librarian Future Institute of Engineering and Management

ABSTRACT

The 'keyword-based search system' by its name implies that the document is presented with the correct keyword. Keyword-content relations need to be thoroughly analyzed in order to select a document's keywords. In this case, the current method used, even after the use of ontology, has not been adequate. That's why many keywords are being selected for a particular document. And users are bounded to use all those keywords in search of a single document, resulting in a decrease in the precision of retrieval day by day. So in this case it is not possible to get high precision and to reduce quantity of keywords without selecting high quality keywords. The purpose of this paper is to examine how informative a particular keyword is, and to select the highest and most informative keyword and to exclude the least informative keyword through which, 'keyword-based search system' can get fewer but higher quality or weighted keywords as a helping hand to solve 'High Recall and Low Precision'.

Keywords

Grammatical-Hierarchical Logic; Parts of Speech; Used Meaning; Keyword-Based Search; Weighted Keyword; KWIC

1. INTRODUCTION

A keyword-based search system is a complete and sophisticated method to find web documents from an electronic data base. Finding a document by keyword is one of the most widely used and user-friendly method available today. A keyword is the word used to convey the original meaning of the sentence. So, it can be said that keywords need to be very informative in terms of a sentence. For a single document, the keyword is usually selected from the 'title' of the document, but can also sometimes be selected from the text, table of contents etc. So the keyword can be called the spokesperson for a document. That is why keywords are taken as the backbone of KWIC indexing or keywords-based search system.

2. THOUGHT OF TOPIC

Keywords need to be a lot more informative and usable in order to give long-term life to the keyword-based search system and also to KWIC indexing. Therefore, in these methods, there are regular developments due to the addition of new features. For this purpose, 'ontology' i.e. 'ontology based semantic indexing' is being used very much basically for semantic analysis on keywords to convert words to meanings and creation of synonymous words of keywords by contextual definition [10]. But in terms of practical experience, many critics said that, 'Search of a topic may have to be done by several keywords'. It can be seen here that the selection of large quantities of keywords neglects the quality of the keywords. In fact, the use of too many keywords reflects a lot of documents, which are not directly relevant or unnecessary to the user's search. So, this huge amount of keywords and the huge retrieval of documents served by it are driving this keyword-based search system to its core problem quickly i.e. "High Recall and Low Precision". This problem can only be solved if high quality or most informative keywords will be selected by proper analysis of keyword-content relation against any document.

Problems of Keyword-based search system:-

- Selection of keywords is made by the mandatory rule.
- Keyword-Content relation against any document not properly judge.
- Search of a topic may have to be done under several Keywords.
- Searcher vary often lead to "High Recall & Low Precision".
- Search time is high enough.

If I analyze the main criticism of this system, then I can say that the root of the problem lies in the beginning i.e. in the stage of keyword selection process. As we all know, there is a mandatory rule (Stop-List) [15, 16] for keyword selection. that is, everything must be selected as a keyword without a 'Stop-List'. This mandatory rule is the real problem area, because, Luhn was pointed out in his original proposal for KWIC and other types of automatic indexing that, the importance of this rule or the reason of its establishment was to differentiate significant words from non-significant [17, 18]. He said that, "since significance is difficult to predict, it is more practicable to isolate it by rejecting all obviously nonsignificant or 'common' words, with the risk of admitting certain words of questionable value. Such words may subsequently be eliminated or tolerated as 'noise'. A list of non-significant words would include articles, conjunctions, prepositions, auxiliary verbs, certain adjectives, and words such as 'report', 'analysis', 'theory', and the like. [13]. So, this rule does not give us any option for checking the keyword-content relation to select the most effective keywords and fired the less informative keywords. Actually this rule had made to support much for selection of 'context' oriented keywords than 'content' related one [14]. Therefore, Excessive keywords are made by taking non-useful keywords i.e. information providing capability is not properly evaluated, even by the use of ontology, that is, the keyword-content relation against any document. It can also be said that there is no systematic way out to make hierarchy of keywords according to their weight earned by degree of content

representation, even not using by ontology. As a result, there is no scope for deciding which keywords to accept, how many keywords to accept. So, the main purpose of this paper is to reduce the amount of keywords in the **keyword selection process** through a new scientific judgmental method so that the **mandatory rule (Stop-List)** can be overcome and establish a **content related selection process** for KWIC indexing and also keyword-based search system.

