Utilization and Application of Weighted Keyword in Retrieval

Sandip Ghosh

Abstract: Keyword improvements must be addressed to address the core issue of ‘Keyword-Based Search System’ i.e. “High Recall and Low Precision”, which will in turn lead to systematic modification of the ‘KWIC indexing system. And this improvement can only be done by imposing a weight that is possible through the ‘Grammatical-Hierarchical Logic’ method. So, in this paper it is trying to establish a modified KWIC named ‘Key-Word Weighted-In-Intra Contextual-Content’ on the one hand, and on the other hand trying to draft a mechanism of an advanced ‘Search Engine’ outline with the modified indexing method named ‘KWWIICC’, by which most relevant and precise document will be provided as per the users requirement.

Keywords: Weighted Keyword, KWIC, KWWIICC, Search Engine, Semantic Web, Retrieval

I. INTRODUCTION

The weighted keyword can be regarded as the only option to solve the main problem of the ‘Keyword-Based Search because, not only context verification is important for System’ or ‘KWIC indexing system’ i.e. “High Recall and Low Precision”. Keyword selection, content analysis is also required. This is the main background of the selection of weighted keyword by ‘Grammatical-Hierarchical Logic’ method (1). Document retrieval (9) can be much more precise by using the weighted keywords. Therefore, it is important to consider that weighted keyword will be the only deciding factor in ‘KWIC indexing’ or ‘Keyword-Based Search System’.

II. THOUGHT OF TOPIC

The practical part is equally justified with the theoretical part of one method. Because of the combination of these two, one of the methods is complete. Therefore, there is a need to explain the practical side of the ‘Grammatical-Hierarchical Logic’ method (1) in a scientific way. The keyword's weight can be used to determine the importance or the internal value of that keyword, thus creating an opportunity to reduce the amount of keyword. But, the selection method of weighted keywords i.e. Grammatical-Hierarchical Logic method cannot be completed as long as a new or modified indexing system and an advanced mechanism of document retrieval are being expressed. Therefore, practical evaluation of the weighted keyword in document retrieval has become very important.

III. LITERATURE REVIEW

The main problem of “Keyword-Based Search System i.e. “High Recall and Low Precision” can only be solved by minimizing the amount of keywords. And this can only be done by improving the quality of the keyword like ‘Weighted Keyword’. As a result the precision will increase (1). By the 1950, computer usage in data storage began to increase. In the year 1961, H.P.Luhn invented the data storage system using the keyword, called Keyword In-Context. It is an automatic indexing system. It is an indexing created by words, where each word is released with its own list of strings (2).

In KWIC Indexing, there are many keywords. So, for short search huge results are come whose access becomes so difficult. So, in 2005, Mika Kaki invented ‘fKWIC’ where filtered interface can be presented by the most used keywords (3).

Many options have been used for modification of traditional KWIC indexing system, such as ‘AKWIC’. Here output format of KWIC is modified for getting a smoother system indeed (4).

So, over time there has been an attempt to modify the KWIC indexing system. Someone has modified the output format; some have made the filtering of the search interface. That is, by maintaining the structure of KWIC indexing system, it is undergoing systematic changes. ‘Intelligent semantic search engines’ are introduced for overcoming the problems of traditional search engines i.e. difficult to retrieve and takes long time, and targeted to give required information within small times and with high accuracy (11).

The context guided information retrieval process is extraction of semantic keyword and clustering automatically generation of new, augmented queries. The result is semantically ranked, again, using context (12).

Semantic web technologies are a crucial role to retrieve meaningful information intelligently. These are called generically search engine (13).

Traditional ‘Keyword-Based Search’ system is lacking of semantic. To overcome this issue it will be considered the context (concept) using semantic search terms to index the search engine (14).

Ontology queries with keyword search will be a better querying mechanism for information retrieval. It is more dependable to study semantic model and query for semantic matching conditions (15).

So, the main purpose of the current modern search engine is accuracy in document retrieval but limited to minimum time usage. It will only be possible if ontology based semantic web technology will act in front. It will support the user’s query to match the needed documents accurately.

IV. OBJECTIVE

This paper describes how the weighted keyword can be useful in document retrieval. How to use this ‘Grammatical-Hierarchical Logic’ method successfully with a new search mechanism, how to simplify and precisely retrieve documents, how to modify the indexing method, how to save time in retrieval.

