IJCSE International Journal of Computer Sciences and Engineering Open Access
Research Paper Vol.-6, Issue-9, Sept. 2018 E-ISSN: 2347-2693

Appraisal of MLIR systems using Weight Based Precision Metrics

Pothula Sujatha[1*], Prasad Koti T[2], Dhavachelvan P[3]

[1, 2,3]Department of Computer Science, Pondicherry University, Puducherry, India.

[*]*Corresponding Author: spothula@gmail.com*

Available online at: www.ijcseonline.org

Accepted: 23/Sept./2018, Published: 30/Sept./2018

Abstract— Multilingual Information Retrieval System (MLIR) allows users to provide queries in one language and extract the relevant content in multiple languages. Appraising the quality of these systems is a promising task. A wide variety of metrics are available for estimating the performance of IR systems, Precision and Recall are considered as the basic measures among them. However, less number of metrics is available in the literature to analyze the performance of MLIR Systems. This paper demonstrates the significance of MLIR systems when the retrieved documents are in various languages and the weights assigned by the user based on his preference languages. This is achieved by comparing the performances of IR and MLIR using the proposed weight based Precision oriented metrics. In addition, four essential parameters of the retrieval systems are considered to compare the significance of the proposed metrics with traditional metrics. The analyses of these metrics demonstrate positive and promising results. Statistical Analyses are also performed to show the importance of the proposed metrics. Thus we can conclude that weight based precision oriented metrics plays a vital role in MLIR domain area.

Keywords— Average Precision, Information Retrieval, MLIR, Precision, Normalized Precision, P@k

I. INTRODUCTION

IR system identifies the significant information from a huge collection of information resources depends on the interest of the investigator Christian Fluhr et al. IR takes part in the process of dealing with the representation, storage, organization of and access to information items. Information items contain documents, web pages, online catalogues, structured records and multimedia objects. IR task continues to depict awareness as the information repositories amplify Toyin Enikuomehin et al. (2013). Fundamentally, MLIR Dan Wu et al. (2013) System is a part of IR System. MLIR System applies queries in one language and dig out the relevant contents in multiple languages. Hence, MLIR Systems are used to allow the user to provide query request in one language and dig out the resultant documents in two or more languages. Furthermore, MLIR describes the ability to practice a query for information in any language, explore a collection of objects, including text, images, sound files and send back the most germane objects, translated into the user's language based on the requirement. In MLIR System, Query translation Rogati, et al. (2004) is based on dictionary based method Cleverdon CW et al. (1966). Translation of query and document languages is the two methods to describe MLIR systems. In most cases, researchers tend to choose translation of query because of the reason that it is more efficient than translation of documents. Using web search engine it can personalize query in one language and produce result sets in more than one language. By Hull and Grefenstette, the five benefits of MLIR includes Hull D et al. (1996) are: 1. Monolingual IR, where the query and retrieved documents are in English Davis et al. (1996). 2. IR process can be performed on a collection of documents in numerous languages, where the result sets are retrieved in two or more languages with queries given in one language only. 3. IR on a monolingual document collection can be queried in various document languages. 4. IR on a multilingual document collection can submit queries in a variety of languages and retrieve documents in various languages. 5. Information Retrieval on multilingual documents, where more than one language possibly present in a single document.

Parmatma Yadav et al. (2014) performs the surveying IR System's performance is the essential issue from the decades. Henceforth, the present assessment depended on the model actualized in Cranfield Project. For this model, assessment test collection contains a collection of documents, set of questions and assembled relevant documents reports for every inquiry inside the recovered archives. Assessment depends on the comparison lied between list of documents that the system retrieved and the list of really relevant documents by the user interests Ellen et al. (1995). Mangala et al. (2016) presented a detailed review of various information retrieval systems.

The importance of this paper is based on the following scenarios. If the user wants to search the documents in his/her native language and the retrieval system is retrieving documents in more than one language. For example, if user wanted to search documents in Telugu (user's native language) and chosen Telugu as query language and the systems is retrieving documents in English and Telugu albeit the query language is Telugu. In this type of retrieval system, evaluation of systems performance must be done with the retrieved languages i.e. English and Telugu. At this juncture user has to give weights to the retrieved languages and predict the performance of the retrieval system based on his preferences. Goodness of our proposed metrics are useful to verify that which MLIR systems are good in retrieving the documents effectively in preferred languages, therefore the user can use that particular system based on his/her own query languages.