3. LITERATURE REVIEW

In the 1950s and later, computers were gradually being used as a means of data gathering. In 1961, H.P.Luhn invented the method of data gathering by keywords, whose name is 'Keyword In-context' indexing. This is an automatic indexing system. The 'KWIC' is an indexing system created by 'Word' where every word is released, with its own list of strings [1].

Many keywords are seen in the indexing of the KWIC. So, in response to a short search, large results are available, whose access becomes very difficult. So, Mika Kaki invented 'fKWIC' in 2005, in which the interface can be filtered by the most used keywords [2].

In the indexing process of traditional KWIC, many options have been used for modification, such as 'AKWIC'. Here we tried to modify the output format of the KWIC so that the system could be smoother [3].

So, for enhancement of KWIC rather to say keyword based searching process time by time new modification is being made. Someone is made for filtering the searching interface and someone is for modifying the output format. Here come another method namely 'Ontology' which act on retrieval by semantic analysis of keywords and also conceptual analyze of the keywords.

In 1992 Tom Gruber proposed the following definition "An Ontology is a specification of a conceptualization" [4].

In philosophical discipline, Ontology deals with the nature and structure of 'reality'. Aristtotle defined 'Ontology' as the science of 'being qua being' i.e. the study of attributes that belong to things because of their very nature [5].

In computer science, we refer to ontology as a special kind of information object or computational artifact. The ontology engineer analyzes relevant entities and organizes them into concepts and relations, being represented, respectively, by unary and binary predicates [6].

Pragmatically, a common ontology defines the vocabulary with which quarries and assertions are exchanged among agents. Ontological commitments are agreements to use the shared vocabulary in a coherent and consistent manner [7].

Ontology is a method to provide the common vocabulary for web document searching purpose. It defines as a shared understanding of some domain of interest. It acts by parsing the text from very basic to very advanced using different natural language processing technique [8, 9].

Therefore, ontology based semantic indexing act as a modification of keyword based searching or KWIC indexing system. I try to make a new retrieval structure of extracted keywords by conceptual analysis and prepare the documents collections into an easily accessible representation of documents.

4. OBJECTIVE

This paper highlights the main problem of the keyword-based search system i.e. "High Recall and Low Precision". The source of the problem has been identified to solve this problem i.e. occurring huge amount of keywords, and to quantify as many keywords as possible through the introduction of new method. Above all, remove the **mandatory rule** of **keyword selection process** and modification of keyword-based search system and also KWIC indexing.

5. SCOPE

Here is an attempt to fix a web document's retrievable problem in a keyword-based search system. The searcher chooses to retrieve a highly executable document with keywords in each case. Therefore, search results can be controlled only by keywords. That's why the only way to get a relevant, precise or close connected document is to select qualitative, weighted and informative keywords. This will automatically reduce amount of keywords, at the same time reduce amount of retrieve documents and search time that will make system more user-friendly.

6. METHODOLOGY

To solve the main problem of the keyword-based search system, here, landing on a new method i.e. "Grammatical-Hierarchical Logic" with merging of previous two experiment made by 'Baxendale' and 'Swanson' in their own perspectives. It organises two approaches 1) Parts of Speech approach, 2) Used Meaning (Context Definition) approach, which will be used in **keyword selection process** basically in 'Parse Page' field. At first, performance will be evaluated based on the result from the existing method in the parse page field. Then the result obtained by the new method will show and analyses what modification has happened. Finally, by comparing this two methods, it will be shown how much the system is improving or detoriate.

7. DETAILS OF PAPER

Keyword-based search system is the best option for searching the huge amount of web documents in the present era. Selecting a keyword from the title of a document or somewhere else is a very easy task. But with complexity arising over time, modification to this system continues to get better and upgrade result, as ontology is currently being used i.e. 'ontology based semantic indexing' for defining context of any keyword to open an option for alternative access point by synonyms of selected keywords. Nonetheless, this method is experiencing an ongoing problem that is, raising a large number of irrelevant or less connected keywords in the context of each web document. So first I try to observe how the existing method works, even after using ontology.

