**A Formal Study of Information Retrieval Heuristics**

The research problem shows that retrieval performance comes from the usage of various retrieval heuristics. This paper represents a formal study of retrieval heuristics by defining a set of constraints that are to be satisfied by the retrieval functions. The constraints are applied to various retrieval formulas methods like Pivoted normalization, Okapi method and Dirichlet prior whether they satisfy the constraints or not.

The challenges are none of these retrieval formulas satisfies all the constraints unconditionally. After a constraint is not satisfied, it often indicates non-optimality of the technique, and while a constraint is satisfied only for a range of parameter values, its performance tends to be poor for the values out of range.

In this approach, we formally outline six intuitive and acceptable constraints that any affordable retrieval formulation should satisfy. They seize the usually used retrieval heuristics, which includes TF-IDF weighting making it feasible to apply them analytically to any retrieval method. The constraints we are included for the model to satisfy are Term Frequency Constraints (TFCs), Term Discrimination Constraint (TDC), Length Normalization Constraint (LNC) and TF-Length Constraint (TF-LNC). We apply the six constraints to three retrieval formulas which represent the vector space model, the classical probabilistic retrieval model, and the language modeling approach. we test practically the retrieval function behavior by performing Pivoted Normalization Method, Okapi Method and Dirichlet Prior Method. In Pivot Normalization Method, TFCs and LNC1 are seen to be satisfied. If s is below a certain upper bound, then TF-LNC is satisfied. TDC is only satisfied conditionally. The performance of this method is bad for larger value of s. In Okapi method when IDF is negative then TFCs, LNCs and TF-LNC constraints are violated. Negative IDF is replaced with original IDF from pivoted normalization formula which satisfies all the constraints. when IDF is positive then TFC and LNC constraints are satisfied for okapi method. In Dirichlet prior method TFCs, LNC1 and TF-LNC are satisfied.

All the three methods have failed to satisfy the TDC unconditionally. Due to the possible negative IDF part in the Okapi formula, the performance of Okapi is bad than others for verbose queries. Okapi method performance is not changed by parameter setting than pivoted normalization and Dirichlet Prior Method.

In this Experimental design, the constraints provide good explanation and make it possible for evaluating analytically, existing or new retrieval formula and helps in testing for the new retrieval models. Parameter Sensitivity gives an idea about the performance of retrieval formula on changing the parameter values. The better performance is achieved by the retrieval model which satisfies the more constraints. Performance comparison between the models suggests the way to improve the retrieval formula.

The research has weakness which gives interesting ideas for future directions. The set of test collections or documents can be used to repeat the experiment by removed all stop words by listing them which can affect the desired outcomes. The experiment constraints did not cover all the properties, it would be interesting to discover additional important heuristics for a reasonable retrieval method. This will help us to understand the overall performance of different retrieval methods. we will apply the constraints to different retrieval models in the writings and some smoothing methods for language models as well. however, the outcomes for the retrieval models are not certain and performs differently in certain constraints.