In light of the audited works there is less number of measures accessible to assess these frameworks. Here in IR, assessment estimation is the major procedure of contrasting user query and a gathered arrangement of predefined reports. As of now the Information Evaluation Model is depended on the order of ranked items. As indicated by rank based framework numerous measures are most likely characterized that prompts the upgrade in the execution of IR Models. P Sujatha et al. (2011) have proposed Precision at "k" for evaluating MLIR frameworks. Sakai Sakai (2006) has proposed the IR measurements for extricating very significant reports and Sakai (2004) establishes relationship of the positions among normal P, R-Precision utilizing NTCIR Data. Vu and Gallinnari (2005) summed up AveP for taking care of reviewed importance. P Sujatha et al. (2011) assessed the execution of MLIR System over IR System. Ravisankar et al. (2015) have proposed a factual procedure for surveying MLIR frameworks. R Hackl et al. (2004) have confirmed different indexing and numerous article connections as Multilingual Retrieval Experiments with MIMOR at the University of Hildesheim. Chetana et al. (2011) indicates Normalized Distance Measure for Evaluating MLIR Merging Mechanisms. The user query and the resultant report could grow to give a base to advance the recovery review and accuracy rate that has been uncovered in Yuemin Wang (2014). From the above works one can obviously comprehend that weight of the favoured dialect idea was not utilized by single creator. This paper uncovers the significance of weights given to the languages when we recover data from the MLIR framework. The key commitment of this exploration work is the advancement and trial affirmation of an arrangement of hypothetically grounded measurements of MLIR frameworks.

The structure of the remaining paper is described as follows. Section II explains the evaluation of the IR System based on existing metrics and Section III provides the detailed description of enhanced weight based MLIR Metrics. Section IV shows the experimental results. Section V shows the statistical analysis of experimental results and finally a few concluding remarks is declared

II. EVALUATING THE PERFORMANCE OF EXISTING IR SYSTEM

An important issue in IR is mostly to achieve maximum Precision. Precision is the standard metric and Average precision, normalized precision, etc are called as variants of the precision metric. In this paper, we have been discussing the IR metrics based on precision sometimes called as precision oriented metrics or its variants. Notion of relevance can be used for all evaluations. The basic and traditional approach is based on the binary relevance and retrieval result which is either relevant or not. Hence, based on the relevance, evaluation is segregated as 1.Unranked (Set) Based Evaluation 2.Rank Based Evaluation.

2.1 Unranked (Set) Based Evaluation

Precision, this is the primary and common measure and its formula is specified in Eqn. (1). This is the original metric defined in IR System which accomplishes useful evaluation of IR.

Precision: It Van Rijsbergen et al. (1979) is the fraction of retrieved documents extracted based on the query provided by the user.

$$P_{IR} = \frac{T_p}{T_p + F_p} \qquad (1)$$

where,'Tp' - relevant item retrieved and Tp+Fp - total retrieved items.

2.2 Rank Based Evaluation

The exactness for recovered archives is portrayed by the quantity of genuine positives i.e. the quantity of things effectively recovered as having a place with the positive archives isolated by the aggregate number of reports recovered as having a place with the positive records i.e. the aggregate of genuine positives. False positives, which are things erroneously recovered as having a place with the pertinent reports.

1. Average Precision (AP): AP EmineYilmaz et al. (2008), Tetsuya Sakai et al. (2008) is the average of precisions that is mathematically determined on every extracted relevant document in the form of sequence of ranks as shown in the Eqn. (2).

$$AP = \frac{1}{R}\sum_{1 \le r \le k} isrel(r) \times \frac{count(r)}{r} \qquad (2)$$

where, 'R' denotes the total number of known relevant documents and count(r) denotes the number of relevant document within the top 'r' document of the ranked output.

Clearly, it is precision at rank r $=\frac{count(r)}{r}$, isrel(r) =1 if document at rank r is relevant, otherwise zero. Here 'k' is the size of ranked output. Average Precision is fundamentally used to decide the average execution for a given question. Consider the average over a wide range of inquiries to locate the normal viability of the recovery framework.

2. Precision@K: Olivier Chapelle, et al. (2010) is the precision after 'k' documents have been retrieved. Precision@ k is the percentage of relevant documents in the first 'k' positions and its representation is given in Eqn. (3).

$$P@K = \frac{1}{k}\sum_{i-1}^{k} r_i \quad (3)$$

where, 'k' is the cutoff of retrieved documents and r_{l_i}=1 if document at j[th] position is relevant otherwise zero.

Calculating precision at fixed low levels of retrieved results is vital in IR systems. It has the advantage of not requiring any estimate of the size of the set of relevant documents. The total number of relevant documents for a given query has a strong effect on precision at 'k'.

3. Normalized Precision (NP): NP is the AP, it is determined for all possible cut-off points R. Korfhage (1997).

$$NP_{IR} = \frac{1}{N}\sum_{j-1}^{N} P_j \quad (4)$$

where, 'N' states the number of resultant documents in the collection and 'P_j' implies the precision at cut-off of 'j'th documents. The other metrics that are concentrated in this paper are not based on specific cut-off ranks, but in a sense that they calculate the performance of the retrieval systems over the entire document collection.