In the existing method, the field name 'Parse Page' extracts keywords from the document. There are two steps, namely, 1) Stop Listing, 2) Stemming. At this stage, all the verifications, such as content or conceptual analysis and context definition analysis, are performed to select the best informative keyword.

Stop Listing: Each page has a large amount of stop words. That is, words that do not carry any valuable information about the document, such as 'it', 'can', 'the' etc. Conversely, these bullets can be neglected, and crews are forced to take as keywords as they are considered sufficiently informative. That is, they are more content-based on conceptual analysis.

At the same time, keywords are defined by their context using ontology [11, 12] and the word net module so that an alternative search option can be created by selecting the synonymous words in relation to the keywords.

Stemming – A process to normalize morphologically similar words in a single root form or stem is called stemming. For example, 'connect', 'connected', and 'connection' all is represented by 'connect'. It is used to represent all documents by a particular word irrespective of its various forms for users' beneficiary.

Taken an example to show the operation of existing system along with using 'Ontology' method and its results obtain:

"I am sitting on the deck of the ship"

Here, selected keywords are – 'Sitting', 'Deck', and 'Ship' (after excluding Stop-List). Also some synonyms are selected on keywords as alternative keywords for better search against this document. Also stemmed words are created. All are shown against keywords as below-

Κ	Existing Method (With Ontology)		
	Stop Listing Step: Sitting, Deck, Ship		
Е	Synonym:		
	Sitting => Sit (Define Context)		
Y	Take a chair, Perch, Flop, Flump, Rest		
	Deck => Floor (Define Context)		
W	Ground, Storey, Tier, Level, Entresol.		
	Ship => Boat (Define Context)		
0	Vessel, Craft, Ferry ,Yawl		
	Stemming Step:		
R	Deck ~ Decked out etc.		
	Ship ~ Ship out, Shipping etc.		
D	Sitting ~ Sit up etc.		

But there are some issues with improper judgment in terms of keyword selection, that is, both by conceptual and contextual verification. Keyword-content relation is not properly judged even keywords are not differentiated by any weighted method and no hierarchy is maintained by content representation levels. As a result, we cannot create an alternative option to manipulate amount of keywords to be selected as per our need, nor we can eliminate the mandatory rule of keyword selection. That means, except the stop list, it has to accept everything as a keyword. As a result, the quality of the keyword is decreasing and the quantity is increasing, which increases the amount of retrieval of document so, the precision of the keyword-based search system is decreasing day by day. Here it is trying to modify the keyword selection process because keyword is the backbone of automatic indexing like KWIC and also keyword-based search system. Here it is trying to consider two major experiments made by 1) Baxendale and 2) Swanson and planning to make a new method to select more informative keywords by which it can solve the main problem of the keyword-based search system i.e. "High Recall and Low Precision".

International Journal of Computer Applications (0975 – 8887) Volume 177 – No. 12, October 2019

We can see in the Baxendale's experiment [19, 20, 21, 22] that it had explored both the statistical approaches to automatic selection of index terms (based for example on word frequencies) and the use of word groups, word pairs, contextual associations, and in particular the subject indicating clues of prepositional phrases. That is, she worked on the ward and its nature. She said primarily about the selection of nouns and modifiers i.e. noun-objectives or only nouns. She focuses primarily on frequency counting and contextual measurements, which was semantic, syntactic, or statistical in nature. According to the results obtained, it appears that the semantic association is work between searching and retrieval, syntactic association is work between machine or literature and retrieval and apparently misleading association is work between metal and retrieval. On the other hand, we can see in the Swanson's experiment [23, 24, 25, 26] that how to differentiate significant and non-significant words by applying the weight. Here zero weight meant totally nonsignificant or non-informative words which are not allow selecting as a keyword.