Revised Manuscript Received on March 05, 2020.
Sandip Ghosh, Department of M.Lib Burdwan University, West Bengal

DOI: 10.35940/ijitee.E2849.039520

Published By: Blue Eyes Intelligence Engineering & Sciences Publication
Utilization and Application of Weighted Keyword in Retrieval

At last, how to make this method more user friendly indeed. So, the main objectives of this paper are:
1) Modification of the format of KWIC indexing system
2) Introduce new search mechanism (Search Interface /Engine) for retrieval of documents

V. SCOPE

The weighted keyword, on the one hand, will modify the format of the KWIC indexing system on the other hand improve the quality of document retrieval through a new search engine. As a result, the user can get the precise document as per his requirement and users friendly search engine can be created by a strong indexing system.

VI. METHODOLOGY

For the practical application of the weighted keyword in retrieval, a modified format will be created in contrast to the current KWIC indexing format, which will have the new attribute and become a completed and improved one. Secondly, a new search engine will be created by expressing an advanced mechanism, which is much easier to use in comparison, with better filtering element added and helpful to retrieve precise document.

VII. DETAILS OF PAPER

Weighted keyword selection is justified as a way to solve the main problem of ‘Keyword-Based Search System’ i.e. “High Recall and Low Precision”. The background of Keyword-Based Search System is built on the KWIC indexing system (2). Thus, systematic changes in keyword selection will naturally result in modification of the KWIC indexing system. In the case of automatic indexing, the KWIC indexing (5, 6, 7) plays an important role where the complete indexing system is established by its key component keywords. For this reason, the modification of the KWIC indexing system can only be done through the improvement of keywords. Therefore, the creation of a modified KWIC indexing system (3, 4) can be attempted by using the ‘Grammatical-Hierarchical Logic’ method in this paper. It may also represent an advanced automatic indexing system indeed. This is illustrated by the example below:

Main Example:

“I am sitting(A) on the deck(N) of a fine(N) ship(N)”

Keywords (according to weight) Used Meaning
*Deck (Noun Form) Floor
*Ship (Noun Form) Ship
*Fine (Adjective Form) High Quality
*Sitting (Verb Form) Sit

Keywords Used Meaning
Deck Floor, Recorder, Pack
of cards, Knock Ship, Transport, Sale
Fine High Quality, Good,
Thin, Charge, Punish
Sitting Sit, Period of Time,
Present

Other Examples:
I am playing the deck(N) at every Sunday
I have a deck(N) for playing purpose in train
Lisa threatened to deck(V) her if she didn’t stop filming
Ship(V) the wounded soldiers at their home

The cellular phone is expected to ship(V) at about $500
Twenty pieces of music is a much to take in at one sitting(N)
A special sitting(N) of Basil don Magistrate
The registration of the sitting(A) members
The resignation of the sitting(A) members of the parliament
This was a fine(A) piece of film-making
A fine(A) nylon thread
She had certainly fined down(V)- her face was thinner
A parking fine(N)
She was fined(V) $100 for driving offences

Synonyms of ‘Used Meaning’ of Keywords:

<table>
<thead>
<tr>
<th>Used Meaning</th>
<th>Synonym (Sam as ‘Used Meaning’)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Floor(N)</td>
<td>Ground, Storey, Tier, Level,</td>
</tr>
<tr>
<td>Mezzanine, Entresol</td>
<td>VCR, Video, Tape</td>
</tr>
<tr>
<td>Recorder(N)</td>
<td>Pack, Playing Cards</td>
</tr>
<tr>
<td>Pack of cards(N)</td>
<td>Bang, Hit, Strike, Ram, Jolt</td>
</tr>
<tr>
<td>Knock(V)</td>
<td>Vessel, Craft, Boat</td>
</tr>
<tr>
<td>Ship(N)</td>
<td>Convey, Carry, Transfer, Move, Shift, Send, Deliver, Ferry</td>
</tr>
<tr>
<td>Transport(V)</td>
<td>Deal, Transaction, Disposal</td>
</tr>
<tr>
<td>Sale(V)</td>
<td>Take a chair, Perch, Flop, Flump, Rest</td>
</tr>
</tbody>
</table>

*Pack of cards, Knock |

*Ship, Transport, Sale |

*Fine, High Quality, Good, |

*Thin, Charge, Punish |

<table>
<thead>
<tr>
<th>Sitting</th>
<th>Period of Time, Present</th>
</tr>
</thead>
</table>