III. SPECULATIVE EVALUATION OF PROPOSED MLIR SYSTEM

In MLIR frameworks, as we realize that the inquiry gave by the client in one dialect and the reports separated are in numerous dialects. So the above examined measurements are insufficient for the evaluation of the MLIR Systems since these measurements are ideal to decide the execution of monolingual IR frameworks. In this exploration work, we are giving the question in one dialect and the records extricated are in various dialects by utilizing the accompanying web search tools. All the standard Search motors like Yahoo, Google, Bing giving multilingual pursuit stage to area based result as indicated by need of dialect. So in the proposed work, a quality is allocated to every language utilized as a part of the recovery procedure as weight. Standardized quality is utilized to decide the estimation of weight. Taking into account the given inclinations, language weights are relegated consistently in the scope of 0 to 1. In this manner, in view of the weights we have proposed the Precision situated measurements for surveying MLIR System's execution.

The following sections discuss the Weight based Precision Oriented Metrics:

1. Precision(P_{MLIR})**:** Precision is defined as the fraction of relevant documents that are extracted with weight based on the user query in MLIR System as verified in Eqn. (5).

$$P_{MLIR} = \frac{\sum_{i=1}^{n} D_{Rel_{li} \times W_{l_i}}}{\sum_{i=1}^{n} K_{l_i}} \quad \textbf{(5)}$$

Where $'D_{Rel_{l_i}}'$ is relevant retrieve document in different languages$'W_{l_i}'$ is the weight of the different languages and $'K_{l_i}'$ total retrieved documents in different languages.

2. Average Precision (AP_{MLIR}): It is the term which defines the average of weight based precision that is measured depending on each relevant document in the form of sequence of ranks as showed in Eqn. (6).

$$AP_{MLIR} = \frac{1}{N_{Rel_{l_i}}} \sum_{D_{l_i} \epsilon Rel_{l_i}} Rel(Rank_{D_{l_i}}) \times \frac{count(Rank_{D_{l_i}}) \times W_{l_i}}{Rank_{D_{l_i}}} \quad \textbf{(6)}$$

where $'N_{Rel_{l_i}}'$ is the total number of known relevant document,$'Rank_{D_{l_i}}'$ is the Rank of different language, $'Count(Rank_{D_{l_i}})'$ is the number of relevant document within the top ranked $'Rank'_{D_{l_i}}$document of ranked output. It is Clearly Precision at $'Rank_{D_{l_i}}'$ with weight is

$$\frac{count(Rank_{D_{l_i}}) \times W_{l_i}}{Rank_{D_{l_i}}}$$

3. Precision@K: Precision@k is the weight based precision which is calculated after $'K_{l_i}'$ documents in different languages have been filtered and extracted in MLIR System and it is presented in Eqn. (7).

$$(P@K)_{MLIR} = \frac{1}{K_{l_i}} \sum_{i-1}^{K_{l_i}} r_{l_i} \times W_{l_i} \quad (7)$$

Where $'K'_{l_i}$is cut-off of set of retrieved documents in MLIR System, r_{l_i}=1 if different language documents retrieved is relevant otherwise 0 and $'W_{l_i}'$ is the weights assigned by the user to those languages.

4. Normalized Precision (NP_{MLIR}): It is the term used to define as the Average of weight based Precision in MLIR System, it is calculated for all possible cut-off points in MLIR System and Eqn. (8) gives this measure.

$$NP_{MLIR} = \frac{1}{N_{l_i}} \sum_{j=1}^{N_{l_i}} P_{MLIR_j} \qquad (8)$$

where,' N_{l_i}' is the number of retrieved documents in different languages in MLIR System and 'P_{MLIR_j}' is the weight based precision at cut-off 'j'.

The main theme of our proposed work depends on the weight of the languages retrieved after a query is given to the underlying measurement systems. A table of comparisons is made to show the importance of weight based metrics in MLIR systems in terms of four significant factors of MLIR systems such as effectiveness, efficiency, scalability and performance of the systems as described below:

Comparison between Weight based measurement and Non-weight based measurement

- Weight based measurement have fixed value for each weight of language according to the number of languages where as Non-Weight Based Measurement does not have language weight scheme.
- Weight based measurement varies according to location wise (regional) language preferences but in Non-weight based measurement does not depend on the location.
- Weight of languages also varies according to location and preference where as Non-weight based Measurement there is no such variations.
- In Weight based measurement the value of weight is fixed between bounded values 0 to 1 where in Non-weight based measurement there is no such range of values is applicable.
- Weight based measurement has more relevant results. Non-weight based measurement has less relevant results.
- The performance measurement of weight based is giving effective results. In Non-weight based measurement scheme is less effective.

Efficiency based comparison of MLIR and IR System

- In MLIR System the efficiency is based on the right things as in context of accuracy and it is given good accuracy. MLIR System is displaying more accurate results then the traditional IR System.
- The time taken by MLIR System to retrieve documents is same as IR but the accuracy of relevancy is more.
- Precision metrics gives better results because of highly relevant documents are retrieved is more and whereas in IR System the Precision of IR is less efficient because of less accuracy of retrieve documents.
- Quantitatively the efficiency of MLIR is measure by the effective ratio of output to input. MLIR System gives better output ratio. Where as in IR System the quantitative measurement of efficiency is lower than MLIR because it gives less output ratio.
- MLIR System retrieves all documents in different languages simultaneously so it achieves the relevant documents more quickly than the IR System though it works on same language documents.
- MLIR System using weight based precisions metrics is based on the location related language preferences that gives better results. In IR System there is no weight based measurement of precisions are available.