The above two experiments have dealt with the natural merits i.e. grammatical formation and use value of the word (Context Definition). This can be regarded as the sole consideration for promoting the word to keyword. But both this two methods are not singly completed or self-sufficient. Because, three basic features are required for a automatic (machine manipulated) indexing system as like as KWIC indexing i.e. speed, machine dependent, Less human expertise. Swanson's experiment shows that even though weight is imposed on keywords, it relies entirely on human analysis. On the other hand, Baxendale's experiment reaches all kinds of features but there are some principle mistakes in determining the 'significant' words. She said that, generally only Noun can be taken as keywords. She analyzed this by contextual definition rather than content verification. Content analysis shows that each noun used in the sentence gains a different weight. Thus, high and low significant words are also found in the significant words selected by the experiment at Baxendale. By an example it can be explored as below:

I am sitting on the deck of the ship

Here, two nouns exist i.e. 'Deck' and 'Ship' and they should be taken as keyword. But if I judge these by Coate's "Thing-Part-Material-Action" formula then I can see that 'Thing'= 'Ship' and 'Material'= 'Deck'. As 'Thing' are highly associates with the content of documents than others. So, according to keyword-content relation each noun in the above example earns different weight. So their priority of selection as a keyword should not be equal. It is shown by the chart:

Fable 2:	Content	Verification
-----------------	---------	--------------

Hierarchy of Keywords			
Grammatical Form	Keywords	Content Relation	Weight
Noun	Ship	Direct	2
Noun	Deck	Direct(Less than Ship)	1

Therefore, in large perspective this criterion (weight) should have to be taken granted for exact and precise retrieval of document. For this reason of lacking, the 'significant word' selection formula provided by Baxendale was not accepted for automatic indexing while 'non-significant word' identification formula of Luhn was accepted because of its full proof explanation.

Judging by these points, in this paper it has planned a new method called 'Grammatical- Hierarchical Logic' (weighted significant word or weighted keyword), by combining the above two methods. And by this it is trying to get the minimum amount of weighted keyword by modifying the **keyword selection process** i.e. overcoming the **mandatory rule** (stop-list or non-significant word). As a result, the main problem in keyword-based search system may be solved.

It makes this method by its two approach 1) Parts of Speech 2) Used Meaning (Define Context), [considering ontology] which are both use at the stage of 'Parse Page' to properly judge the quality of keywords and make a space to create an option for selection or exclusion of keywords.

Firstly, if considering the step of 'Stop-List' and judge the parts of speech of the words which are included to stop listing criteria then it can easily explain that all kinds of 'articles, pronouns, prepositions, conjunctions, interjections, adverbs, auxiliary verbs' are treated as stop-list and ignored to select as keywords. So, remaining words (Noun, Adjective, and Verb Form) are has to be selected as keywords. I am taking the 'Parts of Speech' approach along with a scientific verification scheme like 'Thing-Part-Material-Action' formula given by 'E.J.Coates' to verify the quality of keywords i.e. how much informative the keyword is. By an example I can show it.

"Determination of the temperature of combustion of coal particles"

Here, Thing=>Coal, Part=>Particles, Material=>Combustion, Temperature, Action=>Determination. Therefore, grammatical expression of these sections is:

Thing=>Coal=><u>Noun Form</u> [All Kinds of Noun Form and Noun Phrase] (Related Directly)

Part=>Particles=>Addition and Attribute of the 'Thing' [Noun Form, Adjective Form, Adjective Phrase, Noun Phrase]

Material=>Combustion, Temperature=>Indirect relation after than 'Thing' [<u>Material Noun</u>, <u>Noun Phrase</u> and other kinds of <u>Noun Form</u>]

Action=>Determination=> [Verb Form, <u>Abstract Noun</u> (only which created from verb), Verbal Phrase, <u>Verbal Noun</u>, <u>Noun</u> <u>Phrase</u>]

Here, it can see that if it is try to verify the keyword-content relation through the Coates's 'Thing-Part-Material-Action' formula, then it shows 'Thing' is most associated with the content of the document. Again, all the words included in the 'Thing' are 'Noun form' and basically the other two grammatical forms i.e. 'Adjective' and 'Verb' includes in 'Part' and 'Action' respectively. So, all it can say that, keywords belong to 'Thing' is *equal* to keywords with 'Noun Form'. On the other side, according to grammatical rule of 'Parts of Speech', 'Noun' takes the highest position than others. Therefore considering the both aspect i.e. degree of content representation and grammatical rule it can easily allot the weights on the selected keywords by the 'Parts of Speech' approach and can make a hierarchical list of weighted keywords by which it can create an option both for selection of highest informative keywords and manipulation of amount of keywords as per our requirement or decision. It will be an automatic, speedy and easy mechanism for selection of most informative keywords against any document without applying any human expertise. It is shown as below:

Table	3:	Keyword	-Content	Relation
-------	----	---------	----------	----------

Hierarchy of Keywords			
Parts of Speech	Keywords	Content Relation	Weight
Noun	Coal (Thing)	Direct	4
Noun	Combustion (Material)	Direct (Less than Coal)	3
Noun	Temperature (Material)	Direct (Less than Coal)	2
Adjective	Particle (Part)	Semi-Direct	1
Verb	Determination (Action)	General	0

Also, it is seen in the example, words with 'Noun Form' not only exist in 'Thing' but are also can exist in other parts of this formula (i.e. Part-Material-Action) as like as 'Noun Phrase', 'Noun Clouse'. So, if it select only words with 'Noun Form' as shown by the example as underlined ones, which represent highest concept coverage than other form of words, then it will be selected as most 'Highly informative' or 'weighted' keywords for searching purpose. Yet it can also say that, various words with 'Noun Form' used in the sentence have not same informative value (shown in the table-2). Therefore, keywords selection process gradually becoming a decision oriented and flexible one. So, amount of keywords will be reduced enough as like as, if quality will be high quantity will must be low, which will achieve precise and exact retrieval of documents itself.

Secondly, context verification should be used in the 'Stemming' step of the keyword selection stage. When performing semantic analysis on keywords by using ontology, that is, selection of synonyms, contextualization of keywords is done, so that, same meaningful synonyms can be selected as compare to the 'used meaning' earned by the keywords against any document. Here, it is trying to use this context verification or 'Used Meaning' method to check the accuracy of the stem word in the stemming step. By taking an example it can be shown as below:

"I am sitting on the <u>deck</u> of the ship"

&

"Ingrid was decked out in her Sunday best"

Here, in the 1st example 'Deck' is stand for "Floor" by its 'used meaning' and in 2nd example 'Decked out' is stand for "**Decorate**" by its 'contextual definition'. If 'Deck; and 'Decked out' is come as a single rooted word as by 'Deck' along with synonyms by ontology then it will happen an improper judgment to keywords according to their context definition, because both keywords has their different 'used meaning' in their own document. Therefore, 1st and 2nd keyword along with their synonyms will be mismatched to 2nd and 1st document respectively. It is shown as below:

Keyword	Document	Remarks
Deck (Floor) [With	I am sitting on the <u>deck</u> of the ship	Match
Synonyms]	Ingrid was <u>decked out</u> in her Sunday best	Mismatch
Decked out (Decorate)	Ingrid was <u>decked out</u> in her Sunday best	Match
[With Synonyms]	I am sitting on the <u>deck</u> of the ship	Mismatch

Table 4: Keyword-Document Relation

It increase recall (irrelevant) of documents at retrieval stage, simultaneously precision will be decreased. It actually demolishes quality or weight of keywords by not verifying context of keywords. So, using context definition or 'used meaning' approach it can modify this system at the stage of keyword selection to restrict (to some extent) the use of stemming words, if possible.

8. COMPARISON

By the comparison of existing and modified method it can show the difference of these two methods in respect of selection of amount of keywords with their informative criteria. Take an example to properly show these things as below.

"I am sitting on the deck of the ship"

Here, taking all the keywords to try to explain their synonymous words which are considered as its alternative searching option for better representation of documents for user's benefit.

Table 5:	Comparison
----------	------------

	Existing Method (with ontology)	Grammatical- Hierarchical Logic Method (with ontology)
K	Stop Listing : Sitting, Deck, Ship	Stop Listing (Weighted) : 1) Ship 2) Deck, (Ignore 'Sitting' as low Weight & Verb form)
	Stemming :	Synonym:
	Decked out (various 'used meaning' as below)	Ship=> Boat (used meaning) Vessel, Craft, Ferry, Yawl
Б	Decorate, Bedeck,	etc.
E	Adorn, Ornament, Trim, Trick our, Garnish, Cover, Hang, Festoon,	Deck=> Floor (used meaning)
	etc.	Ground, Storey, Tier, Level,
Y	Ship out, Shipping (various 'used meaning' as below)	Mezzanine, Entresoi.
	Separate, Break, Depart, Go, Quit, Split, withdraw etc.	
	Freight, sailing, Passenger, Steam	
	Sit up (various 'used	