Extra Examples:
He lay on the ground
A three storey building
A tier of seats
The level of basement is equal for each house
A mezzanine floor
It is a good entertainment design
A pack of cigarettes
I have a nice playing card
This VCR is a new one
She could post videos of the event on internet
They put four songs on tape
He began to bang the table with his fist
Marius hit him in the mouth
He raised his hand, as if to strike me
He rammed his stick into the ground
A surge in the crowd behind him jolted him forwards
We go to the river with a big vessel
The craft is made by gold
It is a fishing boat
Pipes were laid to convey water to the house
Ticks can carry a nasty disease which affects humans
He intends to transfer the fund’s assets to the treasury
She moved to the door

Published By:
Blue Eyes Intelligence Engineering & Sciences Publication

Retrieval Number: E2849039520 / 2020©BEIESP
DOI: 10.35940/ijitee.E2849.039520

1731
A team from Power Company came to shift the cables. 
I can send a message on behalf of you 
The product should be delivered on time 
The goods should ferry on time 
The cards were dealt for the last hand 
Transaction of gold bar 
Dispose the goods 
Please take a chair Mr. Bose 
A herring gull perched on the rails 
His blond hair flopped over his eyes 
I flumped back into bed 
He needed to rest after the feverish activity 
The time is 11.30 p.m. 
The President is elected for a single five-year term 
He had long periods of depression 
Bicycle hire for the duration of your holiday 
Keep abreast of current events 
Opponents of the existing political systems 
He is a class player 
Ram is one of the world’s most eminent scientists 
He is my superior officer 
The lorry was in excellent condition 
It is her superb performance 
She paid tribute to their magnificent efforts 
The hotel offers first-class accommodation 
I think you look first-rate 
His skinny arms not acted properly 
Her slim figure is perfect for modeling 
Her slender neck is so beautiful 
He held up his bony figures 
Her body is lean, taut, and athletic 
She was slight and delicate looking 
The skeletal plot for a novel 
She was so emaciated she could hardly stand 
They were faced with legal fees of $500 
Land could be sold for a high price 
The reduction of trade barriers and import tariffs 
Police forces receive 49% of their funding via a levy on the rates 
A tax of 2% was levied on all cargoes (*) 
Motorway tolls 
We should go to Andaman but we can’t afford the air fare 
He chastised his colleagues for their laziness 
High-spending councils will be penalized 
He was castigated for not setting a good example 
He was reproved for obscenity 
Officers were reprimanded for poor work 
In the example above it is seen that, three types of documents can be published by any one keyword these are namely direct document, indirect document and additional document, who are being expressed by the ‘own used meaning’, ‘other used meaning’ and ‘synonymous word’ respectively. These types of documents can be called Umbra, Penumbra, and Aliens respectively in the language of information science. But in all cases the contextual definition remains the same, so here such a relation of the context with the keyword can be called ‘Intra Contextual Relation’.

Table 1: Sample, KWWIICC Index

<table>
<thead>
<tr>
<th>KW</th>
<th>UM</th>
<th>VUM</th>
<th>SYNONYM</th>
<th>DOCUMENTS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Deck</td>
<td>Floor</td>
<td>I am sitting on the deck of a fine ship</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ground</td>
<td>He lay on the ground</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Storey</td>
<td>A three storey building</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tier</td>
<td>A tier of seats</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Level</td>
<td>The level of basement is equal for each home</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mezzanine</td>
<td>A mezzanine floor</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Entresol</td>
<td>It is a good entresol design</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Recorder</td>
<td>I am playing the deck on every Sunday</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>VCR</td>
<td>This VCR is a new one</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Video</td>
<td>She could post videos of the event on internet</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tape</td>
<td>They put four songs on tape</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Utilization and Application of Weighted Keyword in Retrieval

Pack of cards | I have a deck of cards for playing cards
Pack | A pack of cigarettes
Playing card | I have a nice playing card
Knock | Lisa threatened to deck her if she didn’t stop filming
Bang | He began to bang the table with his fist
Hit | Marius hit him in the mouth
Strike | He raised his hand, as if to strike me
Ram | He rammed his stick into the ground
Jolt | A surge in crowd behind, jolted him forwards