Performance based comparison between MLIR and IR System

- In MLIR System the performance measurement require notion of relevancy and weight of different languages of retrieved documents. The performance measurement of IR is less accurate because of the only notion of relevancy.
- In MLIR System the resultant documents are retrieved from different language corpuses in parallel. In IR System documents retrieve from the single language corpus only.

Scalability based comparison between MLIR and IR System

- MLIR System can be extended more than three or many languages and weight of languages uniformly can be fixed according to number of languages in documents. In IR System it uses same language as monolingual languages. So it cannot be extended more than one languages.
- In MLIR System it can be added new functionality without any changes in the system. Where as in IR System it is very difficult to make changes.

Effectiveness based comparison between MLIR and IR System

- Effectiveness is based on the more relevant results so here in MLIR System it gives the more relevant results with the weight of languages of different documents. Where as in IR System it measure only based on relevancy so the result is less effective than MLIR System.
- MLIR System is based on the user needs it gives qualitative solutions. Where as in IR System, it gives less qualitative solutions.

The enhanced metrics are analytically evaluated against Weyuker's proposed set of measurement principles. These properties were developed to characterize superlative complexity metrics for computer based programs. Till now no single researcher has defined such axioms for information retrieval systems. In this work, we have taken first step to assess the enhanced metrics and know how these metrics are idealistic or meaningful to the IR research area. Weyuker's properties are relied on a number of operators and relations on the application domain.

Evaluation of Proposed Metrics on the basis of Wayuker's Properties:
Sanjay Misra, et al. (2009) proposed nine properties for procedural languages. These are designed to evaluate software complexity measures and applied on our proposed metrics which are as follows.

Property 1: Let μ be the metric of IR and MLIR systems where P defines the IR and Q defines the MLIR systems respectively. Suppose, 'a' and 'b' are the values derived from IR and MLIR metrics respectively and 'w' is the weight of languages, therefore, μ(P)=a for IR system and μ(Q) =wb for the MLIR system. Here, Q measures the weight based precision and gives comparatively better results. So that μ(P) ≠ μ(Q) =>a≠wb.

Property 2: Let 'c' be the non –negative number such that, μ(P)=c.

Property 3: Sometimes the relevancy of retrieved results may be lesser than IR; sometimes it may be equal for some queries. Thus, there may be some possible cases where, μ(P) = μ(Q).

Property 4: Functionality of both the existing IR System and the proposed MLIR System are same but the complexity will be different because

$$(\exists P)(\exists Q)(P \equiv Q) \& \mu(P) \neq \mu(Q)$$

Here this property is partially correct for the proposed metrics because, it follows both IR and MLIR systems for same input query but result of both system partially same in different documents.

Property 5: It is not satisfied for the given metrics because the concatenation of IR and MLIR systems works like the MLIR System. MLIR System also scalable for extending as single document so that the complexity value may not be increased while both systems are concatenated.

Property 6: Let μ(P) = a, μ(Q) = wb. Let another system μ(R) = c, so it can be written as:
μ(P+R) = a+c, μ(Q+R) = wb+c

From the above property, we can write as μ(P+R) ≠ μ(Q+R) because the weights of languages in MLIR System have different complexity in system.

Property 7: In both IR and MLIR systems, the results are retrieved from queries. The values of metrics change according to the rank of documents. So that μ(P) ≠ μ(Q).

Property 8: Renaming property is not appropriate because of both systems have comparatively dissimilar properties; therefore we cannot rename these systems.

Property 9: This property is not satisfied because the complexity of IR and MLIR Systems are different but when both systems are combined then it may provide equal complexity as of MLIR System because it can work for both as IR and MLIR due to its scalability. So

$$(\exists P)(\exists Q)\mu(P) + \mu(Q) <! \mu(P+Q) => \mu(Q)$$

From the above properties, one can clearly state that MLIR systems are very important and the traditional measures are incapable to demonstrate the effectiveness of these systems. Moreover language weight scheme is very much essential when the user wants to know the language wise precision.

IV. EXPERIMENTATION AND RESULT ANALYSIS

The web crawlers and Google Language Translation Tool have been utilized to test the aftereffects of the Proposed MLIR Metrics. English is utilized as a dialect to give inquiry and the removed archives are in different dialects like German, French and Chinese.