	meaning' as below)	
W	Affect, Amuse, Attract, Concern, Involve, Engage, Please, Tempt etc.	
	Synonym:	
0	Sitting=> Sit (used meaning)	
0	Take a chair, Perch, Flop, Flump, Rest	
	Deck=> Floor (used meaning)	
R	Ground, Storey, Tier, Level, Mezzanine, Entresol.	
	Ship=> Boat (used meaning)	
	Vessel, Craft, Ferry, Yawl etc.	
D		
S		

The comparison shows that, by existing method, this system creates some unnecessary and some mismatching keywords against documents. They are unnecessary in the sense that, they are not properly represents content of documents (For keywords with 'Verb form') and mismatching in the sense that, they have different 'used meaning' or 'context definition' in different documents in respect of the context analysis (For Stemming words). So, for both of the cases amount of keywords are increasing directly or indirectly and it causes high recall and low precision with irrelevant documents. On the other hand, it is seen that keywords are reducing by the new 'Grammatical-Hierarchical Logic' method to a notable amount, side by side keyword selection process is becoming systematic, automatic, easy and decision oriented not mandatory based one. Here also keywords are hierarchically arranged to select than existing method. So by the new modified method of keywords selection process keywords will become more authentic, weighted, informative, qualitative, composed and least enough.

9. UP-GRADATION

Keyword searching system may be up-graded by using this 'Grammatical Logic' method in some aspects:

- 1) Proper judgment of keywords by its content representation.
- 2) Automatic (Scientific) and easy method of keyword selection.
- 3) Weighted keywords to create an option for selecting and excluding keywords as needed.
- 4) Increase quality of keywords which results high precision of retrieval
- 5) Restricted use of stemming words, if possible to reduce number of keywords.

10. DEGRADATION

This method also has drawbacks but it would be overcome easily. These are:

- By the automatic method of keyword extraction all the time well informed keywords would not be selected because keywords are not always well informed or it has dual meaning. So human intuition is required here.
- 'Used meaning' approach includes a large lexicon work for preparing controlled vocabulary against 'used meaning' of keywords and its synonymous words.
- 3) Here a controlled vocabulary is to be maintained so it wouldn't be a freely searching option.

In every case there should have some exception which are not treated by natural way. So these are not taken as limitation or bother to create a new method. Even some limitations are created due to merits of a particular method. So it should take as a working process indeed.

11. CONCLUSION

Just as the investigator received their valuable clues from the place of incidence, it is possible to reduce the volume of keywords from the keyword selection stage only. And this can only be done by analyzing the ability of keywords to serve information. This should be main and only one criterion for keyword selection purpose. In other words, the quality of the keywords i.e. keyword-content relation should be verified every time, otherwise the quantity of keywords will increase and the performance of the retrieval system will decrease. Therefore, if it is try to improve the performance of the keyword-based search system then it should have to be built by the most informative keywords which will be selected only by the verification of keyword-content relationship, not by any mandatory rule (stop-list). This paper shows how to properly judge keywords and how to quantify keywords by selecting high-quality and weighted keywords by landing an automatic, speedy and easy mechanism without using human expertise, so that relevant and precise documents can be retrieved as per user requirement. Therefore, modification made by 'Grammatical- Hierarchical Logic' method even after using 'Ontology' is very much needed to make retrieval system like keyword-based search system more users friendly.

12. REFERENCES

- [1] Fischer, Marguerite, "The KWIC index concept: a retrospective view", April 1966.
- [2] Kaki, Mika, "fKWIC: Frequent based Keyword-incontext index for filtering web searching results", Finland: University of Tempere. 2005
- [3] Mahakuteshwar, H. Y., "Altered keyword in context (aKWIC) indexing", 1980.
- Thomas R Gruber, "What is an ontology?", http://wwwksl.stanford.edu/kst/what-is-an-ontology.html (last checked at 22-06-2019)
- [5] Aristotle, "The Complete Works of Aristotele: The Revised Oxford Translation", volume 1. Princetron University Press, 1984.
- [6] Guarino, N., "What is an ontology?" https://iaoa.org/isc2012/docs/Guarino2009_What _is_an_Ontology.pdf (last checked at 22-06-2019)