Again, the application of the Grammatical-Hierarchical Logic method is the external appearance of the document retrieval process. Since the Grammatical-Hierarchical Logic method is able to solve the major problem of ‘Keyword-Based Search System’, and also able to add the remarkable benefit to the document retrieval process, therefore, a better search engine (17,20) can be built by using the ‘Grammatical-Hierarchical Logic’ method. The features of this search engine are as follows:

1) Advanced search featuring just one search point (Keyword)
2) Shows the results as per the degree of keyword-content relationship (High to Low)
3) Content will matter but limited to the contextual definition.
4) Ontology will be used to define context.
5) User intend oriented

The following is a structural framework of the search engine:

Fig 2: Mechanism of Retrieval

All the search engines are trying to be the most popular by giving closest result to what the searchers are looking for. So the strategy is to provide the relevant content. Keeping this objective in mind the new search engine mechanism has been created by amalgamation of both the content and context analysis, so that the weighted keyword can be selected for indexing by the ‘KWWIICC’ indexing system. Therefore, due to the selection of keywords through the imposition of weights, there is no need for the complicated process of weighting or ranking of documents at the time of document retrieval. But, at the time of retrieval of document, the priority should be set according to the grammatical form to reflect the weight of the keyword which represents the ranking of document according to user’s need or what they looking for.

VIII. RESULT

Describing the usefulness of weighted keyword by a handful example here it is observed that in keyword-base search system preciseness can only be achieved through the weighted keyword. So, the result of this above thematic and experimental study is given below:

1) Obtain a modified “KWIC” indexing system called “Key-Word Weighted-In-Intra Contextual-Content” (KWWIICC) which based on weights i.e. degree of keyword-content relationship but limited to contextual definition.
2) Get a combined indexing method which has the nature of both ‘KWIC’ index and ‘Related Index’.
3) Get a user intends oriented ‘Search Engine’ (Simple) featuring of advanced search.
4) Expel the complicated process of weighting or ranking of documents at the time of retrieval.
5) All the benefits of ‘Semantic Web Search Engine’ can be found within ‘Simple Search Engine’.

IX. FUTURE PROVISION

Although, above search is based on the advanced search but it is featuring just one search point i.e. keyword. There is no other filtering option available like other advance search. This unique and advance feature will help the search engine to become a simple search engine, or general search engine like Google, Yahoo, etc. and may be able to provide the benefits like other ‘Semantic Web Search Engine’ (16,18,19), especially the same type of document will be together, not separated and weight (keyword-content relation) will prevail.

Fig 3: Search Engine
X.  MERIT
Utilization and application of ‘Grammatical-Hierarchical Logic’ method can provide the following benefits.

1)  Obtain Modified ‘KWIC’ or ‘KWWIICC’ indexing system
2)  Different documents can be linked to.
3)  Context can be linked to content.
4)  Easy-to-use advanced search engines can be created.
5)  User Requirements are considered the main consideration.
6)  Document retrieval - this is low, composite and level of content wise (high to low)
7)  Save the search time of the user

XI.  DEMERIT
There are also some demerits which are occurred by the nature of this method. These are as follows:

1)  The search mechanism has become complicated to use relative indexes.
2)  Users’ may be confused to determine relationship with penumbra, alien type document.

These are first time use problem. Repetitive work can help to overcome all these very easily.

XII.  CONCLUSION
The weighted keyword extracted by ‘Grammatical-Hierarchical Logic’ method, on the one hand, is capable of modifying the keyboard selection process while on the other hand it is capable of solving retrieval problem i.e. ‘High Recall and Low Precision’. Therefore, both indexing and searching can be improved and refined by this method. The use of the ‘Grammatical-Hierarchical Logic’ method gives the KWIC indexing system a new look, adding new features, which in turn helps to link-up huge databases in the modern information era. On the other hand, ‘Grammatical-Hierarchical Logic’ method emphasizes the content of the document but limited to the contextual definition for the retrieval purpose which in turn helps to prioritize the intend of the user. So, in the current information exploration era, it is necessary to say that the user-intended, intra-connected, and precise documents can only be obtained through weighted keywords which are extracted by ‘Grammatical-Hierarchical Logic’ method.

REFERENCE

AUTHOR PROFILE
Sandip Ghosh, M.Lib Burdwan University, Persuing Ph.D from WBUT, Publication from “International Journal of Computer Applicastions"