Table 1. Calculated Precision-IR and Precision-MLIR

Query	Precision-IR	Precision-MLIR
1	0.5600	0.6904
2	0.5400	0.5768
3	0.5400	0.6516
4	0.4800	0.5592
5	0.4400	0.592
6	0.4200	0.5774
7	0.4400	0.5706
8	0.5000	0.5638
9	0.4200	0.5974
10	0.6000	0.5714
11	0.5400	0.6508
12	0.5000	0.5438
13	0.5000	0.5216
14	0.4400	0.4846
15	0.5400	0.5296
16	0.5200	0.55
17	0.6400	0.7988
18	0.5800	0.6638
19	0.5400	0.484
20	0.4600	0.4982

In this paper, we have watched huge upgrades in the execution assessment of MLIR System while looking at the evaluation of both MLIR and IR frameworks. The trial results are appeared in Table 1. This table declares the computed accuracy of IR System and weight construct Precision of MLIR System with respect to 20 inquiries as Precision-IR and Precision-MLIR and graphical result is appeared in figure 1. Table 2 states the results of Average Precision of IR and Average Precision of MLIR. The graphical results of both average precisions are shown in figure 2. Table 3 represents the calculated results of P@k of IR System and MLIR System and graphical results of P@K of IR and MLIR System is shown in fig 3. Normalized Precision of IR and MLIR are defined as NP_{IR} and NP_{MLIR} In Table 4 the calculated results of NP_{IR} and NP_{MLIR} are shown and the graphical result of NP_{IR} and NP_{MLIR} can be seen in figure 4.The experimental results shows that the proposed weight based metrics of MLIR system are more effective and meaningful than the IR System.

Table 2. Calculated AP_{IR} and AP_{MLIR}

Query	AP_{IR}	AP_{MLIR}
1	0.6033	0.7089
2	0.7034	0.6258
3	0.7425	0.6391
4	0.5457	0.7051
5	0.7205	0.5602
6	0.5529	0.5535
7	0.6572	0.6031
8	0.6875	0.569
9	0.7585	0.6114
10	0.671	0.6829
11	0.4364	0.6642
12	0.6749	0.5326
13	0.7436	0.5705
14	0.7057	0.4508
15	0.6383	0.7704
16	0.7592	0.5965
17	0.8682	0.8132
18	0.8493	0.7449
19	0.8515	0.4895
20	0.6306	0.5926

From the above table and relative diagram, we can watch that for all the given inquiries accuracy MLIR performs well and gives higher qualities. This show the significance of new measurements in the appraisal of the MLIR frameworks. The calculated results of $P@K_{IR}$ and $P@K_{MLIR}$ is shown in Table 3. Figure 3 represents the results of $P@K_{IR}$ and $P@K_{MLIR}$ graphically where the results of $P@K_{IR}$ is not very effective than $P@K_{MLIR}$ except for the first case. At level 'k' the performance of the retrieval systems is improved because of the multiple languages along with weights. Therefore, the performance of $P@k_{MLIR}$ is increased when compared to $P@k_{IR}$. For all the queries the effectiveness is increased directly proportional to the k^{th} positions in the MLIR systems except in a few cases. Table 4 represents the calculated results of NP_{IR} and NP_{MLIR}. Figure 4 shows that results are effective than NP_{IR}. Last metric which we used for comparing MLIR and IR systems is NP. Like $P@K_{MLIR}$ and Precision-MLIR the performance of NP_{MLIR} is very much appreciated than the NP_{IR}. More than 50% of the queries perform better than NP_{IR}. This result shows the significance of assessing the MLIR systems is highly needed in this research field.

The experimentation and result examination proclaims the hugeness of the dialect weight plan. It is insufficient to acquaint this specific plan with just accuracy and its variation measures. It is additionally important to augment this plan with standard measures and their variations, for example, review, NDM, and so forth. The outcome analysis expresses the necessity of execution evaluation of MLIR frameworks.