- [7] Thomas R. Gruber, "Toward principles for the design of ontologies used for knowledge sharing", International Journal Human-Computer Studies, (43):907–928, 1995.
- [8] Ontology (information Science) Wikipedia. https:// en.wikipedia.org/wiki/Ontology_(information_science)
- [9] Kumar, Sajendra, Rana, Ram Kumar, Sing, Pawan, "Ontology based semantic indexing approach for information retrieval system", International journal of computer applications (0975-8887), vol.4, no.12
- [10] Wu, Dan, Wang, Hui-Lin, "Role of ontology in information retrieval", Journal of electronic science and technology of china. Vol. 4, no. 6, Jun. 2006.
- [11] Munir, Kamran, Anjam M. Sheraz, "The use of ontologies for effective knowledge modeling and information retrieval", Applied computing and informatics. Vol. 14, no. 2, pp.116-126, 2018
- [12] Tulsi R. Lakshmi, R.H. Goudar, Rao M. Sreenivasa, P.D. Desai, "Domain ontology based knowledge representation for efficient information retrieval", Journal of information system and communication, 2012, ISSN: 0976-8742.
- [13] Luhn, H.P. "Keyword-in-contect index for technical literature (KWIC Index)", presented at american chemical society, Division of chemical literature at Atlantic city, N.J. 14 sept. 1959, Rept. No. RC 127, International business machines corp., York-town heights, N.Y. 1959, 16p. Also in Amer. Documentation 11, 288-295 (1960).
- [14] Luhn, H.P. "The automatic derivation of information retrieval en-codements from machine-readable texts," International business machines corp., Yorktown heights, N.Y. 1959, 9p. Also,min A. Kent, "Information retrieval and machine translation", Pt II, 1961, pp. 1021-1028.
- [15] Sedano, John Michael, "Keyword-in-context (KWIC) indexing: Background, statistical evaluation, pros and cons, and applications", university of pittsburgh, 1964
- [16] Helbich, Jan, "Direct selection of keywords for the KWIC index", Information Storage and Retrieval, vol. 5,no. 3, pp. 123-128, Oct 1969.
- [17] Fischer, Marguerite, "The KWIC index concept: A retrospective view", Journal of the association for information science and technology, April 1966, https://doi.org/10.1002/asi.5090170203 (Last checked at 04-09-2019)
- [18] Kraft, Donald H., "A comparison of keyword-in-context (KWIC) indexing of titles with a subject heading classification system", Journal of the association for information science and technology, Jan. 1964, https://doi.org/10.1002/asi.5090250209.(Last checked at 04-09-2019)
- [19] Baxendale, P.B. "Automatic processing for a limited type of document retrieval system", in H.P.Luhn [ed.]. "Automation and scientific communication, short papers, Pt. 1", 1963, pp. 67-68.
- [20] Baxendale, P.B. "An empirical model for computer indexing" in "Machine indexing", American U. 1962, pp. 207-218.

International Journal of Computer Applications (0975 – 8887) Volume 177 – No. 12, October 2019

- [21] Baxendale, P.B. "Machine-made index for technical literature: an experiment", IBM J. Research and development 2, 1958, pp. 354-361.
- [22] Baxendale, P.B. "Man-computer indexing: functions, goals, and realizations", in "Joint man-computer indexing and avbstracting", Mitre, ss-13, 1962, pp. 61-73.
- [23] Swanson, D.R. "Automatic indexing and classification" preprint, Nato Advnaced study institute on automatic document analysis, Venice, 7-20 july 1963, 4 p.
- [24] Swanson, D.R, "Autometic title analysis", paper presented at NATO advanced study institute on automatic document analysis, Venice, 7-20 july 1963 (substantially excerpted from Montgomery and Swanson, "Machine-like indexing by people")
- [25] Swanson, D.R. "The nature of multiple meaning", in H.P. Edmundson [ed]. "Proceedings of the national symposium on machine translation", 1961, pp. 386-393.
- [26] Swanson, D.R. "Research Procedure for automatic indexing", in "Machine indexing", American U, 1962, pp. 281-304.