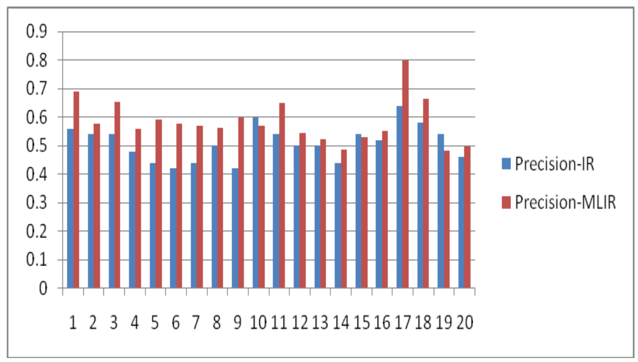

Figure 1. Calculated $precision_{IR}$ and $Precision_{MLIR}$

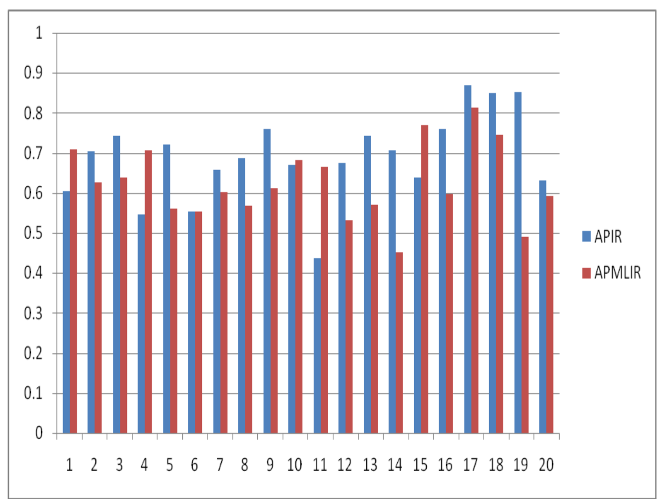

Figure 2. Calculated AP_{IR} and AP_{MLIR}

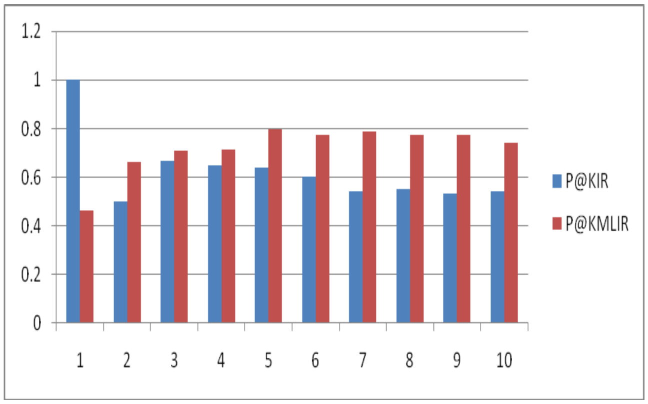

Figure 3. Calculated $P@K_{IR}$ and $P@K_{MLIR}$

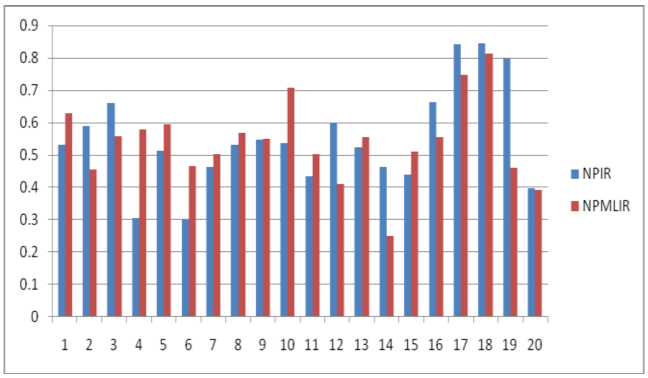

Figure 4. Calculated NP_{IR} and NP_{MLIR}

Table 4. Calculated NP_{IR} AND NP_{MLIR}

Query	NP_{IR}	NP_{MLIR}
1	0.5319	0.628
2	0.5884	0.4563
3	0.6611	0.5585
4	0.3063	0.5788
5	0.5128	0.5937
6	0.2990	0.4645
7	0.4639	0.5021
8	0.5308	0.5693
9	0.5470	0.5511
10	0.5358	0.7086
11	0.4345	0.5014
12	0.6008	0.4104
13	0.5224	0.5549
14	0.4619	0.2513
15	0.4400	0.5102
16	0.6629	0.5545
17	0.8414	0.7479
18	0.8440	0.8125
19	0.7971	0.4615
20	0.3982	0.3925

Table 3. Calculated $P@K_{IR}$ and $P@K_{MLIR}$

Query	$P@K_{IR}$	$P@K_{MLIR}$	P@K
1	1	0.4640	1
2	0.5000	0.6640	5
3	0.6667	0.7087	10
4	0.6500	0.7135	15
5	0.6400	0.7972	20
6	0.6000	0.7750	25
7	0.5429	0.7880	30
8	0.5500	0.7723	35
9	0.5333	0.7747	40
10	0.5400	0.7436	45

V. STATISTICAL ANALYSIS OF EXPERIMENTATION RESULTS

In this section, we present the details of statistical proof regarding our findings. The most suitable statistical tests for the Information retrieval domain are independent t-test and Wilcoxon Signed Rank Test. In the literature a few researchers have used these statistical tests to proof significance of their metrics. Hence, the same tests are performed to observe the significance level of the proposed measurements in this following section.

Independent t-test for Precisions: The independent t-test performs the comparison of the mean computed between two independent groups on the some continuous, dependent variable. The t-test procedure used to perform testing of equality of variances (Levene's test) and the *t*-value for both equal and unequal-variance Mendenhall et al. (1990).

Table 5. Statistics of Independent T-test of Precision of IR and Precision of MLIR

Statistics	N	Mean	Std. Error	t-value	df	P
Precision-IR	20	0.69001	0.011526	-2.070838	38	0.045216
Precision-MLIR	20	0.62421	0.008666			

Table 6. Statistics of Independent T-test of AP_{IR} and AP_{MLIR}

Statistics	N	Mean	Std. Error	t-value	df	P
AP_{IR}	20	0.51	0.003768	-3.36227	38	0.001774
AP_{MLIR}	20	0.58379	0.005865			

Table 7. Statistics of Independent T-test of $P@K_{IR}$ and $P@K_{MLIR}$

Statistics	N	Mean	Std. Error	t-value	df	P
$P@K_{IR}$	10	0.4977	0.023718	-1.76584	8	0.1154
$P@K_{MLIR}$	10	0.5404	0.01602			

Table 8. Statistics of Independent T-test of NP_{IR} and NP_{MLIR}

Statistics	N	Mean	Std. Error	t-value	df	P
NP_{IR}	10	0.62229	0.020821083	-1.23916	38	0.222886
NP_{MLIR}	10	0.7201	0.009859447			

The significance level (p-value) is always is lesser than 0.05 for independent t-test. The above tables shown in Table 5 and Table 6 provides the statistics where the p value is lesser than 0.05. So these results are considerable. On the other hand, Table 7 and 8 provide the p values which are closer to 0.05.

Wilcoxon Signed Rank Test: The use of this test is used to test the median difference in paired data. In paired data, the two groups which are naturally linked are compared and usually arise from individuals being measured in multiple times. In order to carry out this test, the difference between the pair data is calculated. Then rank the difference by their absolute value that is ignoring the sign, 1 is given for the smallest difference and 2 is given for the next smallest and so on. Then ranks of the positive differences and the ranks of the negative differences are summed up or lesser value of positive and negative sign will be taken.

Table 9. Test Statistics: Wilcoxon Signed Rank Test of Precision

Statistics	N	Mean of Rank	Positive Rank	Negative Rank	Sum of rank	Test Z
Precision -IR	20	0.5100	3	17	13 (Positive)	0.00451
Precision -MLIR	20	0.63637			175 (Negative)	

In the Wilcoxon Sign Rank Test method the significance level for above table i.e. p-value of the Precision MLIR metric is lesser than 0.05. The detailed statistics are shown in Table 9, where p-value is always lesser than 0.05. It indicates that our experimental analysis is considerable according to the experimented results.

Table 10. Test Statistics: Wilcoxon Signed Rank Test of AP

Statistics	N	Mean of Rank	Positive Rank	Negative Rank	Sum of rank	Test Z
AP_{IR}	20	0.6900	12	8	126 (Positive)	0.002576
AP_{MLIR}	20	0.6561			74 (Negative)	

In the Wilcoxon Sign Rank Test method the significance level for above table (AP_{MLIR}) is lesser than 0.05.The detailed statistics are shown in Table 10, where p-value is always lesser than 0.05. It indicates that our experimental analysis is considerable according to the experimented results. In the Wilcoxon Sign Rank Test method the p-value lesser than 0.05, it means that $P@K_{MLIR}$ is an important and significant metric to predict the performance of the MLIR systems.The detailed statistics are shown in Table 11, where p-value is always lesser than 0.05. It indicates that our experimental analysis is considerable according to the experimented results.

Table 11 Test Statistics: Wilcoxon Signed Rank Test of P@K

Statistics	N	Mean of Rank	Positive Rank	Negative Rank	Sum of rank	Test Z
P@K	10	0.6222	1	9	10 (+)	0.00347
$P@K_{MLIR}$	10	0.7201			45 (-)	

Table 12. Test Statistics: Wilcoxon Signed Rank Test of NP

Statistics	N	Mean of Rank	Positive Rank	Negative Rank	Sum of rank	Test Z
NP_{IR}	20	0.54901	9	11	99 (Positive)	0.0344
NP_{MLIR}	20	0.5404			101 (Negative)	

In the Wilcoxon Sign Rank Test method the significance level is lesser than 0.05. Therefore, NP_{MLIR} is an essential measure when compared to the counterpart. The thorough statistics are shown in Table 12, where p-value is always lesser than 0.05. It point towards that our experimental analysis is substantial according to the experimented results.

VI. CONCLUSION

In this paper, weight based precision oriented metrics are proposed and these metrics shows the better effectiveness to evaluate the MLIR systems over traditional metrics. Precision is the standard measure from the inception of IR. In this research we have enhanced the precision metrics by introducing language weight scheme. The Proposed MLIR Metrics are: Uniform Weight based Precision, Average Precision, P@k and Normalized Precision. The result analyses demonstrate the outstanding performance of these proposed metrics when compared to the traditional measures such as Precision, Average Precision, P@k and Normalized Precision. In future, we can extend this work to propose Weight based recall, F-measure and MAP for assessing the performance of MLIR systems when the language weight scheme is considered.

ACKNOWLEDGMENTS

This work is a part of the Research Project under the sponsorship of Major Research Project Scheme, UGC, India, and Reference No.: F. No. 41-640/2012 (SR) dt. 16-07-2012. Authors would like to express their sincere thanks for the financial support offered by the Sponsored Agency.

REFERENCES

[1] C. J. van Rijsbergen, et al., "Information Retrieval," Butterworths, London, 2nd edition, 1979.
[2] Chetana Sidige, et al., "Normalized Distance Measure: A Measure for Evaluating MLIR Merging Mechanisms," IJCSI International Journal of Computer Science Issues, Vol. 8, no. 1, pp. 209-214, 2011.
[3] Christian Fluhr, E. Robert, et al., "Chapter 2 Multilingual (or Cross-lingual) Information Retrieval," http://www.cs.cmu.edu/~ref/mlim/chapter2.html
[4] C.W. Cleverdon, et al., "Aslib Cranfield research project - Factors determining the performance of indexing systems," Vol. 2, 1966.
[5] Dan Wu, et al., "Cross-Language Query Expansion Using Pseudo Relevance Feedback," Journal of the Chinese Society for Scientific and Technical Information, Vol. 29, no.2, pp. 232-239, 2013.
[6] Davis, et al., "A TREC evaluation of query translation methods for multi-lingual text retrieval," Proceedings of the 5th Text Retrieval Conference (TREC), 1996.
[7] Ellen M. Voorhees, et al., "Overview of fifth Text Retrieval Conference," Proceedings of the 5th Text Retrieval Conference (TREC), 1995.
[8] EmineYilmaz, et al., "Estimating average precision when judgments are incomplete," Knowledge and Information Systems, Vol. 16, pp. 173-211, 2008.
[9] Hull D, et al. "Querying across Languages a dictionary-based approach to Multilingual Information Retrieval," Proceedings of the 19th Annual ACM Conference on Information Retrieval (SIGIR), pp. 49-57, 1996.
[10] M. Rogati, et al., "Multilingual Information Retrieval Using Open, Transparent Resources in CLEF 2003," Comparative Evaluation of Multilingual Information Access Systems, Vol. 3237, pp. 133-139, 2004.
[11] Mangala Madankar, et al., "Information Retrieval System and Machine Translation: A Review," International Conference on Information Security and Privacy (ICISP2015), Vol. 78, pp. 845-850, 2016.
[12] Olivier Chapelle, et al., "Gradient descent optimization of smoothed information retrieval metrics," Information Retrieval Journal, Vol.13, pp. 261-235.
[13] Parmatma Yadav, et al., "Weight based precision oriented metrics for MLIR systems," IEEE International Conference on Advanced Communication Control and Computing Technologies (ICACCCT), pp. 1114-1119, 2014.
[14] Pothula Sujatha, et al., "Performance Evaluation of Multilingual Information Retrieval (MLIR) System over Information Retrieval (IR) System," IEEE-International Conference on Recent Trends in Information Technology, ICRTIT 2011, pp. 722-727, 2011.
[15] Pothula Sujatha, et al., "Precision at k in multilingual information retrieval," International Journal of Computer Applications, Vol. 24, no. 9, pp. 40-43, 2011.
[16] R. R. Korfhage, "Information Storage and Retrieval," John Wiley & Sons, 1997.
[17] René Hackl, et al., "Multilingual Retrieval Experiments with MIMOR at the University of Hildesheim," Comparative Evaluation of Multilingual Information Access Systems, Vol. 3237, pp.166–173, 2004.
[18] Sanjay Misra, et al., "Weyuker's Properties, Language Independency and Object Oriented Metrics," Computational Science and Its Applications – ICCSA 2009, Vol. 5593, pp 70-81, 2009.
[19] Tetsuya Sakai, et al., "Ranking the NTCIR systems based on multigrade relevance," Proceedings of Asia information retrieval symposium, Vol. 3411, pp. 251-262, 2004.
[20] Tetsuya Sakai, et al., "Give me just one highly relevant document: P-measure," Proceedings of the 29th annual international ACM SIGIR conference on research and development in information retrieval (SIGIR 2006), pp. 695-696, 2006.
[21] Tetsuya Sakai, et al., "On information retrieval metrics designed for evaluation with incomplete relevance assessments,"Information Retrieval, Vol. 11, no.5, pp. 447–470, 2008.
[22] Toyin Enikuomehin and J.S. Sadiku, "Text Wrapping Approach to natural Language Information retrieval using significant Indicator", IAES International Journal of Artificial Intelligence (IJ-AI,) Vol. 2, No. 3, pp. 136-142, 2013.
[23] V. Ravisankar, et al., "A statistical Methodology for Evaluating MLIR Systems: Web Service based Analysis," International Journal of Applied Engineering Research, Vol. 10, no. 2, pp. 3337-3348, 2015.
[24] Vu, Gallinari, "On effectiveness measures and relevance functions in ranking INEX systems," Proceedings of Second Asia Information Retrieval Symposium, AIRS 2005, Vol. 3689, pp. 312–327, 2005.
[25] W. Mendenhall, et al., "Mathematical Statistics with applications," PWS-KENT Publishing Company, 1990.
[26] Yuemin Wang, "Design of Information Retrieval System Using Rough Fuzzy Set," TELKOMNIKA Indonesian Journal of Electrical Engineering, Vol. 12, no. 1, pp. 844-851, 2014.